FOURCASTNET: A DATA-DRIVEN MODEL FOR HIGH-RESOLUTION WEATHER FORECASTS USING ADAPTIVE FOURIER NEURAL OPERATORS

Jaideep Pathak
NVIDIA Corporation
Santa Clara, CA 95051

Shashank Subramanian
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Peter Harrington
Lawrence Berkeley
National Laboratory
Berkeley, CA 94720

Sanjeev Raja
University of Michigan
Ann Arbor, MI 48109

Ashesh Chattopadhyay
Rice University
Houston, TX 77005

Morteza Mardani
NVIDIA Corporation
Santa Clara, CA 95051

Thorsten Kurth
NVIDIA Corporation
Santa Clara, CA 95051

David Hall
NVIDIA Corporation
Santa Clara, CA 95051

Zongyi Li
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

Kamyar Azizzadenesheli
Purdue University
West Lafayette, IN 47907

Pedram Hassanzadeh
Rice University
Houston, TX 77005

Karthik Kashinath
NVIDIA Corporation
Santa Clara, CA 95051

Animashree Anandkumar
California Institute of Technology
Pasadena, CA 91125
NVIDIA Corporation
Santa Clara, CA 95051

ABSTRACT

FourCastNet, short for *Four*ier Fore*Cast*ing Neural *Net*work, is a global data-driven weather forecasting model that provides accurate short to medium-range global predictions at $0.25°$ resolution. FourCastNet accurately forecasts high-resolution, fast-timescale variables such as the surface wind speed and precipitation. It has important implications for planning wind energy resources, predicting extreme weather events such as tropical cyclones, extra-tropical cyclones, and atmospheric rivers. FourCastNet matches the forecasting accuracy of the ECMWF Integrated Forecasting System (IFS), a state-of-the-art Numerical Weather Prediction (NWP) model, at short lead times for large-scale variables, while outperforming IFS for variables with complex fine-scale structure, including precipitation. FourCastNet generates a week-long forecast in less than 2 seconds, orders of magnitude faster than IFS. The speed of FourCastNet enables the creation of rapid and inexpensive large-ensemble forecasts with thousands of ensemble-members for improving probabilistic forecasting. We discuss how data-driven deep learning models such as FourCastNet are a valuable addition to the meteorology toolkit to aid and augment NWP models.

The beginnings of modern numerical weather prediction (NWP) can be traced to the 1920s. Now ubiquitous, they contribute to economic planning in key sectors such as transport, logistics, agriculture, and energy production. Accurate weather forecasts have saved countless human lives by providing advance notice of extreme events. The quality of weather forecasts has been steadily improving over the past decades (c.f. Bauer et al. (2015); Alley et al. (2019)).

There is now increasing interest around developing data-driven DL-based models for weather forecasting owing to their orders of magnitude lower computational cost as compared to state-of-the-art NWP models (Schultz et al., 2021; Balaji, 2021; Irrgang et al., 2021; Reichstein et al., 2019). Many studies have attempted to build data-driven models for forecasting the large-scale circulation of the atmosphere, either trained on climate model outputs, general circulation models (GCM) (Scher & Messori, 2018; 2019; Chattopadhyay et al., 2020), reanalysis products (Weyn et al., 2019; 2020; 2021; Rasp et al., 2020; Rasp & Thuerey, 2021a; 2020; Chattopadhyay et al., 2021; Arcomano et al., 2020; Chantry et al., 2021; Grönquist et al., 2021), or a blend of climate model outputs and reanalysis products (Rasp & Thuerey, 2021a).

By training on reanalysis data or observations, data-driven models have great potential to improve weather predictions by overcoming model biases present in NWP models(Schultz et al., 2021; Balaji, 2021). They also enable the generation of large ensembles at low computational cost for probabilistic forecasting and data assimilation (Chattopadhyay et al., 2021; Weyn et al., 2021; Chattopadhyay et al., 2020).

Most data-driven weather models, however, use low-resolution data for training, usually at the $5.625°$ resolution as in Rasp & Thuerey (2021b) or $2°$ as in Weyn et al. (2020). However, the coarsening procedure leads to the loss of crucial, fine-scale physical information. For data-driven models to be truly impactful, it is essential that they generate forecasts at resolutions equal to or greater than current state-of-the-art numerical weather models, which are run at $\approx 0.1°$ resolution. Forecasts at $5.625°$ spatial resolution, for instance, result in a mere 32×64 pixels grid representing the entire globe leading to limited practical utility. High-resolution models can resolve the formation and dynamics of high-impact extreme events such as tropical cyclones, which are inadequately represented on a coarser grid.

Our approach: We develop FourCastNet, a Fourier-based neural network forecasting model, to generate global data-driven forecasts of key atmospheric variables at a resolution of $0.25°$, or about 30 km $\times$ 30 km near the equator and a global grid size of 720×1440 pixels. This allows us, for the first time, to make a direct comparison with the high-resolution Integrated Forecasting System (IFS) model of the European Center for Medium-Range Weather Forecasting (ECMWF).

Figure 1 shows an illustrative global near-surface wind speed forecast at a 96-hour lead time generated using FourCastNet. We highlight key high-resolution details that are resolved and accurately tracked by our forecast, including Super Typhoon Mangkhut and three named cyclones heading towards the eastern coast of the United States (Florence, Issac, and Helene).

FourCastNet uses a Fourier transform-based token-mixing scheme (Guibas et al., 2022) with a vision transformer (ViT) backbone (Dosovitskiy et al., 2021). This approach is based on the recent Fourier neural operator that learns in a resolution-invariant manner and has shown success in modeling challenging partial differential equations (PDE) such as fluid dynamics (Li et al., 2021). Combining a ViT backbone with Fourier-based token mixing yields a state-of-the-art high-resolution model that resolves fine-grained features, models long-range dependencies accurately, and scales well with resolution and size of dataset[1].

In summary, FourCastNet makes four significant contributions to data-driven weather forecasting:

1. FourCastNet predicts challenging variables with complex fine-scale structure such as surface winds and precipitation, with unparalleled accuracy at forecast lead times of up to one week. No deep learning (DL) model thus far has attempted to forecast surface winds on global scales. Additionally, DL models for precipitation on global scales have been inadequate for resolving fine-scale structures.
2. FourCastNet, at $0.25°$ resolution, has eight times greater resolution than state-of-the-art DL-based global weather models. FourCastNet accurately resolves extreme weather patterns such as tropical cyclones and atmospheric rivers that have been inadequately represented by prior DL models owing to their coarser grids.

[1] We estimate that FourCastNet could be trained on currently available GPU hardware in about two months with 40 years of global 5-km data, if such data were available.

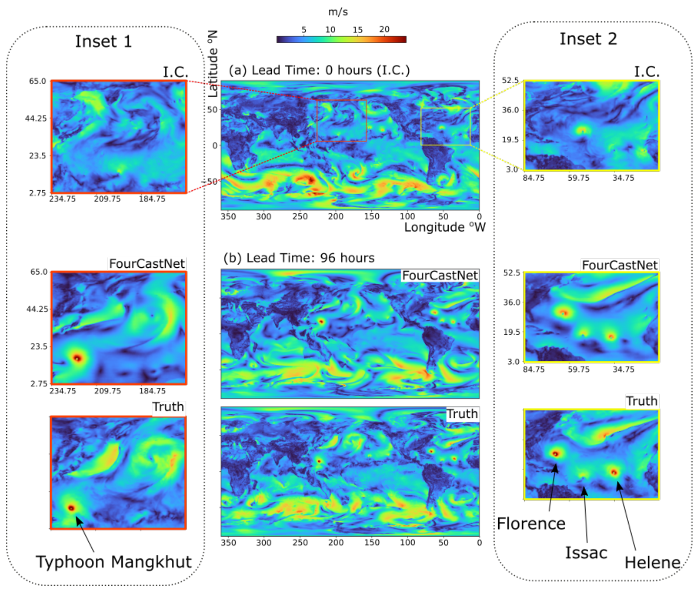

Figure 1: Illustrative example of a global near-surface wind forecast generated by FourCastNet at a resolution of 0.25°. We initialize FourCastNet with an initial condition from the out-of-sample test dataset (September 8, 2018 at 00:00 UTC). Starting from this initial condition, the model was allowed to run freely for 16 time-steps of six hours each corresponding to a 96-hour forecast. Panel (a) shows the wind speed at model initialization. Panel (b) shows the model forecasts at forecast lead time of 96 hours (upper panel) and the corresponding true wind speeds at that time (lower panel). FourCastNet is able to forecast the wind speeds 96 hours in advance with remarkable fidelity and correct fine-scale features. The forecast accurately captures the formation, track, and intensification of Super Typhoon Mangkhut and three named hurricanes (Florence, Issac, and Helene) forming in the Atlantic Ocean and approaching the eastern coast of North America (see Inset 2).

Vertical Level	Variables	Vertical Level	Variables
Surface	U_{10}, V_{10}, T_{2m}, sp, $mslp$	$1000hPa$	U, V, Z
$850hPa$	T, U, V, Z, RH	$500hPa$	T, U, V, Z, RH
$50hPa$	Z	Integrated	$TCWV$

Table 1: Prognostic Variables modeled by the DL model. Abbreviations are as follows. U_{10} (V_{10}): zonal (meridonal) wind velocity 10m from the surface; T_{2m}: Temperature at 2m from the surface; T, T, V, Z. RH: Temperature, zonal velocity, meridonal velocity, geopotential, relative humidity respectively at specified vertical level; $TCWV$: Total Column Water Vapor.

3. FourCastNet's predictions are comparable to the IFS model on metrics of Root Mean Squared Error (RMSE) and Anomaly Correlation Coefficient (ACC) at lead times of up to three days. After that, predictions of all modeled variables lag close behind IFS at lead times of up to a week. The IFS model has been developed over decades, contains over 150 variables each with more than 50 vertical levels in the atmosphere, and is guided by physics. In contrast, FourCastNet models 20 variables at five vertical levels, and is purely data driven.

4. FourCastNet's reliable, rapid (about 45,000 times faster than traditional NWP), and computationally inexpensive (12,000 times cheaper than NWP) forecasts facilitate the generation of very large ensembles. This enables estimation of well-calibrated and constrained uncertainties in extremes with higher confidence than current NWP ensembles, which have at most 50 members owing to their high computational cost. Fast generation of 1,000-member ensembles dramatically changes what is possible in probabilistic weather forecasting.

1 TRAINING AND INFERENCE

We train the FourCastNet model on the ERA5 (Hersbach et al., 2020) reanalysis dataset from the years 1979 to 2015. Further methodological details are provided in Appendix A. We generate forecasts of a few core atmospheric variables listed in Table 1 and the total precipitation by using our trained models in autoregressive inference mode. The model is initialized with an initial condition ($\mathbf{X}_{\text{true}}(j)$) from the year 2018 [2], which is part of the out-of-sample held out dataset, for a large number of different initial conditions and allowed to freely run iteratively for τ time-steps to generate forecasts $\{\mathbf{X}_{\text{pred}}(j+i\Delta t)\}_{i=1}^{\tau}$. We also use the IFS forecasts for the year 2018 from The International Grand Global Ensemble (TIGGE) archive for comparative analysis. The archived IFS forecasts, with initial conditions matching the times of corresponding initial conditions for the FourCastNet model forecast, are used to compare our model's accuracy to that of the IFS model.

2 RESULTS

Figure 1 qualitatively shows the skill of FourCastNet when forecasting near-surface wind speeds over the entire globe at a resolution of 0.25°-lat-long. The wind speeds are computed as the magnitude of the surface wind velocity components ($\sqrt{(U_{10}^2 + V_{10}^2)}$) included in the FourCastNet model backbone. We initialize the FourCastNet model with an initial condition from the out-of-sample test dataset. Starting from this initial condition (September 8, 2018 at 00:00 UTC), the model is allowed to run freely for 16 time-steps in inference mode (Figure 4(d)). Figure 1(a) shows the wind speed at model initialization. Figure 1(b) shows the model forecasts at a lead time of 96 hours (upper-panel) and the corresponding true wind speeds at that time (lower-panel). We note that the FourCastNet model is able to forecast the wind speeds up to 96 hours in advance with remarkable fidelity and accurate fine-scale features. Notably, this figure illustrates the FourCastNet forecast of the formation and track of a super-typhoon named Mangkhut along with three simultaneous hurricanes (Florence, Issac and Helene) in the Atlantic ocean.

2.1 HURRICANES

A rapidly available, computationally inexpensive atmospheric model that could could forewarn the possibility of hurricane formation and track the path of the hurricane would be of great utility for mitigating loss of life and property damage. As the stakes for mis-forecasting such extreme weather phenomena are very high, more rigorous studies need to be undertaken before DL can be considered a mature technology for hurricane forecasting. As a case-study we consider a hurricane that occurred in 2018 (a year that is part of our out-of-sample dataset), namely hurricane Michael. Michael was a category 5 hurricane on the Saffir -Simpson Hurricane Wind Scale that made landfall in Florida causing catastrophic damage (Beven II et al., 2019). Within a short period of roughly 72 hours starting October 7, 2018, Michael went from a tropical depression to a category 5 hurricane at landfall.

We applied our trained model as described in Section A.2 (with no further changes) to study its potential for forecasting the formation, rapid intensification and tracking of hurricane Michael. We started with the initial condition at calendar time 00:00 hours on October 7, 2018 UTC. This state was perturbed with Gaussian noise to generate an ensemble of $E = 100$ perturbed initial conditions. Figure 2 shows the track of the hurricane and the intensification as forecast by the 100-member FourCastNet ensemble using the Mean Sea Level Pressure to estimate the position of the hurricane's eye and the minimum pressure at its center. Figure 2(a) shows the mean position of the minima of Mean Sea Level Pressure using a 100 member ensemble forecast generated by FourCastNet (red circles). The corresponding ground truth according to ERA5 reanalysis is indicated on the same plot (blue squares) over a trajectory spanning 108 hours. The shaded ellipses in the figure have a width and height equal to the 90th percentile spread in the longitudinal and latitudinal positions respectively of the hurricane eye as indicated by the MSLP minima in the 100-member FourCastNet ensemble. Fig. 2(b) shows the 850hPa wind speed forecast using FourCastNet and corresponding ground truth at lead times of 18hr, 36hr, 54hr and 72hr.

[2] The year 2018 was chosen from the out-of-sample dataset due to ready availability of IFS forecasts for that year from the TIGGE archive.

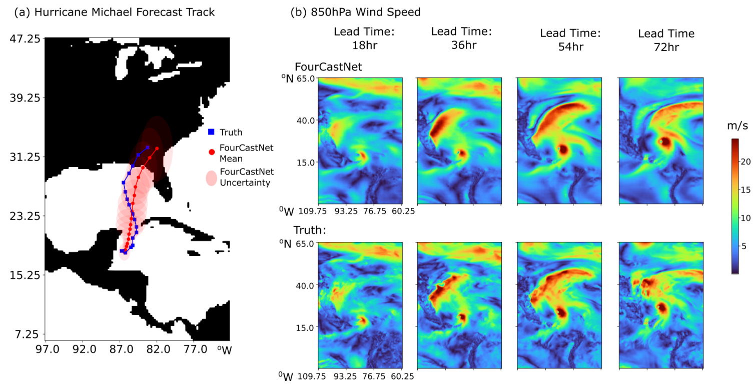

Figure 2: The FourCastNet model has excellent skill on forecasting fine-scale, rapidly changing variables relevant to a hurricane forecast. As an illustrative example, we have chosen Hurricane Michael which underwent rapid intensification during the course of its four day trajectory. Panel (a) shows the mean position of the minima of Mean Sea Level Pressure (indicating the eye of hurricane Michael) as forecast by a 100 member ensemble forecast using FourCastNet (red circles) and the corresponding ground truth according to ERA5 reanalysis (blue squares) for 108 hours starting from the initial condition at 00:00 hours on October 7, 2018 UTC. The shaded ellipses have a width and height equal to the 90th percentile spread of the longitudinal and latitudinal positions respectively of the hurricane eye as indicated by the MSLP minima in the 100-member FourCastNet ensemble. Panel (b) shows the 850hPa wind speed forecast and corresponding ground truth at forecast lead times of 18hr, 36hr, 54hr and 72hr lead times.

2.2 Quantitative comparison to IFS

We illustrate the forecast skill of our model for several initial conditions from the out-of-sample dataset (consisting of the year 2018) and generate a forecast for each initial condition. For each forecast, we evaluate the latitude-weighted Anomaly Correlation Coefficient (ACC) for all of the variables included in the forecast. We report the mean ACC for key variables along with the first and third quartile values of the ACC at each forecast time step, to show the dispersion of these metrics over different initial conditions. As a comparison, we also compute the same ACC metrics for the corresponding IFS forecast with time-matched initial conditions.

Figure 3(a-f) shows the latitude weighted ACC for the FourCastNet model forecasts (Red line with markers) and the corresponding matched IFS forecasts (Blue line with markers) for the variables (a) U_{10}, (b) TP, (c) T_{2m}. The ACC values are averaged over several initial conditions (U_{10}, TP: 180 I.C; T_{2m}: 40 I.Cs). The shaded regions around the ACC curves indicate the region between the first and third quartile values of the corresponding quantity at each time step. In general, the FourCastNet predictions are very competitive with IFS, with our model achieving similar ACC and RMSE over a horizon of several days. At shorter lead times ($\sim$ 48hrs or less), we actually outperform the IFS model for key variables like precipitation, winds, and temperature. Remarkably, we achieve this accuracy using only part of the full variable set available to the IFS model, and we do so at a fraction of the compute cost (see Appendix A.5 for a detailed speed comparison between models).

3 Discussion, Conclusions and Future Work

FourCastNet's predictions are orders of magnitude faster than traditional NWP models. This can facilitate large ensemble forecasts of thousands of members generated in seconds. Current NWP ensembles have at most approximately 50 members owing to their high computational cost. Such large-ensemble probabilistic weather forecasting can improve reliability of early warnings of extreme weather events. The unprecedented accuracy in short-range forecasts of precipitation and its extremes has potentially massive benefits for society such as enabling a rapid response to flooding. For the wind energy industry, FourCastNet's rapid and reliable high-resolution wind forecasts can help plan for fluctuations in wind power output. Due to the absence of a data-assimilation component, FourCastNet cannot yet generate up-to-the-minute weather forecasts. If observations are avail-

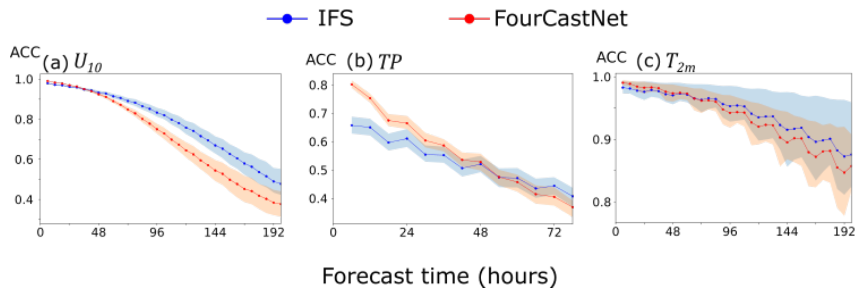

Figure 3: Latitude weighted ACC for the FourCastNet model forecasts (red line with markers) and the corresponding matched IFS forecasts (blue line with markers) averaged over several forecasts initialized using initial conditions in the out-of-sample testing dataset corresponding to the calendar year 2018 for the variables (a) U_{10}, (b) TP, (c) T_{2m}. The ACC values are averaged over many initial conditions (T_{2m}: 40 I.Cs; TP, U_{10}: 180 I.Cs) over a full year to account for seasonal variability in forecast skill. The appropriately colored shaded regions around the ACC curves indicate the region between the first and third quartile values of the corresponding quantity at each time step.

able, however, such a component could be readily incorporated. This will enable real-time weather prediction by initializing the model with real-time observations. FourCastNet's skill improves as the number of modeled variables increases. A larger model trained on more variables, perhaps even entire 3D atmospheric fields, may extend prediction horizons still further and with better uncertainty estimates. In the near future, FourCastNet could be trained on the entire nine petabyte ERA5 dataset to predict all currently predicted variables in NWP at every atmospheric level. Although the cost of training such a model will be huge, fast inference would enable rapid prediction of the full 3D fields in a few seconds. Such an advancement will likely revolutionize weather prediction.

Acknowledgements

We would like to acknowledge helpful comments and suggestions by Peter Dueben from ECMWF. We thank the researchers at ECMWF for their open data sharing and maintaining the ERA5 dataset without which this work would not have been possible. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. A.K. and P. Hassanzadeh were partially supported by ONR grant N00014-20-1-2722. We thank the staff and administrators of the Perlmutter computing cluster at NERSC, NVIDIA Selene computing cluster administrators, Atos, and Jülich Supercomputing Center for providing computing support. JP and KK would like to thank Sanjay Choudhry and the NVIDIA Modulus team for their support. JP, SS, P. Harrington and KK would like to thank Wahid Bhimji for helpful comments.

References

Richard B Alley, Kerry A Emanuel, and Fuqing Zhang. Advances in weather prediction. *Science*, 363(6425):342–344, 2019.

Troy Arcomano, Istvan Szunyogh, Jaideep Pathak, Alexander Wikner, Brian R Hunt, and Edward Ott. A machine learning-based global atmospheric forecast model. *Geophysical Research Letters*, 47(9):e2020GL087776, 2020.

V Balaji. Climbing down charney's ladder: machine learning and the post-dennard era of computational climate science. *Philosophical Transactions of the Royal Society A*, 379(2194):20200085, 2021.

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

Peter Bauer, Tiago Quintino, Nils Wedi, Antonio Bonanni, Marcin Chrust, Willem Deconinck, Michail Diamantakis, Peter Düben, Stephen English, Johannes Flemming, et al. *The ecmwf scalability programme: Progress and plans*. European Centre for Medium Range Weather Forecasts, 2020. doi:10.21957/gdit22ulm. URL `https://www.ecmwf.int/node/19380`.

J.L. Beven II, R. Berg, and A. Hagen. Tropical cyclone report hurricane michael, April 2019.

Matthew Chantry, Hannah Christensen, Peter Dueben, and Tim Palmer. Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft ai. *Philosophical Transactions of the Royal Society A*, 379(2194):20200083, 2021.

Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001958, 2020.

Ashesh Chattopadhyay, Mustafa Mustafa, Pedram Hassanzadeh, Eviatar Bach, and Karthik Kashinath. Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving spatial transformers in a case study with era5. *Geoscientific Model Development Discussions*, pp. 1–23, 2021.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2021.

Peter Grönquist, Chengyuan Yao, Tal Ben-Nun, Nikoli Dryden, Peter Dueben, Shigang Li, and Torsten Hoefler. Deep learning for post-processing ensemble weather forecasts. *Philosophical Transactions of the Royal Society A*, 379(2194):20200092, 2021.

John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive Fourier Neural Operators: Efficient token mixers for transformers. *International Conference on Representation Learning (to appear)*, April 2022.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. ISSN 1477-870X.

Christopher Irrgang, Niklas Boers, Maike Sonnewald, Elizabeth A Barnes, Christopher Kadow, Joanna Staneva, and Jan Saynisch-Wagner. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, 3(8):667–674, 2021.

Eugenia Kalnay, Masao Kanamitsu, Robert Kistler, William Collins, Dennis Deaven, Lev Gandin, Mark Iredell, Suranjana Saha, Glenn White, John Woollen, et al. The ncep/ncar 40-year reanalysis project. *Bulletin of the American meteorological Society*, 77(3):437–472, 1996.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Stephan Rasp and Nils Thuerey. Purely data-driven medium-range weather forecasting achieves comparable skill to physical models at similar resolution. *arXiv preprint arXiv:2008.08626*, 2020.

Stephan Rasp and Nils Thuerey. Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, 13(2):e2020MS002405, 2021a.

Stephan Rasp and Nils Thuerey. Data-driven medium-range weather prediction with a resnet pre-trained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, pp. e2020MS002405, 2021b.

Stephan Rasp, Peter D Dueben, Sebastian Scher, Jonathan A Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11):e2020MS002203, 2020.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, et al. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

Sebastian Scher and Gabriele Messori. Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, 144(717):2830–2841, 2018.

Sebastian Scher and Gabriele Messori. Weather and climate forecasting with neural networks: using general circulation models (gcms) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809, 2019.

MG Schultz, C Betancourt, B Gong, F Kleinert, M Langguth, LH Leufen, Amirpasha Mozaffari, and S Stadtler. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A*, 379(2194):20200097, 2021.

Jonathan A Weyn, Dale R Durran, and Rich Caruana. Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11(8):2680–2693, 2019.

Jonathan A Weyn, Dale R Durran, and Rich Caruana. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12(9):e2020MS002109, 2020.

Jonathan A Weyn, Dale R Durran, Rich Caruana, and Nathaniel Cresswell-Clay. Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *arXiv preprint arXiv:2102.05107*, 2021.

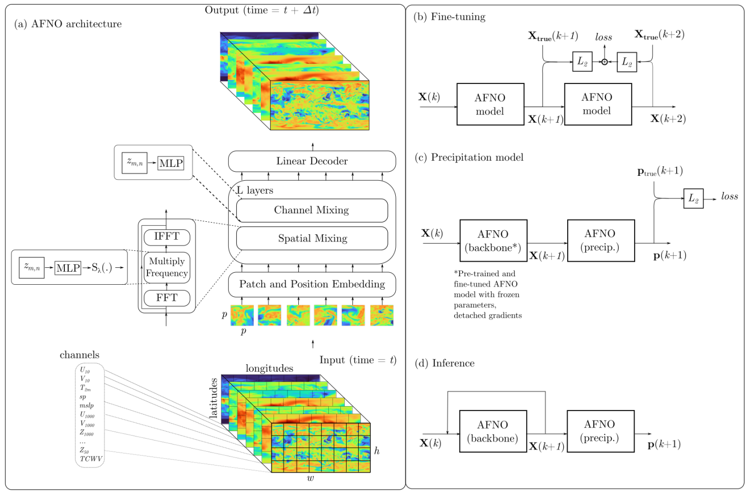

Figure 4: (a) The multi-layer transformer architecture that utilizes the Adaptive Fourier Neural Operator with shared MLP and frequency soft-thresholding for spatial token mixing. The input frame is first divided into a $h \times w$ grid of patches, where each patch has a small size $p \times p \times c$. Each patch is then embedded in a higher dimensional space with high number of latent channels and position embedding is added to form a sequence of tokens. Tokens are then mixed spatially using AFNO, and subsequently for each token the latent channels are mixed. This process is repeated for L layers, and finally a linear decoder reconstructs the patches for the next frame from the final embedding. The right-hand panels describe the FourCastNet model's additional training and inference modes: (b) two-step fine-tuning, (c) backbone model that forecasts the 20 variables in Table 1 with secondary precipitation diagnostic model (note that $\mathbf{p}(k+1)$ denotes the 6 hour accumulated total precipitation that falls between $k+1$ and $k+2$ time steps) (d) forecast model in free-running autoregressive inference mode.

A Appendix

A.1 FourCastNet: Model Description

To produce our high-resolution forecasts, we choose the Adaptive Fourier Neural Operator (AFNO) model (Guibas et al., 2022). This particular neural network architecture is appealing as it is specifically designed for *high-resolution* inputs and synthesizes several key recent advances in DL into one model. Namely, it combines the Fourier Neural Operator (FNO) learning approach of Li et al. (2021), which has been shown to perform well in modeling challenging PDE systems, with a powerful ViT backbone. The AFNO model is unique in that it frames the mixing operation as continuous global convolution, implemented efficiently in the Fourier domain with FFTs, which allows modeling dependencies across spatial and channel dimensions flexibly and scalably. With such a design, the spatial mixing complexity is reduced to $\mathcal{O}(N \log N)$, where N is the number of image patches or tokens. This scaling allows the AFNO model to be well-suited to high-resolution data at the current 0.25° resolution considered in this paper as well as potential future work at an even higher resolution. In the original FNO formulation, the operator learning approach showed impressive results solving turbulent Navier-Stokes systems, so incorporating this into a data-driven atmospheric model is a natural choice.

We refer the reader to the original AFNO paper (Guibas et al., 2022) for more details on the AFNO architecture and provide an illustration in Fig. 4(a).

A.2 Training

The ECMWF provides a publicly available, comprehensive reanalysis (Kalnay et al., 1996) dataset called ERA5 (Hersbach et al., 2020) which consists of hourly estimates of several atmospheric variables at a latitude and longitude resolution of 0.25° from the surface of the earth to roughly 100 km altitude from 1979 to the present day. We use a subset of the ERA5 dataset to train FourCastNet.

In order to model complex atmospheric interactions, we choose a few variables (Table 1) to represent the instantaneous state of the atmosphere. Each of the 20 variables is represented as a 2D field of shape (721×1440) pixels. Thus, a single training data point at an instant in time containing all 20 variables is represented by a tensor of shape ($721 \times 1440 \times 20$). While the ERA5 dataset is available at a temporal resolution of 1 hour, we choose to sub-sample the dataset and use snapshots spaced 6 hours apart to train our model. We divide the dataset into three sets, namely training, validation and out-of-sample testing datasets. The training dataset consists of data from the year 1979 to 2015 (both included). The validation dataset contains data from the years 2016 and 2017. The out-of-sample testing dataset consists of the years 2018 and beyond.

We collectively denote the modeled variables by the tensor $\mathbf{X}(k\Delta t)$, where k denotes the time index and Δt is the temporal spacing between consecutive snapshots in the training dataset. We will consider the ERA5 dataset as the truth and denote the *true* variables by $\mathbf{X}_{\text{true}}(k\Delta t)$. With the understanding that Δt is fixed at 6 hours throughout this work, we omit Δt in our notation for convenience where appropriate. The training procedure consists of two steps, pre-training and fine-tuning. In the pre-training step, we train the AFNO model using the training dataset in a supervised fashion to learn the mapping from $\mathbf{X}(k)$ to $\mathbf{X}(k+1)$. We then fine-tune the model using a two-step rollout to ensure stability. The end to end training takes about 16 hours wall-clock time on a cluster of 64 Nvidia A100 GPUs.

A.3 Precipitation Model

The total precipitation (TP) in the ERA5 re-analysis dataset is a variable that represents the the accumulated liquid and frozen water that falls to the Earth's surface through rainfall and snow. Compared to the variables handled by our backbone model, TP exhibits more sparse spatial features than the other prognostic variables. For these reasons, we treat the total precipitation (TP) as a diagnostic variable and denote it by $\mathbf{p}(k\Delta t)$. Total precipitation is not included in the 20 variable dataset used to train the backbone model[3]. Rather, we train a separate AFNO model to diagnose TP using the outputs of the backbone model, as indicated in Figure 4(c). In addition, once trained, our diagnostic TP model could potentially be used in conjunction with other forecast models (either traditional NWP or data-driven forecasts).

A.4 Inference

We generate forecasts of the core atmospheric variables in Table 1 and the total precipitation by using our trained models in autoregressive inference mode as shown in Figure 4(d). The model is initialized with an initial condition ($\mathbf{X}_{\text{true}}(j)$) from the year 2018 [4] out-of-sample held out dataset for N_f different initial conditions and allowed to freely run iteratively for τ time-steps to generate forecasts $\{\mathbf{X}_{\text{pred}}(j+i\Delta t)\}_{i=1}^{\tau}$. We also use the IFS forecasts for the year 2018 from The International Grand Global Ensemble (TIGGE) archive for comparative analysis. The archived IFS forecasts, with initial conditions matching the times of corresponding initial conditions for the FourCastNet model forecast, are used for comparing our model's accuracy to that of the IFS model.

A.5 Computational Cost of FourCastNet

To estimate the forecast speed of the IFS model, we use figures provided in Bauer et al. (2020) as a baseline. In Ref. (Bauer et al., 2020), we see that the IFS model computes a 15-day, 51-member ensemble forecast using the "L91" 18km resolution grid on 1530 Cray XC40 nodes with dual socket Intel Haswell processors in 82 minutes. The IFS model archived in TIGGE, which we compare the FourCastNet predictions with in Section 2.2, also uses the L91 18km grid for computation (but is archived at the ERA5 resolution of 30km). Based on this information, we estimate the compute-cost and energy consumption of the IFS for a 100-member 24-hour ensemble forecast[5]. The FourCastNet model can compute a 100-member 24-hour forecast in 7 seconds by using a single node on the

[3] This approach is similar to previous work (Rasp & Thuerey, 2021b), which trained a separate model for precipitation than for the other atmospheric variables.

[4] The year 2018 was chosen from the out-of-sample dataset due to ready availability of IFS forecasts for that year from the TIGGE archive.

[5] A dual-socket Intel Haswell node draws a Thermal Design Power (TDP) of 270 Watts

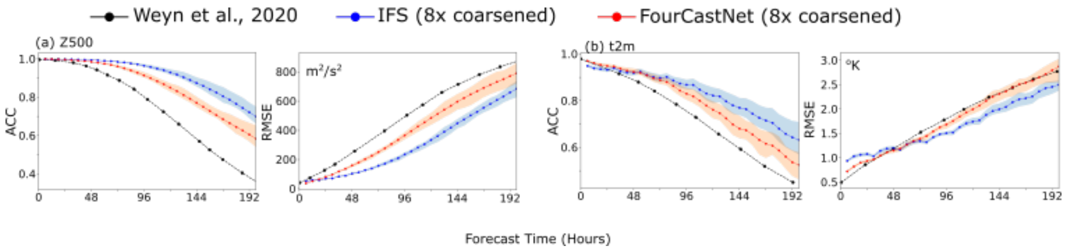

Figure 5: Comparison of ACC and RMSE metrics between the (downsampled) FourCastNet predictions, (downsampled) IFS, and baseline state-of-the-art DLWP model (Weyn et al., 2020) for (a) Z_{500} and (b) T_{2m}. We observe that the FourCastNet predictions show significant improvement over the baseline model. We also note that the FourCastNet generates predictions that have a higher resolution by a factor of 8, and is thus able to resolve many more important fine-scale features than the DLWP model.

Perlmutter HPC cluster which contains 4 A100 GPUs per node and has a peak power consumption of 1kW. We attempt to account for the resolution difference between the 18km L91 model and the 30km FourCastNet model by additionally reporting the inference time for an 18km FourCastNet model by interpolating the model parameters and inputs. These calculations, tabulated in Table. 2 show that the FourCastNet model at 30km resolution is about 145,000 times faster than the IFS whereas an interpolated FourCastNet model at 18km resolution would be 45,000 times faster than the IFS.

Latency and Energy consumption for a 24-hour 100-member ensemble forecast				
	IFS	FCN - 30km (actual)	FCN - 18km (extrapolated)	IFS / FCN(18km) Ratio
Nodes required	3060	1	2	**1530**
Latency (Node-seconds)	984000	7	22	**44727**
Energy Consumed (kJ)	271000	7	22	**12318**

Table 2: The FourCastNet model is about 45,000 times faster than the IFS model on a node-hour basis and uses about 12,000 times less energy to generate a forecast.

A.6 Comparison Against State-of-the-art DL Weather Prediction

To the best of our knowledge, the current state-of-the-art DL weather prediction model is the DLWP model of Weyn et al. (2020). The authors work with a coarser resolution of 2° and forecast variables relating to geopotential heights, geopotential thickness, and 2-m temperature (see (Weyn et al., 2020) for further details). The FourCastNet model predicts more variables than the DLWP model at a resolution that is higher than the DLWP model by a factor of 8, allowing us to resolve important phenomena such as hurricanes, extreme precipitation and atmospheric rivers. We downsample the FourCastNet outputs eight times in each direction using bilinear interpolation to bring them to a resolution comparable to that of the DLWP model. Coarsening our forecasts and making them less effective in order to accommodate a prior benchmark at a lower resolution is not fair to our model but we provide this comparison for completeness. We re-compute our ACC and RMSE metrics for the two variables reported in the DLWP results, Z_{500} and T_{2m}. We also note that the ACC metric in the DLWP work was computed using daily climatology so we modify our ACC computation using the same definition for a fair comparison. We observe (Figure 5) that even at the lower resolution of the DLWP work, the FourCastNet model predictions show significant improvement over the current state-of-the-art DLWP model in both variables.