MODELLING ATMOSPHERIC DYNAMICS WITH SPHERICAL FOURIER NEURAL OPERATORS

Boris Bonev *
NVIDIA

Thorsten Kurth *
NVIDIA

Christian Hundt *
NVIDIA

Jaideep Pathak *
NVIDIA

Maximilian Baust *
NVIDIA

Karthik Kashinath *
NVIDIA

Anima Anandkumar †
Caltech & NVIDIA

ABSTRACT

Fourier Neural Operators (FNOs) have established themselves as an efficient method for learning resolution-independent operators in a wide range of scientific machine learning applications. This can be attributed to their ability to effectively model long-range dependencies in spatio-temporal data through computationally efficient global convolutions. However, the use of discrete Fourier transforms (DFTs) in FNOs leads to spurious artifacts and pronounced dissipation when applied to spherical coordinates, due to the incorrect assumption of flat geometry. To address the issue, we introduce Spherical FNOs (SFNOs), which use the generalized Fourier transform for learning operators on spherical geometries. We demonstrate the effectiveness of the method for forecasting atmospheric dynamics, producing stable auto-regressive results for a simulated time of one year (1,460 steps) while retaining physically plausible dynamics. This development has significant implications for machine learning-based climate dynamics emulation, which could play a crucial role in accelerating our response to climate change.

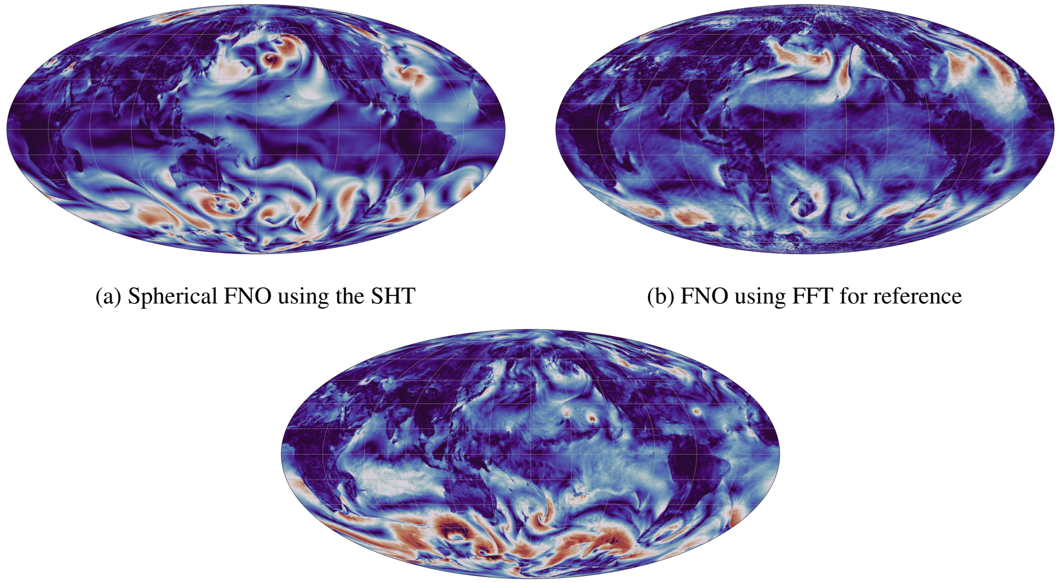
1 INTRODUCTION

Climate change is arguably the greatest challenges facing humanity today. Modeling Earth's complex weather and climate accurately, in a computationally efficient manner, has wide-ranging implications for science and society across the enterprise of climate prediction, mitigation, and adaptation. Weather and climate modeling has traditionally relied on principled physics- and process-based numerical simulations that solve the partial differential equations (PDEs) governing the fluid dynamics, thermodynamics, and other physics of the Earth system. These equations are discretised and solved on a grid, but the wide range of spatial and temporal scales, as well as complex nonlinear interactions across these scales, necessitate fine grids and high resolution making these computations extremely expensive.

Machine learning (ML) provides alternative approaches to modeling weather and climate, and more generally spatio-temporal dynamics, by describing the time evolution of the system as a learned transition map between states of the time-discretized physical system exclusively from raw data. While this enables a unified treatment of the full system, the physics is deduced from data alone without imposing the strong inductive bias of aforementioned physics-based models. Hence, purely data-driven ML-based methods have struggled to faithfully represent the dynamics of physical systems, especially those with long-range correlations in space and time.

Fourier Neural Operators Li et al. (2020) and their variants Guibas et al. (2021); Wen et al. (2022) possess the advantage of learning mappings between function spaces, which act globally on the entire domain. This is in contrast to other architectures, which employ local operations such as convolutions and use hierarchies to model non-local interactions Falk et al. (2019); McCormick (1987). A drawback of FNOs is that Discrete Fourier Transforms (DFTs) assume periodic boundary

*{bbonev, tkurth, chundt, jpathak, mbaust, kkashinath}@nvidia.com
†anima@caltech.edu



(a) Spherical FNO using the SHT

(b) FNO using FFT for reference

(c) Ground Truth

Figure 1: Year-long rollout (1,460 autoregressive steps) of absolute wind speed 10m above the surface depicting stable behavior over exceptionally long timescales for an ML model, which has important implications for ML-based climate modeling. In contrast, the 2D FFT-based architecture shows spurious waves, artifacts and excessive diffusion. SFNO captures weather patterns and their dynamics and statistics well beyond the two-week predictability horizon of the atmosphere. cf. attached video.

conditions leading to an incorrect identification of the north and south pole as well as incorrect longitudinal periodicity on the two-dimensional sphere $\mathbb{S}^2$.

Our approach: We extend the FNO approach to respect spherical geometry and its associated symmetries. To do so, we project functions onto the spherical harmonics, which generalizes the Fourier transform Driscoll & Healy (1994) on $\mathbb{S}^2$. This approach has the advantage that the basis functions fundamentally satisfy equivariance properties Driscoll & Healy (1994); Cohen et al. (2018) with respect to rotations. In particular, rotating the input to such an operator commutes with the operator itself, which is a strong inductive bias to the learned operator.

The proposed method is applied to the Earth Reanalysis 5 dataset (ERA5) Hersbach et al. (2020), and to the Shallow Water Equations (SWE) on the rotating sphere. Our method leads to greatly increased long-term stability, with autoregressive inference remaining stable for over one year (1,460 steps) as opposed to 25 days (100 steps) with a comparable, FFT-based method (cf. attached video, Figure 1). As each autoregressive step takes around 200ms on a single NVIDIA A6000 GPU, these developments open the door to long-range ensemble inference and uncertainty quantification, well beyond weather timescales to subseasonal-to-seasonal (S2S) prediction and potentially, towards climate prediction.

2 Spherical Fourier Neural Operators

Generalized Fourier transform: Our construction of equivariant mappings between function spaces on the sphere is an extension of the FNO framework. FNOs learn resolution-independent representations using global convolution kernels κ

$$\mathcal{K}\big[u\big](x) = \int_{\mathcal{M}} \kappa(x-y) \cdot u(y)\,\mathrm{d}y = \mathcal{F}^{-1}\big[\mathcal{F}[\kappa] \cdot \mathcal{F}[u]\big](x), \tag{1}$$

which can be expressed in terms of the continuous Fourier transform $\mathcal{F}$. When sampling a finite domain on a uniform grid, $\mathcal{F}$ can be expressed as the DFT, allowing FNOs to express long-range dependencies using efficient global convolutions via the FFT.

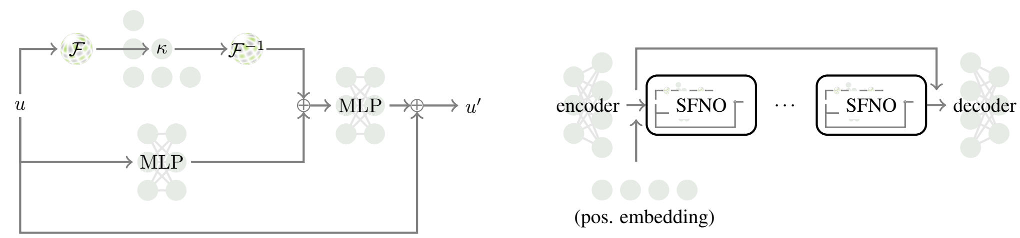
The Spherical FNO (SFNO) layer extends this approach to adapt the Fourier transform while respecting the symmetry of the sphere. In flat geometry, the Fourier transformation is a change of basis by projecting a function $u \in L^2(\mathbb{R}^n)$ onto planar waves $b_k(x) = \exp\big(i\langle k, x\rangle\big)$ effectively encoding translation equivariance. In the spherical setting, the set of basis functions can be obtained from the eigenfunctions of the Laplace-Beltrami operator, which are the spherical harmonics defined as

$$Y_l^m(\theta, \varphi) := (-1)^m c_l^m P_l^m(\cos\theta) e^{\mathrm{i}m\varphi}. \tag{2}$$

Here P_l^m and c_l^m denote the associated Legendre polynomials and normalization factors[1]. Among all the possible bases of $L^2(\mathbb{S}^2)$, the spherical harmonics uniquely exploit the symmetries of the sphere Driscoll & Healy (1994). The Fourier transform on the sphere is then obtained as the decomposition of $L^2(\mathbb{S}^2)$ into linear combinations of the spherical harmonics:

$$u(\theta, \phi) = \sum_{l \in \mathbb{N}} \sum_{|m| \le l} \hat{u}(l, m) Y_l^m(\theta, \phi) \quad \text{with} \quad \hat{u}(l, m) = \int_{\mathbb{S}^2} \overline{Y_l^m} \cdot u \, \mathrm{d}\Omega \,, \tag{3}$$

where $\mathrm{d}\Omega = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$ is the volume form of the sphere. The map $\mathcal{F} : u \to \hat{u}$, which maps $u \in L^2(\mathbb{S}^2)$ to the harmonic coefficients $\hat{u}(l, m)$ is called the spherical harmonics transform (SHT) or generalized Fourier transform Driscoll & Healy (1994).



(a) SFNO block

(b) overall SFNO architecture

Figure 2: (a) The structure of a single SFNO block. (b) Overall SFNO architecure.

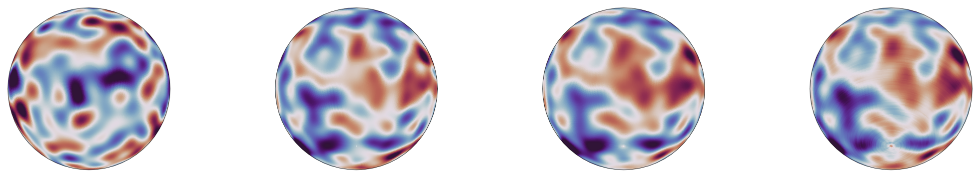
SFNO Network Topology Figure 2a depicts the layout of a single SFNO block. At the core lies the Fourier layer $\mathcal{F}^{-1} \circ \kappa \circ \mathcal{F}$, which uses the SHT $\mathcal{F}$. As with FNOs, κ is a complex-valued, learnable filter, with complex-valued weight matrices for each frequency. The resulting Fourier layer can be regarded as a global convolution on the sphere Li et al. (2020). Alternatively, we propose the use of a complex-valued neural networks which acts "frequency-wise" in the Fourier domain. We use the SFNO block to perform up- and down-scaling, which reduces the memory-footprint of our models. This is done by truncating the frequencies in the forward transform $\mathcal{F}$ and evaluating the inverse $\mathcal{F}^{-1}$ at a higher resolution when up-scaling. The remaining components of the network are also chosen with symmetries in mind. We use MLPs which act point-wise on u, thus respecting rotational equivariance in the continuous setting.

Figure 2b depicts the structure of the entire network, which contains an encoder network, multiple spherical FNO blocks, and a decoder network. To maintain equivariance properties, the encoder and decoder networks are point-wise MLPs with a single hidden layer and GELU activations Hendrycks & Gimpel (2016). These inflate and deflate the channel dimension to the embedding dimension. As F is close to identity, we add a skip connection around the SFNO blocks. Finally, a position embedding is added to learn spatial dependencies.

3 Numerical Experiments

Spherical Shallow Water Equations: The SWE on the rotating sphere (see e.g. Giraldo (2001); Nair et al. (2005); Bonev et al. (2018)) are a system of non-linear hyperbolic PDEs well-suited to model planetary fluid phenomena such as atmospheric dynamics, tsunami propagation, and tidal flows Nair et al. (2005); Bonev et al. (2018).

[1]For a detailed introduction of the spherical harmonics see Abramowitz et al. (1964).



(a) initial condition (b) ground truth, $t = 5h$ (c) SFNO, $t = 5h$ (d) FNO, $t = 10h$

Figure 3: Solutions to the Shallow Water Equations on the rotating Sphere predicted by the SFNO, the FNO and a classical, spectral solver. From left to right: the initial condition, ground truth data computed with the spectral solver, predictions generated with SFNO and FNO respectively.

Table 1: Auto-regressive results applying the Shallow Water equations on the rotating sphere at a spatial resolution of 256×512 and a temporal resolution of 1 hour. The validation loss is reported at 1 and 10 autoregressive steps. For reference, we use the classical solver to compute results at the same resolution. The error is estimated by comparing to a high-fidelity solution at triple the resolution.

METHOD	EMB. DIM.	L^2 LOSS 1H	L^2 LOSS 10H	EVAL TIME [S]	PARAMETERS
UNET	-	$2.961 \cdot 10^{-3}$	$1.462 \cdot 10^{-1}$	0.011	$3.104 \cdot 10^{7}$
FNO LINEAR	256	$8.280 \cdot 10^{-4}$	$9.958 \cdot 10^{-3}$	0.156	$4.998 \cdot 10^{7}$
FNO NON-LIN.	256	$8.298 \cdot 10^{-4}$	$9.784 \cdot 10^{-3}$	0.212	$3.920 \cdot 10^{7}$
SFNO LINEAR	256	$7.741 \cdot 10^{-4}$	$\mathbf{7.239 \cdot 10^{-3}}$	0.218	$3.518 \cdot 10^{7}$
SFNO NON-LIN.	256	$\mathbf{7.673 \cdot 10^{-4}}$	$1.558 \cdot 10^{-2}$	0.321	$3.920 \cdot 10^{7}$
CLASSICAL SOLVER	-	$1.891 \cdot 10^{-2}$	$3.570 \cdot 10^{-2}$	1.299	-

We train our models on the SWE by generating random data on the fly using a classical, spectral solver Giraldo (2001). The models[2] are trained for 20 epochs using a single autoregressive step, where each timestep corresponds to a single hour in the system. Results are shown for U-Net, FNO and SFNO architectures in Table 1. For reference, we estimate the error of the classical solver by computing solutions at triple the resolution to assess the error in the training data.

Weather Prediction / Atmospheric Dynamics

We demonstrate the utility of the proposed method for the task of medium-range weather forecasting (up to two weeks) and long-timescale rollouts (up to 1 year). To do so, we train our model[3] on the ERA5 dataset Hersbach et al. (2020) on a subset of atmospheric variables sub-sampled at a temporal frequency of 6 hours and at the native spatial resolution of the ERA5 dataset (0.25 degrees lat-long). Models are trained following a protocol similar to that outlined in Pathak et al. (2022): an initial training stage, using a single autoregressive step and a second, fine-tuning stage, in which two to four autoregressive steps are used.

4 DISCUSSION

The proposed method pushes the frontier of data-driven deep learning for weather and climate prediction because of the following key properties:

Respecting spherical geometry is essential to ensure that topological boundary conditions are realized correctly and therefore leads to stable autoregressive predictions, as depicted in Figure 1 and Figure 3. On the ERA5 dataset, the SFNO achieves predictive skill comparable to IFS on weather timescales (up to two weeks), while demonstrating unprecedented long-term stability of over a year (**cf.** video of long rollout in supplementary material). This is an essential property for enabling the creation of ML-based digital twins.

[2]models for the SWE have 4 (S)FNO blocks, an embedding dimension of 256 and a down-scaling factor of 3

[3]models for ERA5 have 12 (S)FNO blocks, an embedding dimension of 256 and a down-scaling factor of 6

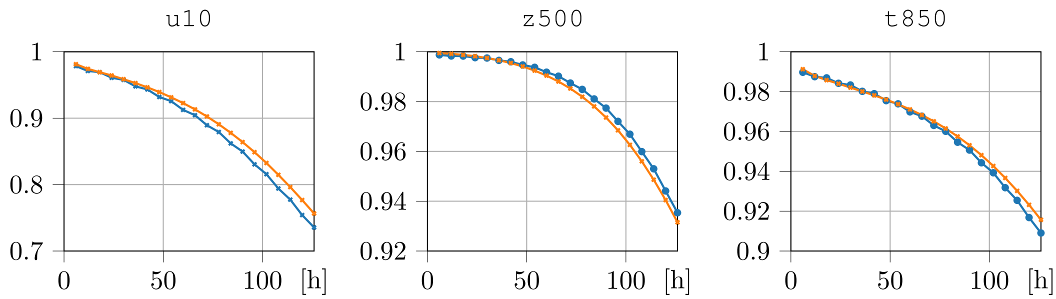


Figure 4: Comparison of forecast skill (ACC, higher is better) as a function of prediction lead time (in hours) between IFS (blue) and the SFNO (orange): u10 (surface winds in the latitudinal direction); z500 (geopotential height at 500 hPa) and t850 (temperature at pressure level 850 hPa). The shaded area signifies the standard deviation over the entire sample for which ACC scores were measured.

Computational efficiency More importantly, one year-long rollouts of the SFNO are computed in 12.8 minutes on a single NVIDIA A6000 GPU, compared to one hour (wall-clock time) for a year-long simulation of IFS on 1000 dual-socket CPU nodes Bauer et al. (2020). With the caveat of differing hardware, this corresponds to a speedup of close to 5000x.

The high accuracy, long-term stability, and immense speedup over classical methods bear a vast promise for the application of Spherical Fourier Neural Operators in the holy grail of weather and climate prediction: sub-seasonal-to-seasonal forecasting using large ensembles. It is foreseeable that such methods could one day lead to ML-based climate prediction.

Acknowledgements

We would like to thank our colleagues Thomas Müller, Kamyar Azizzadenesheli, Nikola Kovachki, Jean Kossaifi and Mike Pritchard for helpful discussions and proof-reading the manuscript.

References

Milton Abramowitz, Irene A Stegun, et al. *Handbook of mathematical functions*, volume 55. Dover New York, 1964.

Peter Bauer, Tiago Quintino, Nils Wedi, Antonino Bonanni, Marcin Chrust, Willem Deconinck, Michail Diamantakis, Peter Düben, Stephen English, Johannes Flemming, Paddy Gillies, Ioan Hadade, James Hawkes, Mike Hawkins, Olivier Iffrig, Christian Kühnlein, Michael Lange, Peter Lean, Olivier Marsden, Andreas Müller, Sami Saarinen, Domokos Sarmany, Michael Sleigh, Simon Smart, Piotr Smolarkiewicz, Daniel Thiemert, Giovanni Tumolo, Christian Weihrauch, Cristiano Zanna, and Pedro Maciel. The ecmwf scalability programme: Progress and plans, 02/2020 2020. URL `https://www.ecmwf.int/node/19380`.

Boris Bonev, Jan S. Hesthaven, Francis X. Giraldo, and Michal A. Kopera. Discontinuous galerkin scheme for the spherical shallow water equations with applications to tsunami modeling and prediction. *Journal of Computational Physics*, 2018. ISSN 10902716. doi: 10.1016/j.jcp.2018.02.008.

Taco S. Cohen, Mario Geiger, Jonas Koehler, and Max Welling. Spherical cnns. 1 2018. URL `http://arxiv.org/abs/1801.10130`.

J.R. Driscoll and D.M. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15:202–250, 6 1994. ISSN 01968858. doi: 10.1006/aama.1994.1008. URL `https://linkinghub.elsevier.com/retrieve/pii/S0196885884710086`.

Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1):67–70, 2019.

Francis X Giraldo. A spectral element shallow water model on spherical geodesic grids, 2001.

John Guibas, Morteza Mardani, Zongyi Li, Andrew Tao, Anima Anandkumar, and Bryan Catanzaro. Adaptive fourier neural operators: Efficient token mixers for transformers, 2021. URL `https://arxiv.org/abs/2111.13587`.

Dan Hendrycks and Kevin Gimpel. Gaussian Error Linear Units (GELUs). 2016.

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.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2020.

Stephen F McCormick. *Multigrid methods*. SIAM, 1987.

Ramachandran D. Nair, Stephen J. Thomas, and Richard D. Loft. A discontinuous galerkin global shallow water model. *Monthly Weather Review*, 133:876–888, 4 2005. ISSN 0027-0644. doi: 10.1175/MWR2903.1. URL `https://journals.ametsoc.org/doi/10.1175/MWR2903.1`.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. 2 2022. URL `http://arxiv.org/abs/2202.11214`.

Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.