AXISYMMETRY-INFORMED NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Spherical neural operators have advanced operator learning on the sphere by exploiting group symmetries, with equivariance and invariance providing powerful inductive biases. However, such transformation laws are limited in characterizing the diverse physical responses. In this work, we explore a complementary principle: directly constructing neural operators from known physical response structures. We instantiate this principle through axisymmetry, a common physical structure in which system responses are organized around a distinguished axis. Starting from an axisymmetric Green's function, we derive its corresponding operator solution, which admits a structured spectral convolution characterized by a preferred axis and Legendre polynomials while retaining grid-invariance. This leads to the proposed Axisymmetry-informed Neural Operator (AsiNO). Across spherical physical dynamics, geophysical prediction and cortical modeling, AsiNO achieves superior operator learning, cross-scale generalization and spherical modeling with substantially improved parameter efficiency. These results demonstrate that physical response structure, beyond group symmetry, provides an effective principle for constructing neural operators on the sphere.

1 INTRODUCTION

Neural operators have emerged as a powerful paradigm for learning mappings between function spaces, offering a data-driven alternative to solving complex physical systems (Kovachki et al., 2024). This perspective has motivated a rapidly growing family of operator-learning methods, including DeepONet, Fourier Neural Operator (FNO), and their subsequent extensions to increasingly complex physical domains (Lu et al., 2019; 2021; Bhattacharya et al., 2021; Li et al., 2020a; Azizzadenesheli et al., 2024). A central challenge underlying these developments is how to incorporate physical priors into the learned operator by exploiting spectral structure, geometry, and symmetry, thereby reducing the burden of learning an otherwise unconstrained mapping purely from data (Kovachki et al., 2023; 2024).

This challenge becomes particularly important for physical systems defined on the sphere, which arise naturally in geophysical dynamics, climate modeling and biomedicine. Spherical neural operators have been developed to respect spherical geometry and widely applied in geophysics (Lin et al., 2023; Mahesh et al., 2024a;b; Liu et al., 2024; Hu et al., 2025). Especially, SFNO constructs rotation-equivariant operators through spherical harmonics (Bonev et al., 2023), and GSNO further combines equivariant and invariant responses to model systems beyond strict equivariance (Tang et al., 2026). These formulations primarily incorporate physical priors through *global group symmetries*: equivariant and invariant transformation behaviours.

Despite their effectiveness, global group symmetries alone cannot fully characterize diverse spherical physics. While equivariance and invariance specify *how a system transforms*, they do not fully determine *how the system responds*. Physical systems sharing the same transformation laws may exhibit distinct response structures, which can be explicitly characterized by their governing physics. Relying solely on group symmetry therefore leaves such physical priors implicit, potentially introducing unnecessary degrees of freedom and limiting operator generalization and forecasting stability (Kovachki et al., 2023; Kurz et al., 2024; Gu et al., 2025; Tang et al., 2026). This limitation motivates the exploration of alternative operator constructions that directly exploit known physical response structures beyond global group symmetries.

To this end, we explore a complementary principle through *axisymmetry*, a common physical response structure in which interactions are organized around a distinguished axis. Starting from the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Green's function representation of an axisymmetric system, we show that an axis-conditioned response satisfies $G^A(u, R) = G^A(u \cdot (R\mathbf{d}))$, where $\mathbf{d}$ denotes the preferred response axis. This yields an *Axisymmetric Green's response*, from which we derive an operator solution admitting a structured spectral convolution characterized by a preferred axis and Legendre polynomials. Based on this solution, we develop the **Axisymmetry-informed Neural Operator (AsiNO)**, a novel spherical neural operator model. Across spherical physical dynamics, geophysical prediction, and cortical surface modelling, AsiNO achieves strong predictive performance and cross-scale generalization with substantially fewer parameters than existing methods. These results demonstrate that a simple, physically motivated response structure can provide an effective and parameter-efficient foundation for spherical operator construction. Our main contributions are summarized as follows:

1. We introduce a complementary principle for neural operator design that directly exploits known physical response structures beyond group symmetries.
2. We derive an axisymmetric operator solution from an axis-conditioned operator system, yielding a structured spectral convolution characterized by a preferred axis and Legendre polynomials.
3. We design Axisymmetry-informed Neural Operator (AsiNO) based on the derived axisymmetric operator solution, enabling efficient operator learning on the sphere while retaining grid-invariance.
4. Across spherical physical dynamics, geophysical prediction and cortical modelling, AsiNO demonstrates strong performance, cross-scale generalization, and parameter efficiency.

2 Related Work

Neural operators have been developed with diverse representations and architectures for learning mappings between function spaces (Kovachki et al., 2024), including Branch–trunk-based (Lu et al., 2021; He et al., 2024; 2023), Fourier-based (Li et al., 2020a; Tran et al., 2021; Liu et al., 2023; Cao et al., 2024), graph-based (Li et al., 2020b;c; 2025), and Attention-based (Pathak et al., 2022; Hao et al., 2023; Wu et al., 2024) methods. However, these methods neglect intrinsic spherical geometry. To address this, spherical neural operators, such as SFNO (Bonev et al., 2023) and GSNO (Tang et al., 2026), exploit group symmetries to achieve effective operator learning on spherical domains.

More broadly, group symmetry has become an important source of inductive bias for learning on structured domains. Translation, scaling, dilation, and other transformation structures have been incorporated into neural architectures (Xu et al., 2014; Sosnovik et al., 2021; Rahman & Yeh, 2023; Chen et al., 2023), forming an important foundation of geometric deep learning (Cohen & Welling, 2016; Bronstein et al., 2021). In particular, group-equivariant networks (Cohen & Welling, 2016; Cohen et al., 2018) and scalable spherical convolutions such as DISCO (Ocampo et al., 2022) explicitly exploit rotational symmetries, while related studies investigate approximate, relaxed, or more flexible equivariance for modelling physical systems (Finzi et al., 2021; Huang et al., 2022; Wang et al., 2022; Duval et al., 2023; Pertigkiozoglou et al., 2024; Zheng et al., 2024; Li et al., 2024). Together, these developments establish transformation properties, particularly equivariance and invariance, as effective inductive biases for spherical learning and operator construction.

3 Preliminaries

3.1 Spherical Harmonics and Spectral Representations

Let $\mathbb{S}^2 = \{u \in \mathbb{R}^3 : \|u\|_2 = 1\}$ denote the unit sphere. Spherical harmonics provide an orthonormal basis for square-integrable functions on $\mathbb{S}^2$ (Müller, 2006). For degree $\ell \geq 0$ and order $-\ell \leq m \leq \ell$, we denote the spherical harmonic basis by $Y_\ell^m : \mathbb{S}^2 \to \mathbb{C}$, satisfying:

$$\int_{\mathbb{S}^2} Y_\ell^m(u)\overline{Y_{\ell'}^{m'}(u)}\, du = \delta_{\ell\ell'}\delta_{mm'}. \tag{1}$$

Accordingly, a spherical signal $f \in L^2(\mathbb{S}^2)$ admits the harmonic expansion (Seeley, 1966):

$$f(u) = \mathcal{F}_{\mathbb{S}^2}^{-1}[\widehat{f}_\ell^m] = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} \widehat{f}_\ell^m Y_\ell^m(u), \quad \widehat{f}_\ell^m = \mathcal{F}_{\mathbb{S}^2}[f] = \int_{\mathbb{S}^2} f(u)\overline{Y_\ell^m(u)}\, du \tag{2}$$

where $\mathcal{F}_{\mathbb{S}^2}$ denotes the spherical harmonic transform, and the inverse spherical harmonic transform ($\mathcal{F}_{\mathbb{S}^2}^{-1}$) reconstructs the spatial signal from $\{\widehat{f}_\ell^m\}$. Under a rotation $R \in SO(3)$, harmonic coefficients of the same degree transform among themselves through the Wigner D-matrix (Seeley, 1966; Driscoll & Healy, 1994):

$$Y_\ell^m(R^{-1}u) = \sum_{a=-\ell}^{\ell} D_{am}^{(\ell)}(R) Y_\ell^a(u). \tag{3}$$

The orthogonality of Wigner matrices under the normalized Haar measure on $SO(3)$ gives (Driscoll & Healy, 1994):

$$\int_{SO(3)} D_{am}^{(\ell)}(R)\overline{D_{bn}^{(\ell')}(R)}\, dR = \frac{1}{2\ell+1}\delta_{\ell\ell'}\delta_{ab}\delta_{mn}. \tag{4}$$

Besides, the spherical harmonic addition theorem connects harmonic bases with Legendre polynomials (Müller, 2006):

$$P_\ell(u \cdot v) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_\ell^m(u)\overline{Y_\ell^m(v)}, \tag{5}$$

which provides the spectral representation of a response whose value depends only on the angle between a spherical location and a preferred physical axis.

3.2 Spherical Neural Operators

Spherical neural operators aim to learn mappings between functions defined on $\mathbb{S}^2$. We consider a spherical operator system $\mathcal{O}$ (Tang et al., 2026):

$$\mathcal{O}[g](u) = f(u), \qquad u \in \mathbb{S}^2, \tag{6}$$

where f is the input source, and g is the corresponding response. Given a reference $n \in \mathbb{S}^2$, the solution is represented through the system's Green's response as:

$$g(u) = \int_{SO(3)} G(u, R) f(Rn)\, dR, \tag{7}$$

where $G(u, R)$ characterizes the response at u induced by an input source at Rn, and dR denotes the normalized Haar measure on $SO(3)$. The Green's response determines the structure of the resulting operator solution, providing a mathematical basis for incorporating physical priors into spherical operator learning. Especially, SFNO (Bonev et al., 2023) employs rotationally equivariant spectral convolution, corresponding to a Green's response of the form $G^E(u, R) = G^E(R^{-1}u)$. GSNO (Tang et al., 2026) further incorporates an invariant Green's response $G^I(u, R) = G^I(u)$ alongside the equivariant response, allowing the learned operator to capture both equivariant and invariant transformation behaviours. These formulations establish a direct connection between the structure of a Green's response and its corresponding spherical operator solution.

4 Methodology

Building on the Green's response formulation in Section 3.2, we investigate how known physical response structures can directly inform spherical operator construction. We instantiate this principle through axisymmetry, deriving a structured operator solution from an axisymmetric Green's function and translating it into the proposed Axisymmetry-informed Neural Operator (AsiNO). Figure 1 illustrates the theoretical framework and the resulting neural architecture.

4.1 Physical-Response-Informed Operator Construction

The Green's function formulation in Eq. 7 establishes a direct connection between the physical response of a system and its corresponding operator solution. While existing spherical neural operators primarily construct their responses through global transformation properties, known physical structures can also be imposed directly on the Green's function to determine the resulting operator formulation.

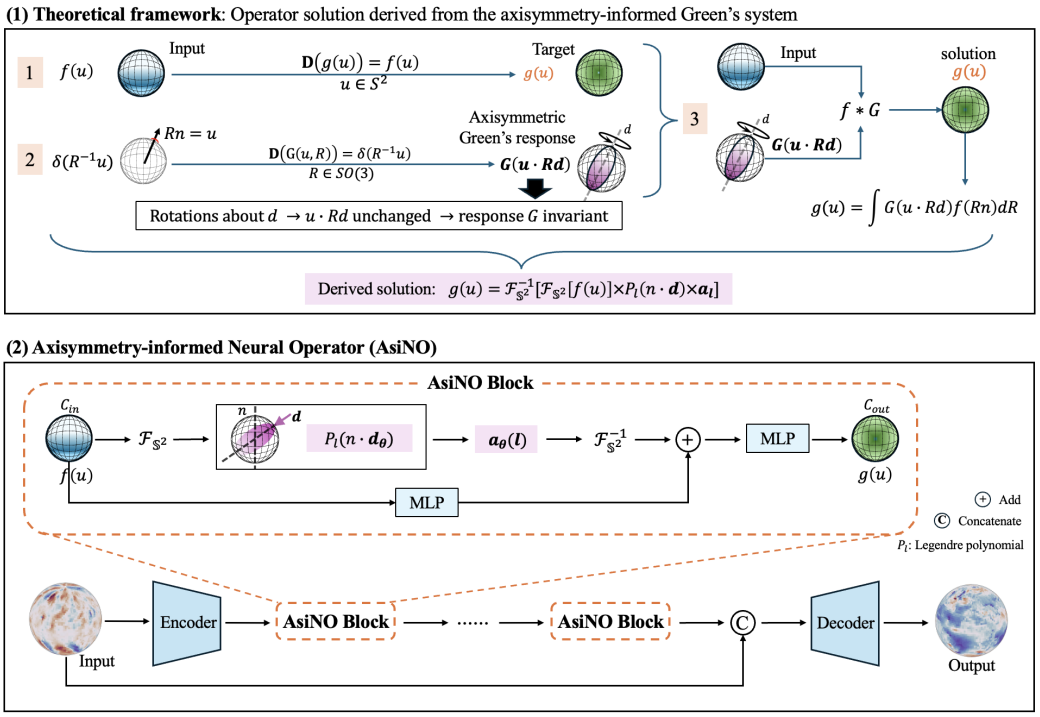

Figure 1: **Overview of AsiNO. (1) Theoretical framework.** Starting from a spherical operator system, axisymmetric Green's response is characterized by a preferred axis $\mathbf{d}$. Rotations about $\mathbf{d}$ leave the response unchanged, yielding a structured operator solution characterized by Legendre polynomials. **(2) AsiNO architecture.** The derived operator solution is implemented as AsiNO block. Following SFNO and GSNO, AsiNO adopts the same MLP, residual pathways, and encoder–decoder architecture to enable a controlled comparison of the underlying spectral operators.

Specifically, let $\mathcal{P}$ denote a known physical response prior that constrains the admissible Green's function $G^{\mathcal{P}}(u, R)$. Substituting this response into Eq. 7 yields the corresponding physics-informed operator:

$$(\mathcal{G}_{\mathcal{P}} f)(u) = \int_{SO(3)} G^{\mathcal{P}}(u, R) f(Rn)\, dR. \tag{8}$$

Here, the physical prior determines the structural form of the response, while its remaining parameters can be learned from data. This formulation allows known physical knowledge to directly constrain the operator solution rather than requiring a general spectral parameterization to recover its structure from observations.

In this work, we investigate *axisymmetry*, a common physical response structure characterized by a distinguished axis and rotational invariance around that axis. Such directional organization arises in diverse physical and biomedical systems, including anisotropic transport, cortical geometry, and white-matter fiber architectures. As illustrated in Figure 1(1), we introduce this structural prior into the Green's function and derive its corresponding spherical operator solution.

4.2 Axisymmetric Green's Response and Operator Solution

Axisymmetric Green's response. Let $\mathbf{d} \in \mathbb{S}^2$ denote a preferred physical axis and $R \in SO(3)$ a rotation acting on this axis. We define an axisymmetric Green's response whose value depends only on the relative angle between the spherical location u and the rotated axis $R\mathbf{d}$:

$$G^A(u, R) = G^A(u \cdot R\mathbf{d})\,. \tag{9}$$

This response is axisymmetric because any rotation Q_ϕ about $\mathbf{d}$ satisfies $Q_\phi \mathbf{d} = \mathbf{d}$, and consequently $G^A(u, RQ_\phi) = G^A(u, R)$. Thus, rotations around the preferred axis ($\mathbf{d}$) leave the Green's response unchanged, whereas rotations that alter its orientation may produce different responses.

Since Eq. 9 depends only on the angular relationship between u and $R\mathbf{d}$, it admits a Legendre expansion:

$$G^A(u, R) = \sum_{\ell \geq 0} g_\ell P_\ell(u \cdot R\mathbf{d}), \tag{10}$$

where P_ℓ denotes the Legendre polynomial of degree ℓ, and g_ℓ determines the corresponding spectral response strength. This representation explicitly incorporates the preferred-axis structure into the Green's function.

Axisymmetric operator solution. Substituting Eq. 10 into the Green's response formulation in Eq. 7 and applying the spherical harmonic transform gives

$$\widehat{g}_\ell^m = \int_{SO(3)} f(Rn) \left[\int_{\mathbb{S}^2} G^A(u, R) \overline{Y_\ell^m(u)}\, du \right] dR. \tag{11}$$

Using the spherical harmonic addition theorem in Eq. 5 and harmonic orthogonality, the inner integral reduces to

$$\int_{\mathbb{S}^2} G^A(u, R) \overline{Y_\ell^m(u)}\, du = g_\ell \frac{4\pi}{2\ell + 1} \overline{Y_\ell^m(R\mathbf{d})}. \tag{12}$$

The remaining integral over $SO(3)$ is evaluated using the rotational transformation of spherical harmonics and the orthogonality of Wigner D-matrices in Eqs. 3–4, yielding the axisymmetric operator solution:

$$\widehat{g}_\ell^m = \frac{g_\ell}{2\ell + 1} P_\ell(n \cdot \mathbf{d}) \widehat{f}_\ell^m. \tag{13}$$

The complete derivation is provided in Appendix A. Absorbing the degree-dependent normalization into the learnable spectral response $\alpha_\ell = g_\ell/(2\ell + 1)$ yields the compact Axisymmetric Green's response (GA):

$$g(u) = \mathcal{F}_{\mathbb{S}^2}^{-1}[\alpha_\ell P_\ell(n \cdot \mathbf{d}) \mathcal{F}_{\mathbb{S}^2}[f](\ell, m)]\,. \tag{14}$$

Equation 14 establishes a structured spectral convolution in which the input dependence is represented by the spherical harmonic coefficients, the spectral response strength by α_ℓ, and the preferred-axis structure by $P_\ell(n \cdot \mathbf{d})$. The resulting formulation directly translates the axisymmetric physical response into an operator solution, providing the mathematical foundation for the proposed AsiNO architecture.

4.3 Axisymmetry-informed Neural Operator

Based on the derived GA solution, we develop the Axisymmetry-informed Neural Operator (AsiNO). As illustrated in Figure 1(2), the core innovation lies in replacing the conventional spherical spectral convolution with the derived GA operator. The remaining network components follow the established SFNO and GSNO architectures, enabling a controlled evaluation of the proposed operator formulation.

Multi-channel GA spectral convolution. Let $v \in \mathbb{R}^{B \times C \times V}$ denote a hidden spherical representation with batch size B, channel dimension C, and V spherical samples. To accommodate multiple response orientations, each channel c is associated with a learnable unit axis $\mathbf{d}_c = \widetilde{\mathbf{d}}_c / \|\widetilde{\mathbf{d}}_c\|_2$, where $\widetilde{\mathbf{d}}_c \in \mathbb{R}^3$. Following Eq. 14, the corresponding axis-conditioned spectral response is defined as

$$\kappa_{\ell c}^A = a_\theta(c, \ell) P_\ell(n \cdot \mathbf{d}_\theta(c)), \tag{15}$$

where $a_\theta(c, \ell)$ denotes the learnable spectral coefficient for channel c and harmonic degree ℓ. Each channel therefore learns a distinct axisymmetric response, with its preferred axis shared across all harmonic degrees. The multi-channel GA convolution is implemented through spherical harmonic transforms:

$$z^A = \mathcal{F}_{\mathbb{S}^2}^{-1}\left[\kappa_\theta^A \odot \mathcal{F}_{\mathbb{S}^2}[v]\right], \tag{16}$$

where $\odot$ denotes channel-wise spectral multiplication. The Legendre factor encodes the preferred-axis structure, while the learnable coefficients determine the spectral response strength. This formulation preserves global spherical interactions without independently parameterizing dense channel transformations at every harmonic degree.

Table 1: Autoregressive forecasting performance on SWE. All models are trained and evaluated at 128×256. We report the mean relative error↓ of geopotential height (H), vorticity (V), and divergence (D) and their average at rollout steps 10, 20, and 30.

Method	Params↓ (M)	Speed↓ (s/epoch)	Step	H↓	V↓	D↓	Average↓
ClimaX (Nguyen et al., 2023)	3.786	**30.3**	10	0.00473 ± 0.00214	0.78292 ± 0.51863	1.56325 ± 1.34013	0.78363
			20	0.00914 ± 0.00587	2.00737 ± 1.56553	4.48302 ± 4.75312	2.16651
			30	0.01329 ± 0.00818	3.43658 ± 2.58837	8.80148 ± 8.46161	4.08378
FourCastNet (Pathak et al., 2022)	1.369	38.4	10	0.00528 ± 0.00044	0.74393 ± 0.01800	1.14013 ± 0.02375	0.62978
			20	0.00683 ± 0.00054	1.44105 ± 0.03855	1.77436 ± 0.03879	1.07408
			30	0.00721 ± 0.00053	2.14434 ± 0.05843	2.27595 ± 0.04241	1.47584
SFNO (Bonev et al., 2023)	0.755	33.9	10	0.00957 ± 0.00540	0.46131 ± 0.16823	0.49995 ± 0.13026	0.32361
			20	0.04153 ± 0.02465	1.64813 ± 0.91464	1.96941 ± 1.10233	1.21969
			30	0.15281 ± 0.08968	6.13693 ± 3.39188	8.00383 ± 3.94867	4.76452
GSNO (Tang et al., 2026)	1.206	34.1	10	$\mathbf{0.00186 \pm 0.00059}$	$\mathbf{0.14615 \pm 0.02520}$	0.31560 ± 0.02543	**0.15454**
			20	0.00549 ± 0.00288	0.39878 ± 0.19293	0.57692 ± 0.28003	0.32706
			30	0.02121 ± 0.02009	1.75293 ± 1.67637	2.64954 ± 2.57501	1.47456
AsiNO (Ours)	**0.077**	33.8	10	0.00190 ± 0.00026	0.23982 ± 0.01062	$\mathbf{0.28370 \pm 0.01194}$	0.17514
			20	$\mathbf{0.00228 \pm 0.00031}$	$\mathbf{0.36770 \pm 0.01396}$	$\mathbf{0.35241 \pm 0.01564}$	**0.24080**
			30	$\mathbf{0.00301 \pm 0.00046}$	$\mathbf{0.48549 \pm 0.01755}$	$\mathbf{0.51207 \pm 0.02129}$	**0.33352**

AsiNO block and network architecture. As shown in Figure 1(2), the AsiNO block implements the derived GA spectral convolution within the established spherical neural operator architecture. Following SFNO and GSNO (Bonev et al., 2023; Tang et al., 2026), we retain the same point-wise MLPs, residual pathways, activation functions, and normalization operations, replacing only the underlying spectral operator. Given the hidden representation $v^{(k)}$ at layer k, the block is formulated as: $v^{(k+1)} = \mathrm{MLP}\big(\mathcal{F}_{\mathbb{S}^2}^{-1}\big[\kappa_\theta^A \odot \mathcal{F}_{\mathbb{S}^2}[v^{(k)}]\big] + \mathrm{MLP}(v^{(k)})\big)$. The GA pathway captures global spherical interactions through the structured physical response, while the point-wise pathway provides local nonlinear transformations and channel mixing.

The complete network follows the encoder–operator–decoder architecture of SFNO and GSNO (Bonev et al., 2023; Tang et al., 2026). Specifically, an encoder maps the input spherical field into a latent representation, which is processed by multiple AsiNO blocks. The resulting features are concatenated with the original input and mapped to the output field through a decoder. By keeping the overall network architecture and training configurations consistent with the baselines, we isolate the contribution of the GA spectral operator and enable a fair operator-level comparison. Overall, AsiNO translates a known axisymmetric physical response into a learnable and efficient spherical neural operator through the derived Green's function solution. By preserving the established network architecture and modifying only the spectral operator, the proposed formulation provides a direct assessment of physical-response-informed operator construction beyond group symmetry.

Parameter efficiency. For C latent channels and L retained harmonic degrees, a general dense degree-dependent spectral transformation requires $\mathcal{O}(LC^2)$ learnable parameters. In contrast, the GA spectral convolution uses $\mathcal{O}(LC)$ learnable spectral coefficients and $\mathcal{O}(C)$ axis parameters. Channel mixing and transformation are performed through the shared point-wise MLPs rather than independent dense spectral matrices, yielding a structured and parameter-efficient spectral operator.

5 Experiments and Results

We evaluate AsiNO across spherical physical dynamics, real-world weather forecasting, cross-scale generalization, and cortical surface modelling. Our experiments aim to investigate whether the proposed axisymmetry-informed operator can effectively represent diverse spherical systems while improving prediction accuracy and parameter efficiency. We compare AsiNO with ClimaX (Nguyen et al., 2023), FourCastNet (Pathak et al., 2022), SFNO (Bonev et al., 2023), and GSNO (Tang et al., 2026). To isolate the contribution of the proposed operator formulation, grid-invariant AsiNO, SFNO and GSNO use the same network architecture (without positional embeddings) and experimental settings, ensuring **operator-level cross-scale comparison**. Detailed experimental settings, evaluation metrics, and additional results are provided in Appendix B.

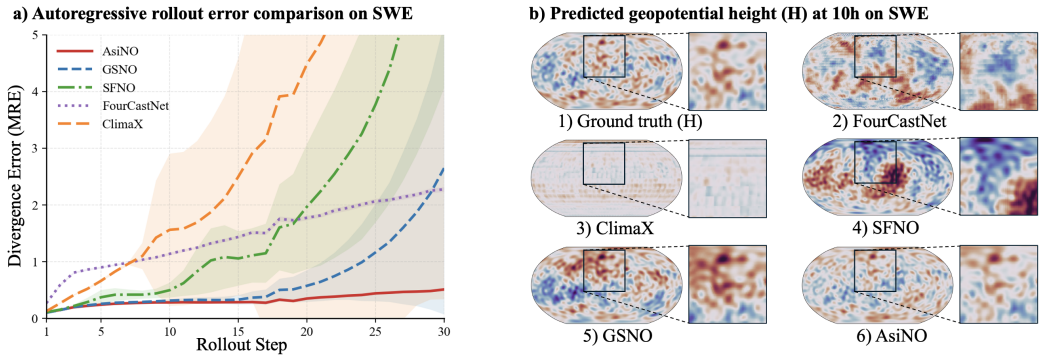

Figure 2: In-distribution comparison on 128×256 SWE: a) Autoregressive rollout error comparison of Divergence; b)Predicted geopotential height (H) at 10h. Zoom in for clear comparison.

Table 2: Zero-shot generalization on SWE (mean relative error↓ at different steps). All models are trained at 128×256 and directly evaluated at 256×512 without retraining or fine-tuning.

Method	Step	H↓	V↓	D↓	Average↓
SFNO (Bonev et al., 2023)	10	0.01482 ± 0.00679	0.89493 ± 0.05594	1.10155 ± 0.16723	0.67043
	20	0.06559 ± 0.03142	1.38511 ± 0.41651	3.23713 ± 0.94024	1.56261
	30	0.22009 ± 0.10635	3.77474 ± 2.27230	12.01847 ± 4.62596	5.33777
GSNO (Tang et al., 2026)	10	0.00209 ± 0.00035	0.81569 ± 0.00608	0.53944 ± 0.01071	0.45241
	20	0.00399 ± 0.00112	0.83986 ± 0.00709	0.55584 ± 0.06240	0.46656
	30	0.00798 ± 0.00429	0.90638 ± 0.05196	1.03639 ± 0.33076	0.65025
AsiNO	10	$\mathbf{0.00207 \pm 0.00031}$	$\mathbf{0.80903 \pm 0.00566}$	$\mathbf{0.53082 \pm 0.00987}$	**0.44733**
	20	$\mathbf{0.00230 \pm 0.00031}$	$\mathbf{0.83787 \pm 0.00584}$	$\mathbf{0.48463 \pm 0.01062}$	**0.44160**
	30	$\mathbf{0.00290 \pm 0.00044}$	$\mathbf{0.85419 \pm 0.00582}$	$\mathbf{0.54547 \pm 0.01309}$	**0.46752**

5.1 Spherical Shallow Water Equations

The spherical shallow water equations (SWE) describe nonlinear fluid dynamics on a rotating sphere and provide a representative benchmark for learning spherical physical evolution. Following the experimental setting of SFNO (Bonev et al., 2023), we generate numerical simulations of the rotating SWE and learn an autoregressive mapping between consecutive physical states. The model predicts three physical variables: geopotential height (H), vorticity (V), and divergence (D). All models are trained and evaluated at a spatial resolution of 128×256, with prediction accuracy assessed over multiple forecasting horizons. We additionally evaluate zero-shot cross-scale transfer to 256×512 without retraining. The simulation protocol, model configurations, and evaluation metrics are detailed in Appendix B.1.

Prediction accuracy and stability. Table 1 show that AsiNO achieves consistently strong predictive performance and stability across the three physical variables and multiple forecasting horizons, demonstrating the effectiveness of axisymmetry-informed spectral learning for nonlinear spherical dynamics. Figure 2 further verifies the accurate long-term prediction of AsiNO. Additional predictions at different spatial and temporal scales are provided in Figure 5 and 6 (Appendix). Beyond individual forecasting horizons, we further investigate long-term autoregressive stability by evaluating prediction errors over 30 consecutive rollout steps. As shown in Appendix 7, AsiNO maintains stable error evolution across the physical variables, indicating that the structured operator formulation can effectively support repeated temporal propagation.

5.2 Weather Forecasting

We further evaluate AsiNO on WeatherBench (WB) (Rasp et al., 2020), a multi-scale benchmark for data-driven global weather forecasting based on atmospheric reanalysis data. Following the experimental protocol of GSNO (Tang et al., 2026), we consider six atmospheric variables as the prediction targets: 10u, 10v, 2T, T600, U600, and V600. All models use the same settings, and are trained and evaluated at 32×64 resolution for 72-hour and 120-hour forecasting. Detailed dataset description

Table 3: Weather forecasting performance comparison of different methods measured by anomaly correlation coefficient (ACC↑, %, higher is better). All models are trained at 32×64, and evaluated at 32×64 and 64×128 WeatherBench.

Model	Params	Horizon	10u	10v	2T	T600	U600	V600	Avg.
In-distribution test (32×64)									
SFNO (Bonev et al., 2023)	3.7M	72h	59.79 ± 4.59	58.05 ± 4.64	91.80 ± 4.75	83.97 ± 7.20	68.34 ± 4.04	60.32 ± 5.10	70.38
		120h	37.27 ± 8.60	30.39 ± 8.60	87.19 ± 7.78	71.22 ± 15.92	45.90 ± 8.89	26.42 ± 9.10	49.73
GSNO (Tang et al., 2026)	4.0M	72h	61.14 ± 4.86	59.16 ± 4.77	91.86 ± 4.57	84.53 ± 7.28	69.82 ± 4.14	61.03 ± 5.19	71.26
		120h	37.29 ± 9.18	30.67 ± 8.87	87.37 ± 7.67	71.39 ± 15.79	46.04 ± 9.30	26.68 ± 9.40	49.91
FourCastNet (Pathak et al., 2022)	5.3M	72h	56.44 ± 4.84	51.99 ± 4.72	88.73 ± 6.58	81.79 ± 8.46	65.74 ± 4.40	53.17 ± 5.45	66.31
		120h	30.77 ± 7.90	22.61 ± 7.86	82.54 ± 10.94	66.90 ± 18.00	38.95 ± 9.01	17.57 ± 7.73	43.22
ClimaX (Nguyen et al., 2023)	5.4M	72h	58.18 ± 5.12	54.03 ± 5.16	89.78 ± 6.36	82.84 ± 8.51	67.61 ± 4.54	58.12 ± 5.79	68.43
		120h	28.57 ± 9.94	22.85 ± 8.17	77.42 ± 14.60	65.13 ± 19.61	34.97 ± 10.94	14.59 ± 9.02	40.59
AsiNO	1.2M	72h	$\mathbf{63.85 \pm 4.20}$	$\mathbf{62.26 \pm 4.43}$	$\mathbf{91.91 \pm 4.61}$	$\mathbf{85.84 \pm 6.61}$	$\mathbf{72.08 \pm 3.60}$	$\mathbf{64.75 \pm 4.86}$	**73.45**
		120h	$\mathbf{41.34 \pm 7.95}$	$\mathbf{34.67 \pm 8.33}$	$\mathbf{87.80 \pm 7.44}$	$\mathbf{73.81 \pm 13.88}$	$\mathbf{50.48 \pm 7.85}$	$\mathbf{32.23 \pm 9.73}$	**53.39**
Cross-scale generalization ($32 \times 64 \rightarrow 64 \times 128$)									
SFNO (Bonev et al., 2023)	3.7M	72h	47.75 ± 5.64	51.09 ± 4.62	88.52 ± 6.49	79.39 ± 9.94	59.55 ± 5.14	51.86 ± 5.60	63.03
		120h	30.01 ± 8.22	26.58 ± 7.93	83.92 ± 9.60	65.75 ± 17.98	40.39 ± 8.85	23.91 ± 8.41	45.09
GSNO (Tang et al., 2026)	4.0M	72h	46.68 ± 5.05	44.78 ± 4.88	87.65 ± 6.84	78.06 ± 9.51	57.54 ± 4.74	50.29 ± 5.56	60.83
		120h	30.43 ± 8.51	24.00 ± 7.64	83.70 ± 9.43	67.43 ± 16.75	40.51 ± 8.83	22.44 ± 8.33	44.75
AsiNO	1.2M	72h	$\mathbf{59.52 \pm 4.55}$	$\mathbf{57.72 \pm 4.50}$	$\mathbf{90.83 \pm 5.25}$	$\mathbf{84.53 \pm 7.40}$	$\mathbf{67.89 \pm 4.08}$	$\mathbf{60.11 \pm 4.95}$	**70.10**
		120h	$\mathbf{40.67 \pm 8.08}$	$\mathbf{33.58 \pm 8.09}$	$\mathbf{87.29 \pm 7.89}$	$\mathbf{73.57 \pm 14.13}$	$\mathbf{49.87 \pm 7.91}$	$\mathbf{31.26 \pm 9.71}$	**52.71**

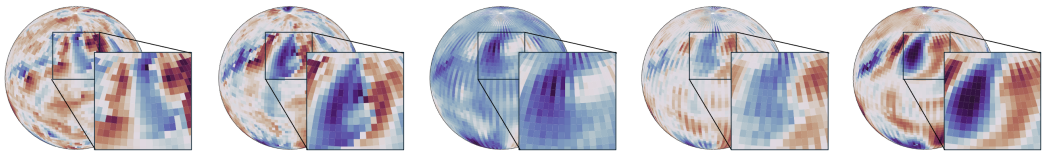

Figure 3: Cross-scale generalization on WeatherBench: 72-hour meridional wind velocity predictions at 10m from the surface (10V) by different methods trained at 32×64 resolution and evaluated at 64×128 resolution. Zoom in for clear comparison.

and training configurations are presented in Appendix B.2. We report the anomaly correlation coefficient (ACC) in Table 3 and mean squared error (MSE) in Table 8 (Appendix). These results show that AsiNO consistently achieves best performance with substantially fewer parameters (1.2M) than the compared methods, demonstrating that the axisymmetry-informed operator provides an effective representation for complex real-world atmospheric dynamics.

5.3 Cross-scale Generalization

An important property of neural operators is their ability to transfer learned mappings across spatial discretizations. We investigate whether the axisymmetry-informed operator preserves this property by evaluating AsiNO at resolutions higher than those used during training. All grid-invariant models are directly evaluated on the target resolution without additional training or fine-tuning.

Cross-scale spherical physical dynamics. Models trained at 128×256 are directly evaluated at 256×512 SWE. Table 2 reports the corresponding zero-shot prediction errors, while Figure 6 in Appendix provides qualitative comparisons at multiple forecasting horizons. We additionally examine the error evolution of H, V, D, and their average over 30 autoregressive steps in Figure 7 (Appendix). These evaluations verify that the learned axisymmetric spectral response remains effective when applied to a finer spatial discretization.

Cross-scale Atmospheric dynamics. We directly transfer models trained at 32×64- to 64×128-resolution WB. The cross-scale ACC results are reported in Table 3, with the corresponding MSE results provided in Table 8 (Appendix). Figure 3 visualizes the 10v predictions at the 72-hour forecasting horizon, comparing the spatial structures of different methods. Additional visualizations across atmospheric variables and forecasting horizons are provided in Figure 8 and 9 (Appendix).

Table 4: Cortical parcellation on NAMIC and Mindboggle. Dice scores, parameters (M) and training speeds (samples/s) of different methods are reported. Bold is the best result.

Dataset	Metric	SPHARM-Net (Ha & Lyu, 2022)	SFNO (Bonev et al., 2023)	GSNO (Tang et al., 2026)	AsiNO
NAMIC	Params / Speed	4.32 / 7.87	6.33 / 3.60	6.57 / 3.45	**0.85 / 9.00**
	Dice (%) ↑	85.26 ± 0.70	85.41 ± 0.79	85.29 ± 0.93	$\mathbf{85.68 \pm 0.81}$
Mindboggle	Params / Speed	4.31 / 7.82	6.33 / 3.55	6.57 / 3.43	**0.85 / 8.75**
	Dice (%) ↑	88.35 ± 0.63	88.29 ± 0.72	88.20 ± 0.71	$\mathbf{88.59 \pm 0.52}$

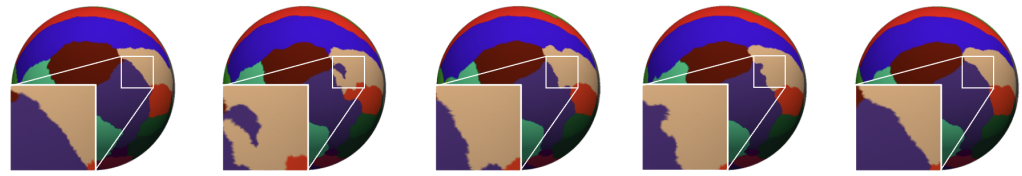

Figure 4: **Cortical surface parcellation in spherical space.** Parcellation results on NAMIC, comparing the predictions of different methods with the ground truth (GT). Zoom in for clarity.

Together, these results demonstrate that the proposed axisymmetry-informed operator supports cross-scale transfer across both simulated physical dynamics and real-world atmospheric forecasting. These results provide empirical evidence that incorporating a structured physical response into spherical operator learning is compatible with generalization across spatial and temporal scales.

5.4 Cortical Surface Modelling

We finally evaluate AsiNO on cortical surface parcellation using the NAMIC (Desikan et al., 2006) and Mindboggle (Klein & Tourville, 2012) datasets. Following the preprocessing and training configurations of SPHARM-Net (Desikan et al., 2006; Ha & Lyu, 2022), we perform cortical parcellation on spherical surface representations. We compare AsiNO with SPHARM-Net, SFNO, and GSNO under identical data partitions and training configurations. For a controlled operator-level comparison, SFNO, GSNO, and AsiNO share the same resnet-type network architecture (Bonev et al., 2023), differing only in their spectral operator formulations. We perform five independent runs with different random seeds and report the mean and standard deviation of Dice scores. Additional experimental details are provided in Appendix B.4.

As shown in Table 4, AsiNO achieves the highest Dice scores on both datasets with significantly higher parameter efficiency and running speed. Figure 4 further presents cortical parcellation in spherical space. Additional visualizations of the reconstructed cortical surfaces in native space are provided in Figure 10 (Appendix). Together, these results demonstrate that axisymmetry-informed operator effectively extends to large-scale biomedical spherical modeling, achieving improved prediction accuracy and efficiency.

6 Conclusion

In this work, we introduce physical-response-informed operator construction as a complementary principle beyond group-symmetry-based spherical neural operators, and then explore axisymmetric structure as a inductive bias for operator construction. By incorporating axisymmetry into the operator system, we derive a structured spectral operator solution characterized by a preferred axis and Legendre polynomials. Based on this solution, we develop the Axisymmetry-informed Neural Operator (AsiNO), a grid-invariant spherical operator model. Experiments across spherical physical dynamics, geophysical prediction, and cortical modeling demonstrate the effective operator learning, long-horizon forecasting stability, cross-scale generalization, efficient cortical modeling and substantially improved parameter efficiency of AsiNO. These findings highlight the potential of directly incorporating known physical response structures into neural operator construction, opening new directions for developing physics-informed and application-specific operators beyond global transformation laws.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and improving the clarity of the manuscript. All core research activities, including research conception, methodological development, experimental design, implementation, and result analysis, were conducted by the authors. All AI-assisted text was carefully reviewed and verified by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Appendix includes the implementation details, experimental configurations, dataset access links, and data preprocessing procedures to facilitate the reproduction of our results. All experiments were conducted using publicly available datasets, with the corresponding data sources and access instructions documented in the appendix.

We also provide the `source code` for the proposed model and its implementation in the **Supplementary Material** to further strengthen the reproducibility.

REFERENCES

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, pp. 1–9, 2024.

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model reduction and neural networks for parametric pdes. *The SMAI journal of computational mathematics*, 7:121–157, 2021.

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*, 362:425–448, 2018.

Boris Bonev, Thorsten Kurth, Christian Hundt, Jaideep Pathak, Maximilian Baust, Karthik Kashinath, and Anima Anandkumar. Spherical fourier neural operators: Learning stable dynamics on the sphere. In *International conference on machine learning*, pp. 2806–2823. PMLR, 2023.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Qianying Cao, Somdatta Goswami, and George Em Karniadakis. Laplace neural operator for solving differential equations. *Nature Machine Intelligence*, 6(6):631–640, 2024.

Gengxiang Chen, Xu Liu, Yingguang Li, Qinglu Meng, and Lu Chen. Laplace neural operator for complex geometries. *arXiv preprint arXiv:2302.08166*, 2023.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *arXiv preprint arXiv:1801.10130*, 2018.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

James R Driscoll and Dennis M Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in applied mathematics*, 15(2):202–250, 1994.

Alexandre Agm Duval, Victor Schmidt, Alex Hernández-Garcıa, Santiago Miret, Fragkiskos D Malliaros, Yoshua Bengio, and David Rolnick. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pp. 9013–9033. PMLR, 2023.

Alan Robert Edmonds. *Angular momentum in quantum mechanics*, volume 4. Princeton university press, Princeton, 1996.

Marc Finzi, Gregory Benton, and Andrew G Wilson. Residual pathway priors for soft equivariance constraints. *Advances in Neural Information Processing Systems*, 34:30037–30049, 2021.

Jianghang Gu, Ling Wen, Yuntian Chen, and Shiyi Chen. An explainable operator approximation framework under the guideline of green's function. *Journal of Computational Physics*, pp. 114244, 2025.

Seungbo Ha and Ilwoo Lyu. Spharm-net: Spherical harmonics-based convolution for cortical parcellation. *IEEE Transactions on Medical Imaging*, 41(10):2739–2751, 2022.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning*, pp. 12556–12569. PMLR, 2023.

Junyan He, Seid Koric, Shashank Kushwaha, Jaewan Park, Diab Abueidda, and Iwona Jasiuk. Novel deeponet architecture to predict stresses in elastoplastic structures with variable complex geometries and loads. *Computer Methods in Applied Mechanics and Engineering*, 415:116277, 2023.

Junyan He, Seid Koric, Diab Abueidda, Ali Najafi, and Iwona Jasiuk. Geom-deeponet: A point-cloud-based deep operator network for field predictions on 3d parameterized geometries. *Computer Methods in Applied Mechanics and Engineering*, 429:117130, 2024.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 146(730):1999–2049, 2020.

Yifan Hu, Fukang Yin, Weimin Zhang, Kaijun Ren, Junqiang Song, Kefeng Deng, and Di Zhang. Spherical multigrid neural operator for improving autoregressive global weather forecasting. *Scientific Reports*, 15(1):11522, 2025.

Wenbing Huang, Jiaqi Han, Yu Rong, Tingyang Xu, Fuchun Sun, and Junzhou Huang. Equivariant graph mechanics networks with constraints. *arXiv preprint arXiv:2203.06442*, 2022.

Arno Klein and Jason Tourville. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in neuroscience*, 6:171, 2012.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Nikola B Kovachki, Samuel Lanthaler, and Andrew M Stuart. Operator learning: Algorithms and analysis. *Handbook of Numerical Analysis*, 25:419–467, 2024.

Jason Kurz, Brett Bowman, M Seman, Chad Oian, and Taufiquar Khan. A physics-informed kernel approach to learning the operator for parametric pdes. *Neural Computing and Applications*, 36 (36):22773–22787, 2024.

Zhihao Li, Haoze Song, Di Xiao, Zhilu Lai, and Wei Wang. Harnessing scale and physics: A multi-graph neural operator framework for pdes on arbitrary geometries. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 729–740, 2025.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020b.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems*, 33:6755–6766, 2020c.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

Kenghong Lin, Xutao Li, Yunming Ye, Shanshan Feng, Baoquan Zhang, Guangning Xu, and Ziye Wang. Spherical neural operator network for global weather prediction. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(6):4899–4913, 2023.

CC Liu, K Hsu, MS Peng, DS Chen, PL Chang, LF Hsiao, et al. Evaluation of five global ai models for predicting weather in eastern asia and western pacific. npj climate and atmospheric science, 7 (1), 221, 2024.

Ning Liu, Siavash Jafarzadeh, and Yue Yu. Domain agnostic fourier neural operators. *Advances in neural information processing systems*, 36:47438–47450, 2023.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Ankur Mahesh, William Collins, Boris Bonev, Noah Brenowitz, Yair Cohen, Joshua Elms, Peter Harrington, Karthik Kashinath, Thorsten Kurth, Joshua North, et al. Huge ensembles part i: Design of ensemble weather forecasts using spherical fourier neural operators. *arXiv preprint arXiv:2408.03100*, 2024a.

Ankur Mahesh, William Collins, Boris Bonev, Noah Brenowitz, Yair Cohen, Peter Harrington, Karthik Kashinath, Thorsten Kurth, Joshua North, Travis OBrien, et al. Huge ensembles part ii: Properties of a huge ensemble of hindcasts generated with spherical fourier neural operators. *arXiv preprint arXiv:2408.01581*, 2024b.

Claus Müller. *Spherical harmonics*, volume 17. Springer, Berlin, 2006.

Ramachandran D Nair, Stephen J Thomas, and Richard D Loft. A discontinuous galerkin global shallow water model. *Monthly weather review*, 133(4):876–888, 2005.

Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*, 2023.

Jeremy Ocampo, Matthew A Price, and Jason D McEwen. Scalable and equivariant spherical cnns by discrete-continuous (disco) convolutions. *arXiv preprint arXiv:2209.13603*, 2022.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Stefanos Pertigkiozoglou, Evangelos Chatzipantazis, Shubhendu Trivedi, and Kostas Daniilidis. Improving equivariant model training via constraint relaxation. *Advances in Neural Information Processing Systems*, 37:83497–83520, 2024.

Md Ashiqur Rahman and Raymond A Yeh. Truly scale-equivariant deep nets with fourier layers. *Advances in Neural Information Processing Systems*, 36:6092–6104, 2023.

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.

Robert T Seeley. Spherical harmonics. *The American Mathematical Monthly*, 73(4P2):115–121, 1966.

Ivan Sosnovik, Artem Moskalev, and Arnold Smeulders. Disco: accurate discrete scale convolutions. *arXiv preprint arXiv:2106.02733*, 2021.

Hao Tang, Hao Chen, and Chao Li. Generalized spherical neural operators: Green's function formulation. In *International Conference on Learning Representations*, volume 2026, pp. 11180–11201, 2026.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. *arXiv preprint arXiv:2111.13802*, 2021.

Rui Wang, Robin Walters, and Rose Yu. Approximately equivariant networks for imperfectly symmetric dynamics. In *International Conference on Machine Learning*, pp. 23078–23091. PMLR, 2022.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. *arXiv preprint arXiv:2402.02366*, 2024.

Yichong Xu, Tianjun Xiao, Jiaxing Zhang, Kuiyuan Yang, and Zheng Zhang. Scale-invariant convolutional neural networks. *arXiv preprint arXiv:1411.6369*, 2014.

Zinan Zheng, Yang Liu, Jia Li, Jianhua Yao, and Yu Rong. Relaxing continuous constraints of equivariant graph neural networks for broad physical dynamics learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4548–4558, 2024.

A THEORETICAL DETAILS AND ANALYSIS OF ASINO

This appendix provides the mathematical details underlying the Axisymmetric Green's response (GA) and the resulting Axisymmetry-informed Neural Operator (AsiNO). Building on the spherical harmonic representations and Green's response formulation introduced in Section 3, we first establish the mathematical properties required for the derivation, then revisit the equivariant spherical response and derive the proposed axisymmetric operator solution. Finally, we analyze the resulting spectral structure and parameter efficiency.

A.1 MATHEMATICAL FOUNDATIONS

Spherical harmonic representation. As introduced in Section 3.1, the spherical harmonics $\{Y_\ell^m\}$ form an orthonormal basis of $L^2(\mathbb{S}^2)$ (Müller, 2006). For $f \in L^2(\mathbb{S}^2)$, its spherical harmonic expansion is

$$f(u) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \widehat{f}_\ell^m Y_\ell^m(u), \qquad \widehat{f}_\ell^m = \mathcal{F}_{\mathbb{S}^2}[f](\ell, m). \tag{17}$$

The spherical harmonic basis satisfies:

$$\int_{\mathbb{S}^2} Y_\ell^m(u)\overline{Y_{\ell'}^{m'}(u)}\, du = \delta_{\ell\ell'}\delta_{mm'}. \tag{18}$$

Here, $\overline{(\cdot)}$ denotes complex conjugation. The inverse spherical harmonic transform reconstructs the spatial signal from its spectral coefficients. In numerical implementations, the expansion is truncated to a finite number of retained harmonic degrees.

Rotations and Wigner matrices. For a rotation $R \in SO(3)$, the action on a spherical function is defined as $(\mathcal{T}_R f)(u) = f(R^{-1}u)$. Spherical harmonics of the same degree transform through the Wigner D-matrix (Edmonds, 1996):

$$Y_\ell^m(R^{-1}u) = \sum_{a=-\ell}^{\ell} D_{am}^{(\ell)}(R) Y_\ell^a(u). \tag{19}$$

Throughout the derivation, dR denotes the normalized Haar measure on $SO(3)$, satisfying $\int_{SO(3)} dR = 1$. The Wigner matrices obey the orthogonality relation (Edmonds, 1996):

$$\int_{SO(3)} D_{am}^{(\ell)}(R)\overline{D_{bn}^{(\ell')}(R)}\, dR = \frac{\delta_{\ell\ell'}\delta_{ab}\delta_{mn}}{2\ell+1}. \tag{20}$$

The Haar measure is invariant under inversion, so $\int_{SO(3)} F(R)\, dR = \int_{SO(3)} F(R^{-1})\, dR$.

Legendre polynomials and axisymmetry. The spherical harmonic addition theorem connects the harmonic basis with Legendre polynomials:

$$P_\ell(u \cdot v) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} Y_\ell^m(u)\overline{Y_\ell^m(v)}. \tag{21}$$

A spherical response $h(u)$ is axisymmetric about a unit axis $\mathbf{d} \in \mathbb{S}^2$ if it depends only on the angular relationship between u and $\mathbf{d}$, namely $h(u) = h(u \cdot \mathbf{d})$. For any rotation Q satisfying $Q\mathbf{d} = \mathbf{d}$, we have

$$h(Qu) = h((Qu)\cdot\mathbf{d}) = h(u \cdot Q^{-1}\mathbf{d}) = h(u\cdot\mathbf{d}) = h(u). \tag{22}$$

Thus, the response is invariant under rotations about its preferred axis. Such rotations form the stabilizer subgroup $SO(2)_{\mathbf{d}} = \{Q \in SO(3) : Q\mathbf{d} = \mathbf{d}\}$.

An axisymmetric response admits a Legendre expansion $h(u\cdot\mathbf{d}) = \sum_{\ell\geq 0} h_\ell P_\ell(u\cdot\mathbf{d})$ (Müller, 2006). This property provides the mathematical basis for translating an axisymmetric physical response into a spherical spectral operator.

A.2 Green's Response and Equivariant Spherical Operators

We first revisit the conventional equivariant spherical operator from the Green's response perspective. Following Section 3.2, let $\mathcal{D}$ denote a linear spherical system with input f and solution g, satisfying $\mathcal{D}g = f$. For a reference direction $n \in \mathbb{S}^2$, the corresponding Green's response formulation is (Tang et al., 2026):

$$g(u) = (\mathcal{G}f)(u) = \int_{SO(3)} G(u, R) f(Rn)\, dR. \tag{23}$$

Here, $G(u, R)$ describes the response at u associated with the source position Rn. Applying the spherical harmonic transform and interchanging the order of integration gives

$$\widehat{g}_\ell^m = \int_{SO(3)} f(Rn) \left[\int_{\mathbb{S}^2} G(u, R) \overline{Y_\ell^m(u)}\, du \right] dR. \tag{24}$$

Different structural assumptions on $G(u, R)$ lead to different spectral representations of the resulting operator.

Equivariant Green's response. For a rotationally equivariant spherical system, consider a Green's response depending only on the relative angular displacement between the observation direction u and the source direction Rn (Bonev et al., 2023):

$$G^E(u, R) = G^E(u \cdot Rn) = \sum_{L \geq 0} g_L^E P_L(u \cdot Rn). \tag{25}$$

Substituting Eq. 25 into Eq. 24 yields

$$\widehat{g}_\ell^m = \int_{SO(3)} f(Rn) \sum_{L \geq 0} g_L^E \left[\int_{\mathbb{S}^2} P_L(u \cdot Rn) \overline{Y_\ell^m(u)}\, du \right] dR. \tag{26}$$

Applying the addition theorem and harmonic orthogonality, the inner integral becomes

$$\begin{aligned} &\int_{\mathbb{S}^2} P_L(u \cdot Rn) \overline{Y_\ell^m(u)}\, du \\ &= \frac{4\pi}{2L+1} \sum_{a=-L}^{L} \overline{Y_L^a(Rn)} \int_{\mathbb{S}^2} Y_L^a(u) \overline{Y_\ell^m(u)}\, du \\ &= \frac{4\pi}{2\ell+1} \delta_{L\ell} \overline{Y_\ell^m(Rn)}. \end{aligned} \tag{27}$$

Consequently,

$$\widehat{g}_\ell^m = g_\ell^E \frac{4\pi}{2\ell+1} \int_{SO(3)} f(Rn) \overline{Y_\ell^m(Rn)}\, dR. \tag{28}$$

Since the pushforward of the normalized Haar measure under $R \mapsto Rn$ is the uniform probability measure on $\mathbb{S}^2$, the remaining integral satisfies

$$\int_{SO(3)} f(Rn) \overline{Y_\ell^m(Rn)}\, dR = \frac{1}{4\pi} \int_{\mathbb{S}^2} f(x) \overline{Y_\ell^m(x)}\, dx = \frac{\widehat{f}_\ell^m}{4\pi}. \tag{29}$$

Substituting Eq. 29 into Eq. 28 gives

$$\boxed{\widehat{g}_\ell^m = \gamma_\ell^E \widehat{f}_\ell^m, \qquad \gamma_\ell^E = \frac{g_\ell^E}{2\ell+1}.} \tag{30}$$

For multi-channel spherical representations, the scalar spectral response can be generalized to a degree-dependent channel transformation:

$$\widehat{\mathbf{g}}_{\ell m} = W_\ell^E \widehat{\mathbf{f}}_{\ell m}, \qquad W_\ell^E \in \mathbb{C}^{C_{\text{out}} \times C_{\text{in}}}. \tag{31}$$

This provides the spectral form of the conventional equivariant spherical operator introduced in Section 3.2. The response is characterized by a degree-dependent spectral transformation shared across harmonic orders.

A.3 Derivation of the Axisymmetric Green's Response

Axisymmetric physical response. We now introduce the physical response constraint underlying AsiNO. Unlike the equivariant response in Eq. 25, which is organized around the source direction Rn, we consider a system whose response is organized around a distinguished physical axis $\mathbf{d} \in \mathbb{S}^2$.

As defined in Section 4.2, the Axisymmetric Green's response is

$$G^A(u, R) = G^A(u \cdot R\mathbf{d}). \tag{32}$$

For any rotation Q_ϕ about $\mathbf{d}$, we have $Q_\phi \mathbf{d} = \mathbf{d}$, and hence

$$G^A(u, RQ_\phi) = G^A(u \cdot RQ_\phi \mathbf{d}) = G^A(u \cdot R\mathbf{d}) = G^A(u, R). \tag{33}$$

Thus, rotating the preferred axis about itself leaves the response unchanged, whereas rotations that alter its orientation may change the response. For each fixed R, the Green's response is also invariant under rotations of the observation direction about $R\mathbf{d}$.

Since the response depends only on $u \cdot R\mathbf{d}$, it admits the Legendre expansion

$$G^A(u, R) = \sum_{L \geq 0} g_L^A P_L(u \cdot R\mathbf{d}), \tag{34}$$

where g_L^A determines the response strength at harmonic degree L.

Spherical harmonic transformation. Substituting Eq. 34 into the Green's response formulation in Eq. 23 gives

$$g(u) = \int_{SO(3)} f(Rn) \left[\sum_{L \geq 0} g_L^A P_L(u \cdot R\mathbf{d}) \right] dR. \tag{35}$$

Applying the spherical harmonic transform and interchanging the order of integration yields

$$\begin{aligned} \widehat{g}_\ell^m &= \int_{\mathbb{S}^2} g(u) \overline{Y_\ell^m(u)}\, du \\ &= \int_{SO(3)} f(Rn) \left[\sum_{L \geq 0} g_L^A \int_{\mathbb{S}^2} P_L(u \cdot R\mathbf{d}) \overline{Y_\ell^m(u)}\, du \right] dR. \end{aligned} \tag{36}$$

We first evaluate the inner spherical integral and then the outer group integral.

Evaluation of the inner integral. Applying the addition theorem in Eq. 21 and the orthogonality relation in Eq. 18, we obtain

$$\begin{aligned} I_{\text{inner}}(R) &= \sum_{L \geq 0} g_L^A \int_{\mathbb{S}^2} P_L(u \cdot R\mathbf{d}) \overline{Y_\ell^m(u)}\, du \\ &= \sum_{L \geq 0} g_L^A \frac{4\pi}{2L+1} \sum_{a=-L}^{L} \overline{Y_L^a(R\mathbf{d})} \int_{\mathbb{S}^2} Y_L^a(u) \overline{Y_\ell^m(u)}\, du \\ &= \sum_{L \geq 0} g_L^A \frac{4\pi}{2L+1} \sum_{a=-L}^{L} \overline{Y_L^a(R\mathbf{d})} \delta_{L\ell} \delta_{am} \\ &= g_\ell^A \frac{4\pi}{2\ell+1} \overline{Y_\ell^m(R\mathbf{d})}. \end{aligned} \tag{37}$$

Substituting this result into Eq. 36 gives

$$\widehat{g}_\ell^m = g_\ell^A \frac{4\pi}{2\ell+1} \int_{SO(3)} f(Rn) \overline{Y_\ell^m(R\mathbf{d})}\, dR. \tag{38}$$

Evaluation of the group integral. Using the inversion invariance of the Haar measure, the remaining integral can be written as

$$I_{\text{outer}} = \int_{SO(3)} f(R^{-1}n)\overline{Y_\ell^m(R^{-1}\mathbf{d})}\, dR. \tag{39}$$

Expanding the input in spherical harmonics and applying the rotational transformation in Eq. 19 gives

$$\begin{aligned} f(R^{-1}n) &= \sum_{L,M} \widehat{f}_L^M Y_L^M(R^{-1}n) \\ &= \sum_{L,M} \widehat{f}_L^M \sum_{a=-L}^{L} D_{aM}^{(L)}(R) Y_L^a(n), \end{aligned} \tag{40}$$

and

$$\overline{Y_\ell^m(R^{-1}\mathbf{d})} = \sum_{b=-\ell}^{\ell} \overline{D_{bm}^{(\ell)}(R)}\,\overline{Y_\ell^b(\mathbf{d})}. \tag{41}$$

Substituting Eqs. 40 and 41 into Eq. 39, we obtain

$$I_{\text{outer}} = \sum_{L,M} \sum_{a=-L}^{L} \sum_{b=-\ell}^{\ell} \widehat{f}_L^M Y_L^a(n)\overline{Y_\ell^b(\mathbf{d})} \int_{SO(3)} D_{aM}^{(L)}(R)\overline{D_{bm}^{(\ell)}(R)}\, dR. \tag{42}$$

Applying Wigner orthogonality in Eq. 20 gives

$$\begin{aligned} I_{\text{outer}} &= \sum_{L,M} \sum_{a=-L}^{L} \sum_{b=-\ell}^{\ell} \widehat{f}_L^M Y_L^a(n)\overline{Y_\ell^b(\mathbf{d})} \frac{\delta_{L\ell}\delta_{ab}\delta_{Mm}}{2\ell+1} \\ &= \frac{\widehat{f}_\ell^m}{2\ell+1} \sum_{a=-\ell}^{\ell} Y_\ell^a(n)\overline{Y_\ell^a(\mathbf{d})}. \end{aligned} \tag{43}$$

Using the addition theorem once more,

$$\sum_{a=-\ell}^{\ell} Y_\ell^a(n)\overline{Y_\ell^a(\mathbf{d})} = \frac{2\ell+1}{4\pi} P_\ell(n\cdot\mathbf{d}), \tag{44}$$

the outer integral reduces to

$$I_{\text{outer}} = \frac{\widehat{f}_\ell^m}{4\pi} P_\ell(n\cdot\mathbf{d}). \tag{45}$$

Final operator solution. Substituting Eq. 45 into Eq. 38 yields

$$\widehat{g}_\ell^m = \frac{g_\ell^A}{2\ell+1} P_\ell(n\cdot\mathbf{d})\widehat{f}_\ell^m. \tag{46}$$

Defining the degree-dependent spectral coefficient $\alpha_\ell = g_\ell^A/(2\ell+1)$, the derived solution can be expressed as

$$g(u) = \mathcal{F}_{\mathbb{S}^2}^{-1}[\alpha_\ell P_\ell(n\cdot\mathbf{d})\mathcal{F}_{\mathbb{S}^2}[f](\ell,m)]\,. \tag{47}$$

Equation 47 recovers the GA spectral convolution introduced in Eq. 14. The physical response prior determines the Legendre-structured spectral factor, while the coefficients α_ℓ determine the corresponding response strength.

A.4 Spectral Properties of the Derived Operator

Response factorization. The derived solution in Eq. 46 separates the spectral response into three components:

$$\widehat{g}_\ell^m = \underbrace{\widehat{f}_\ell^m}_{\text{input representation}} \quad \underbrace{\alpha_\ell}_{\text{spectral response}} \quad \underbrace{P_\ell(n\cdot\mathbf{d})}_{\text{axisymmetric structure}}. \tag{48}$$

The Legendre factor follows directly from the axisymmetric Green's function rather than being introduced as an arbitrary spectral modulation. This provides a physical-response-based construction of the spectral operator.

Relation to equivariant spherical convolution. Comparing Eqs. 30 and 46, the two spectral responses can be written as

$$\underbrace{\widehat{g}_\ell^m = \gamma_\ell^E \widehat{f}_\ell^m}_{\text{GE}}, \qquad \underbrace{\widehat{g}_\ell^m = \alpha_\ell P_\ell(n \cdot \mathbf{d}) \widehat{f}_\ell^m}_{\text{GA}}. \tag{49}$$

Both formulations act diagonally on spherical harmonic degrees and orders. For a fixed reference direction n and preferred axis $\mathbf{d}$, the GA response can be represented by an equivariant spectral multiplier $\gamma_\ell^E = \alpha_\ell P_\ell(n \cdot \mathbf{d})$.

Therefore, its distinction lies in the construction and parameterization of the response: the Legendre factor is explicitly determined by the axisymmetric Green's function, providing a physically motivated structure for the learned spectral response.

Axis-dependent spectral structure. The preferred axis enters the derived solution through the relative angle between n and $\mathbf{d}$. Writing $\mu = n \cdot \mathbf{d}$, the spectral multiplier becomes $\kappa_\ell^A = \alpha_\ell P_\ell(\mu)$.

For a fixed μ, this defines a degree-dependent response shared across all harmonic orders m. Different choices of μ generate different Legendre modulation profiles across harmonic degrees. The axis therefore controls the spectral response through its angular relationship with the reference direction, rather than introducing independent transformations across harmonic orders.

A.5 Multi-Channel GA and Parameter Efficiency

Multi-channel axisymmetric response. As described in Section 4.3, the GA operator is extended to multi-channel spherical representations by assigning each latent channel c a learnable unit axis $\mathbf{d}_c \in \mathbb{S}^2$. Given an unconstrained vector $\widetilde{\mathbf{d}}_c \in \mathbb{R}^3$, its normalized direction is $\mathbf{d}_c = \widetilde{\mathbf{d}}_c / \|\widetilde{\mathbf{d}}_c\|_2$.

The corresponding spectral response is

$$\kappa_{\ell c}^A = a_{\theta,c}(\ell) P_\ell(n \cdot \mathbf{d}_c), \tag{50}$$

where $a_{\theta,c}(\ell)$ is the learnable spectral coefficient associated with harmonic degree ℓ and channel c. The multi-channel GA convolution is then

$$z^A = \mathcal{F}_{\mathbb{S}^2}^{-1}\left[\kappa_\theta^A \odot \mathcal{F}_{\mathbb{S}^2}[v]\right]. \tag{51}$$

Each channel retains a single preferred axis across harmonic degrees, preserving the axisymmetric response structure derived above. Channel interactions are subsequently handled by the point-wise MLPs in the AsiNO block, following the network formulation in Figure 1.

Spectral parameter complexity. Let L denote the number of retained harmonic degrees and C the hidden channel dimension. A general degree-dependent dense spectral transformation uses a matrix $W_\ell \in \mathbb{C}^{C \times C}$ for each harmonic degree, requiring $\mathcal{O}(LC^2)$ learnable spectral parameters.

In contrast, the GA spectral convolution uses one learnable coefficient for each degree-channel pair, requiring $\mathcal{O}(LC)$ spectral parameters. The learnable axes introduce an additional $\mathcal{O}(C)$ parameters. Thus, the GA spectral response requires

$$N_{\text{GA}} = \mathcal{O}(LC + C). \tag{52}$$

The point-wise MLPs provide channel mixing and are retained from the baseline network architecture. Their parameters are not included in the spectral response counts above. When these shared components are included, the complete block also contains the corresponding MLP parameters.

This reduction in spectral parameterization follows from replacing independent degree-dependent channel interactions with a structured channel-wise response. Since the overall network architecture is kept consistent with SFNO and GSNO, the resulting comparison isolates the effect of the proposed GA spectral operator.

Overall, the derivation establishes a direct connection between an axisymmetric Green's function and a Legendre-structured spherical spectral convolution. By translating a known physical response structure into the operator formulation, AsiNO provides a structured parameterization of spherical operator learning while retaining the established neural architecture for controlled empirical evaluation.

B EXPERIMENTAL DETAILS

Table 5: Chosen Variables Abbreviation

Abbreviation	Description
10U	Zonal wind velocity at 10m from the surface
10V	Meridional wind velocity at 10m from the surface
2T	Temperature at 2m from the surface
U−−	Zonal wind velocity at pressure level −−
V−−	Meridional wind velocity at pressure level −−
T−−	Temperature at pressure level −−

B.1 SPHERICAL SHALLOW WATER EQUATIONS

Dataset and simulation. The spherical shallow water equations (SWE) describe the evolution of a thin fluid layer on a rotating sphere and are widely used to study large-scale geophysical fluid dynamics (Nair et al., 2005; Bonev et al., 2018; 2023). Following SFNO (Bonev et al., 2023), we generate numerical simulations of the rotating SWE using the spherical harmonic-based solver implemented in `torch-harmonics` (GitHub). The physical state consists of geopotential height (H), vorticity (V), and divergence (D), which jointly characterize the evolution of the spherical fluid system.

The numerical simulations are generated at 128×256 resolution, with a temporal interval of one hour between consecutive model states. The numerical solver uses an internal integration time step of 150 seconds, corresponding to 24 integration steps per forecasting interval. The resulting trajectories are used to learn an autoregressive state-transition operator.

Training and evaluation. All the models use the same settings: (1) training: 50 epochs, 256 samples (initial conditions), batch size of 16, Adam optimizer with learning rate of 2×10^{-3}, and spherical weighted mean relative error as loss function; (2) testing: 50 initial conditions (50 samples) and evaluated using mean relative error (MRE). All models are trained at 128×256 resolution and evaluated on the same spatial discretization for in-distribution prediction. We additionally perform zero-shot cross-scale evaluation at 256×512 resolution without retraining or fine-tuning. For autoregressive evaluation, each predicted state is recursively used as the input for the subsequent forecasting step.

We report the mean relative error (MRE) for H, V, and D at the selected forecasting horizons. To assess long-term prediction stability, we additionally evaluate the evolution of prediction errors over 30 consecutive rollout steps. All models are evaluated using the same test trajectories and prediction protocol.

Model configurations. We compare AsiNO with ClimaX, FourCastNet, SFNO, and GSNO. AsiNO adopts the same network backbone and training configuration as the spherical operator baselines, replacing their original spectral operator with the proposed axisymmetry-informed convolution. The model hyperparameters are summarized in Table 6.

B.2 MULTI-SCALE WEATHER FORECASTING

Dataset. WeatherBench (Rasp et al., 2020) provides a multi-scale benchmark for data-driven global weather prediction using atmospheric reanalysis data (GitHub). Following GSNO (Tang et al., 2026), we conduct experiments using atmospheric variables derived from ERA5 (Hersbach et al., 2020) and 24 common variables are chosen (detailed in Table 5), including 10U, 10V, 2T, U50, U50, U250, U500, U600, U700, U850, U925, V50, V250, V500, V600, V700, V850, V925, T50, T250, T500, T600, T700, T850, T925. We consider six variables as the prediction target: 10-meter zonal wind (10u), 10-meter meridional wind (10v), 2-meter temperature (2T), temperature at 600 hPa (T600), zonal wind at 600 hPa (U600), and meridional wind at 600 hPa (V600).

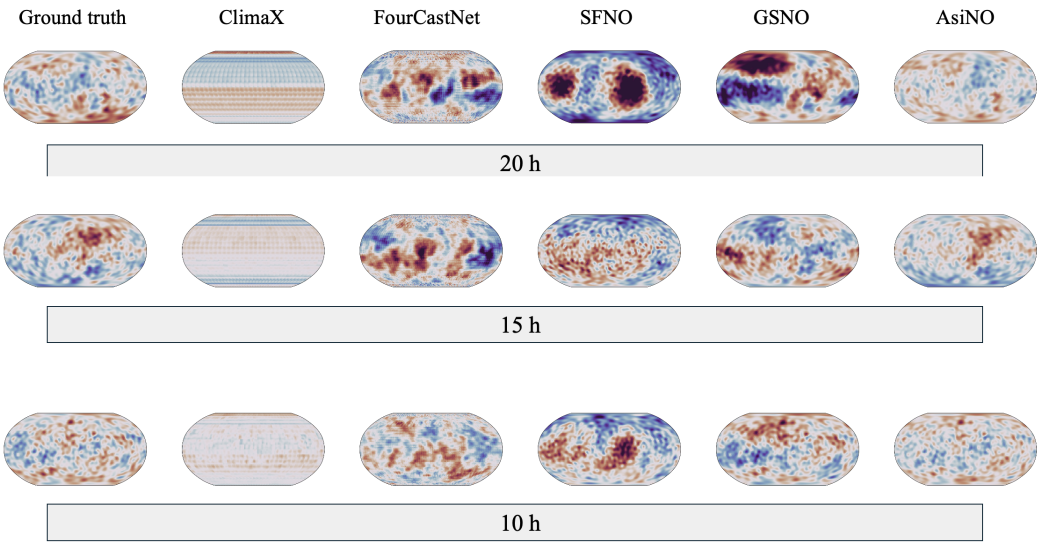

Figure 5: Predicted geopotential height (H) at 10h, 15h and 20h of different methods.

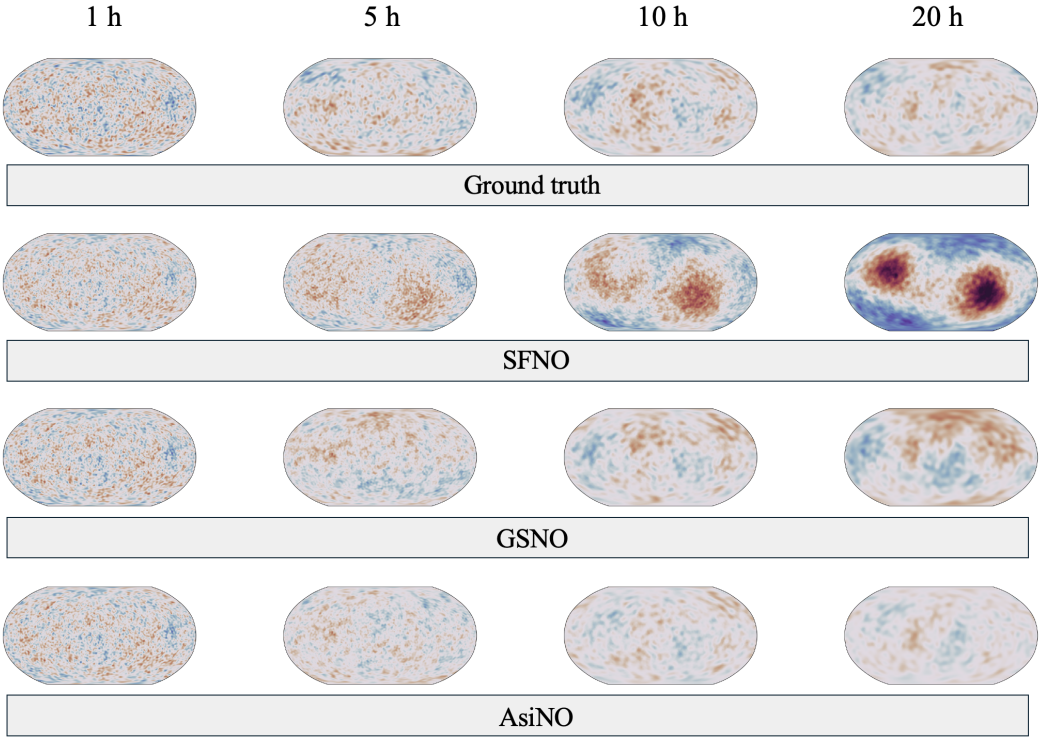

Figure 6: Zero-shot cross-scale generalization on SWE: Predicted geopotential height (H) at multiple time horizons of different methods, trained at 128×256 and evaluated at 256×512 resolution.

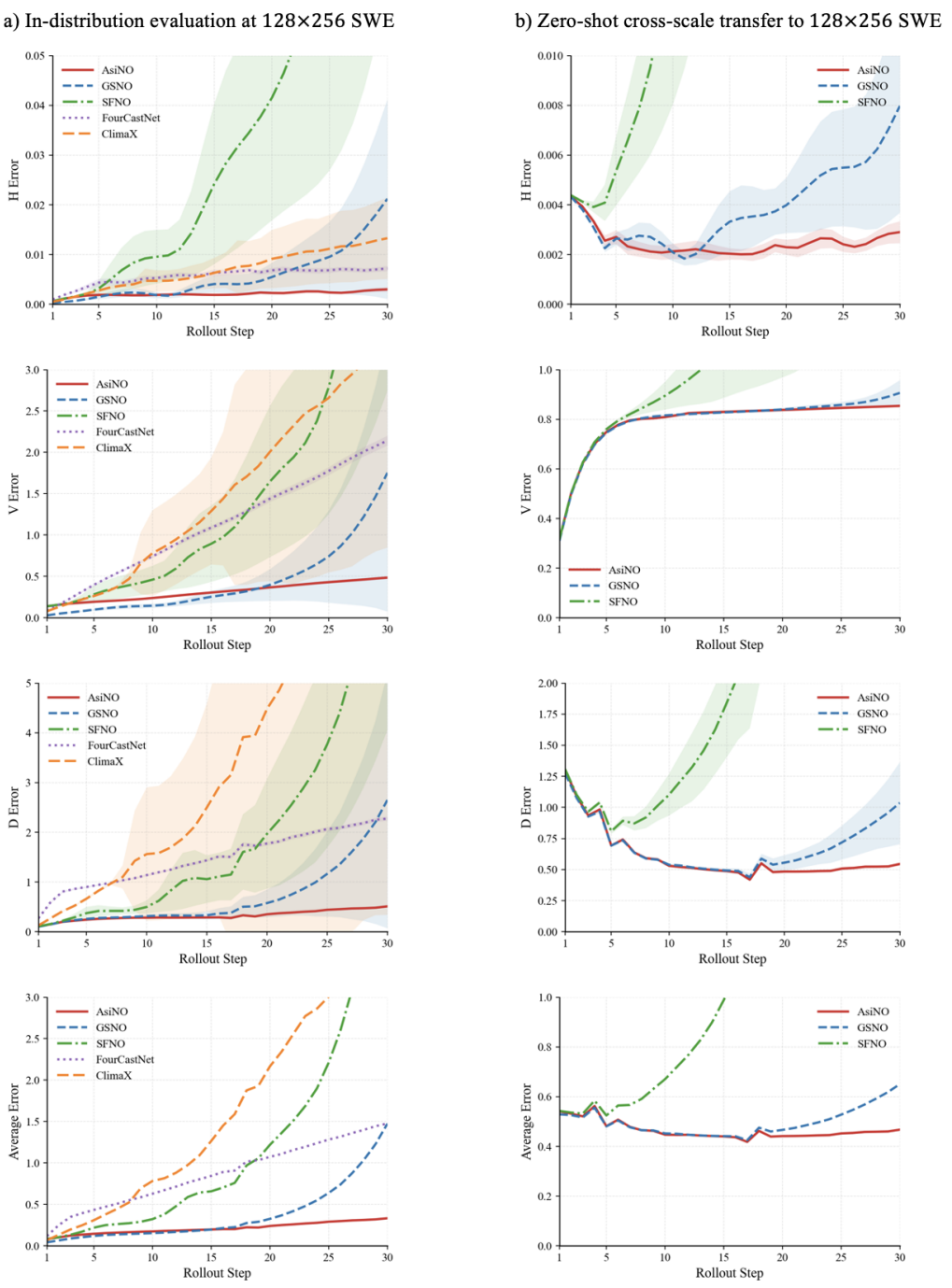

Figure 7: **Autoregressive rollout error maps on SWE.** Comparison of H, V, D, and average prediction errors over 30 rollout steps for in-distribution evaluation at 128×256 and zero-shot cross-scale transfer to 256×512 resolution.

Table 6: Model hyperparameters on spherical shallow water equations (SWE).

Hyperparameters	AsiNO	GSNO	SFNO	FourCastNet	ClimaX
Blocks	4	4	4	8	6
Embedding dimension	64	64	64	128	256
Activation function	GELU	GELU	GELU	GELU	GELU
MLP ratio	2	2	2	2	2
Patch size	–	–	–	4×4	4×4
Parameters (M)	0.077	1.206	0.755	1.369	3.786

Table 7: Model hyperparameters on WeatherBench (WB).

Hyperparameters	AsiNO	GSNO	SFNO	FourCastNet	ClimaX
Blocks	5	5	5	8	6
Embedding dimension	256	256	256	384	256
Activation function	GELU	GELU	GELU	GELU	GELU
MLP ratio	2	2	2	2	2
Patch size	–	–	–	4×4	4×4
Parameters (M)	1.2	4.0	3.7	5.3	5.4

We use $5.625°$ (32×64) and $2.8125°$ (64×128) WB datasets with a temporal resolution of 1 hour for evaluation. Following GSNO (Tang et al., 2026), the training period spans 1979-2015, validation is from 2016, and testing spans 2017-2018. All models are under the same settings: batch size of 64, 50 epochs, MRE as the loss function, Adam optimizer with learning rate of 1×10^{-4}.

Training and evaluation. All models are trained and evaluated at 32×64 resolution following the experimental protocol of GSNO (Tang et al., 2026). We consider forecasting horizons of 72 and 120 hours and evaluate prediction accuracy using ACC and MSE. For cross-scale evaluation, the models trained at 32×64 are directly applied to 64×128 resolution without additional optimization. All methods use the same input variables, prediction targets, and evaluation protocol.

Model configurations. We compare AsiNO with ClimaX, FourCastNet, SFNO, and GSNO. To ensure a controlled operator-level comparison, AsiNO follows the same spherical network architecture as SFNO and GSNO, including the number of operator blocks, embedding dimensions, nonlinear transformations, and residual connections. The primary architectural difference lies in the proposed axisymmetry-informed spectral convolution. Detailed hyperparameters are provided in Table 7.

B.3 Loss function and Evaluation Metrics

We evaluate the prediction accuracy using three metrics: mean relative error (MRE) for spherical shallow water dynamics, and anomaly correlation coefficient (ACC) and mean squared error (MSE) for weather forecasting. All metrics are computed consistently across the compared methods.

Mean Relative Error (MRE). Following the relative-error evaluation adopted in spherical physical dynamics (Bonev et al., 2023), we measure the discrepancy between predicted and reference fields using

$$\mathrm{MRE} = \frac{1}{N} \sum_{i=1}^{N} \frac{\|\widehat{\mathbf{y}}_i - \mathbf{y}_i\|_2}{\|\mathbf{y}_i\|_2}, \tag{53}$$

where $\widehat{\mathbf{y}}_i$ and $\mathbf{y}_i$ denote the predicted and reference fields of the i-th test sample, respectively, and N is the number of test samples. MRE is evaluated separately for geopotential height (H), vorticity (V), and divergence (D), with their arithmetic mean reported as the overall prediction error. Lower MRE indicates higher prediction accuracy.

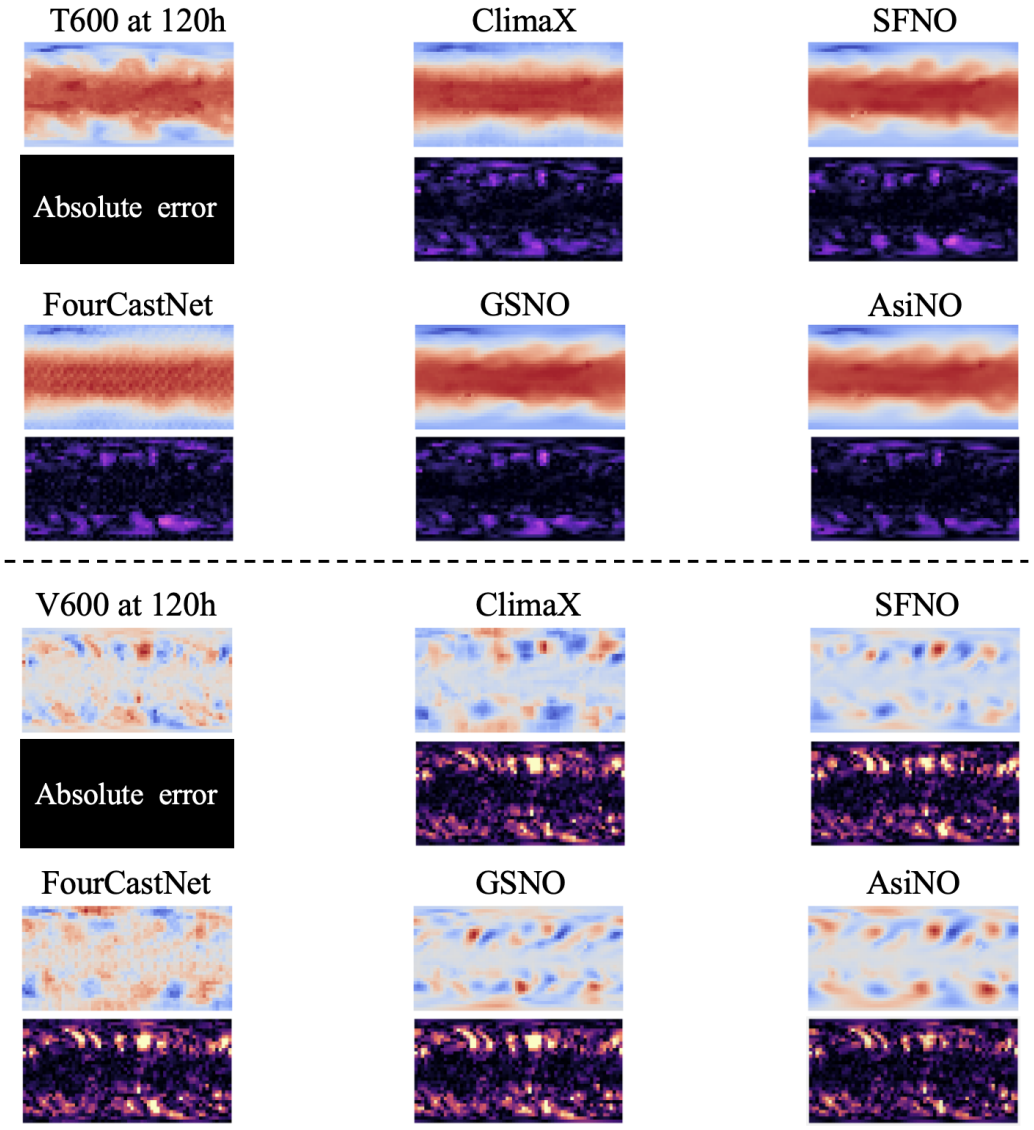

Figure 8: In-distribution prediction results on WeatherBench.: Visualization of different variables at 120h by different methods trained at 32×64 and evaluated at 32×64 resolution. The absolute error maps are provided for comparison (darker is better).

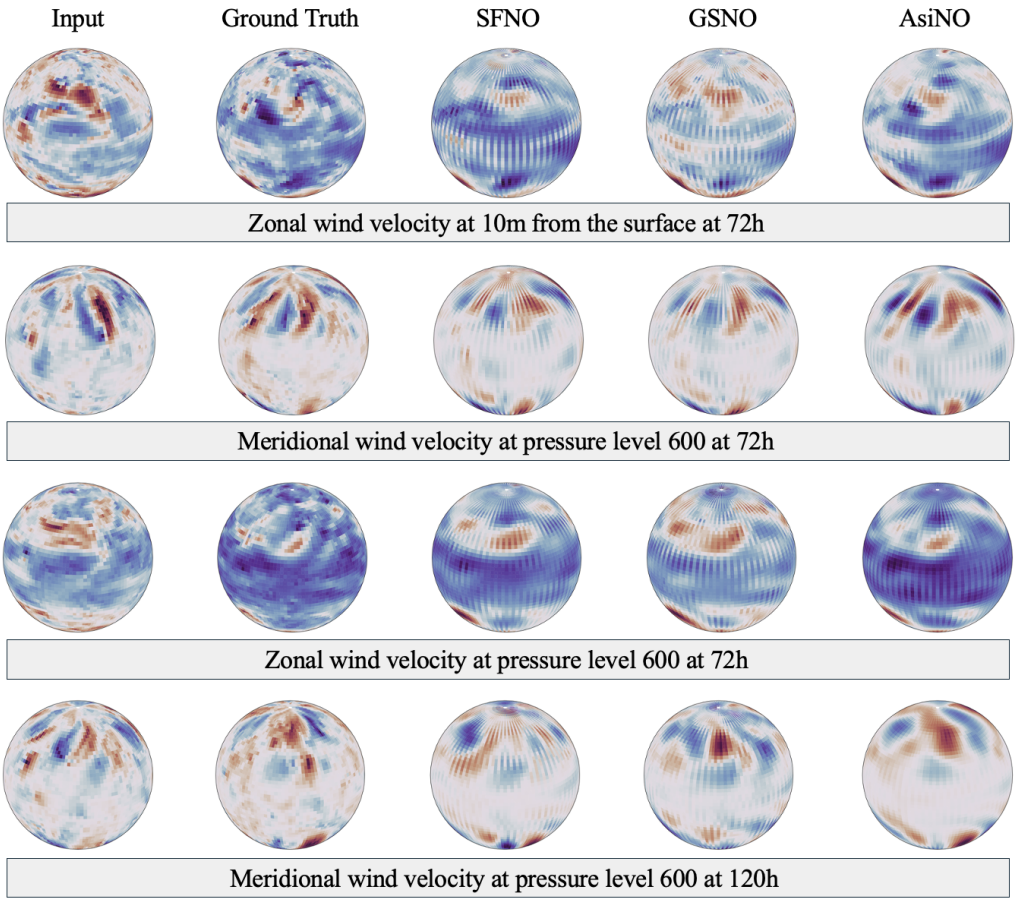

Figure 9: Additional cross-scale generalization results on WeatherBench: Visualization of different variables and prediction horizons by different methods trained at 32×64 and evaluated at 64×128 resolution. Zoom in for clear comparison.

Table 8: Weather forecasting performance measured by weighted mean squared error (MSE). Results are reported as mean ± standard deviation. All models are trained at 32×64, and cross-scale transfer is evaluated at 64×128 without retraining. Lower is better.

Model	Params	Horizon	10u	10v	2T	T600	U600	V600	Avg.
In-distribution (32×64)									
SFNO	3.7M	72h	0.3640 ± 0.0296	0.5109 ± 0.0454	0.0095 ± 0.0015	0.0282 ± 0.0041	0.2898 ± 0.0251	0.5321 ± 0.0541	0.2891
		120h	0.4992 ± 0.0475	0.7274 ± 0.0788	0.0143 ± 0.0026	0.0489 ± 0.0080	0.4447 ± 0.0447	0.8032 ± 0.1013	0.4230
GSNO	4.0M	72h	0.3537 ± 0.0308	0.5022 ± 0.0473	0.0097 ± 0.0014	0.0283 ± 0.0043	0.2813 ± 0.0260	0.5164 ± 0.0576	0.2819
		120h	0.4989 ± 0.0474	0.7280 ± 0.0853	0.0141 ± 0.0024	0.0487 ± 0.0080	0.4421 ± 0.0448	0.8022 ± 0.1099	0.4223
FourCastNet	5.3M	72h	0.4021 ± 0.0330	0.5943 ± 0.0548	0.0127 ± 0.0017	0.0339 ± 0.0045	0.3215 ± 0.0279	0.6154 ± 0.0621	0.3300
		120h	0.5642 ± 0.0482	0.8285 ± 0.0900	0.0189 ± 0.0028	0.0561 ± 0.0089	0.5058 ± 0.0441	0.9181 ± 0.1150	0.4819
ClimaX	5.4M	72h	0.4071 ± 0.0345	0.6015 ± 0.0592	0.0124 ± 0.0016	0.0340 ± 0.0045	0.3231 ± 0.0284	0.6233 ± 0.0642	0.3336
		120h	0.5267 ± 0.0473	0.7417 ± 0.0726	0.0244 ± 0.0039	0.0581 ± 0.0092	0.4891 ± 0.0468	0.8327 ± 0.0980	0.4454
AsiNO	1.2M	72h	$\mathbf{0.3317 \pm 0.0282}$	$\mathbf{0.4687 \pm 0.0437}$	0.0093 ± 0.0014	$\mathbf{0.0268 \pm 0.0037}$	$\mathbf{0.2624 \pm 0.0240}$	$\mathbf{0.4746 \pm 0.0513}$	**0.2622**
		120h	$\mathbf{0.4696 \pm 0.0407}$	$\mathbf{0.6878 \pm 0.0754}$	$\mathbf{0.0136 \pm 0.0024}$	$\mathbf{0.0452 \pm 0.0071}$	$\mathbf{0.4102 \pm 0.0395}$	$\mathbf{0.7542 \pm 0.1002}$	**0.3968**
Cross-scale generalization ($32 \times 64 \rightarrow 64 \times 128$)									
SFNO	3.7M	72h	0.4705 ± 0.0316	0.5782 ± 0.0444	0.0133 ± 0.0017	0.0376 ± 0.0065	0.3628 ± 0.0278	0.6049 ± 0.0545	0.3446
		120h	0.5649 ± 0.0459	0.7542 ± 0.0727	0.0177 ± 0.0033	0.0589 ± 0.0087	0.4859 ± 0.0428	0.8124 ± 0.0984	0.4490
GSNO	4.0M	72h	0.4960 ± 0.0315	0.6837 ± 0.0460	0.0143 ± 0.0014	0.0424 ± 0.0046	0.4046 ± 0.0428	0.6550 ± 0.0624	0.3827
		120h	0.5585 ± 0.0441	0.7920 ± 0.0796	0.0183 ± 0.0028	0.0555 ± 0.0086	0.4815 ± 0.0410	0.8331 ± 0.1034	0.4565
AsiNO	1.2M	72h	$\mathbf{0.3644 \pm 0.0275}$	$\mathbf{0.5111 \pm 0.0423}$	$\mathbf{0.0105 \pm 0.0015}$	$\mathbf{0.0294 \pm 0.0041}$	$\mathbf{0.2944 \pm 0.0243}$	$\mathbf{0.5187 \pm 0.0549}$	**0.2881**
		120h	$\mathbf{0.4773 \pm 0.0399}$	$\mathbf{0.6957 \pm 0.0732}$	$\mathbf{0.0142 \pm 0.0025}$	$\mathbf{0.0460 \pm 0.0075}$	$\mathbf{0.4137 \pm 0.0399}$	$\mathbf{0.7501 \pm 0.0988}$	**0.3995**

Table 9: Computational efficiency comparison for 120h weather forecasting at 32×64 resolution. Training time is reported per epoch, and testing time is measured over the entire test set.

Model	Params (M)	Training (s/epoch)	Testing (s)
AsiNO	1.2	857	26
GSNO	4.0	883	28
SFNO	3.7	836	26
FourCastNet	5.3	797	24
ClimaX	5.4	778	25

Anomaly Correlation Coefficient (ACC). ACC measures the spatial correspondence between predicted and reference atmospheric anomalies relative to climatology (Nguyen et al., 2023). Let $\widehat{\mathbf{y}}$ and $\mathbf{y}$ denote the predicted and reference atmospheric fields, respectively, and let $\mathbf{c}$ represent the corresponding climatological mean. The latitude-weighted ACC is defined as

$$\mathrm{ACC} = \frac{\sum_{j=1}^{M} w_j (\widehat{y}_j - c_j)(y_j - c_j)}{\sqrt{\sum_{j=1}^{M} w_j (\widehat{y}_j - c_j)^2} \sqrt{\sum_{j=1}^{M} w_j (y_j - c_j)^2}}, \tag{54}$$

where M denotes the number of spatial grid points and w_j is the latitude-dependent area weight at location j. ACC evaluates the agreement between predicted and reference anomaly patterns, with higher values indicating better forecasting performance.

Mean Squared Error (MSE). We additionally quantify the magnitude of atmospheric prediction errors using the latitude-weighted MSE (Rasp et al., 2020; Nguyen et al., 2023):

$$\mathrm{MSE} = \frac{\sum_{j=1}^{M} w_j (\widehat{y}_j - y_j)^2}{\sum_{j=1}^{M} w_j}, \tag{55}$$

where the latitude-dependent weights account for the unequal surface areas represented by different grid points on the sphere. Lower MSE indicates more accurate predictions.

For each atmospheric variable, ACC and MSE are evaluated at the 72-hour and 120-hour forecasting horizons. We report the mean and standard deviation across test samples, together with the arithmetic average over the six atmospheric variables.

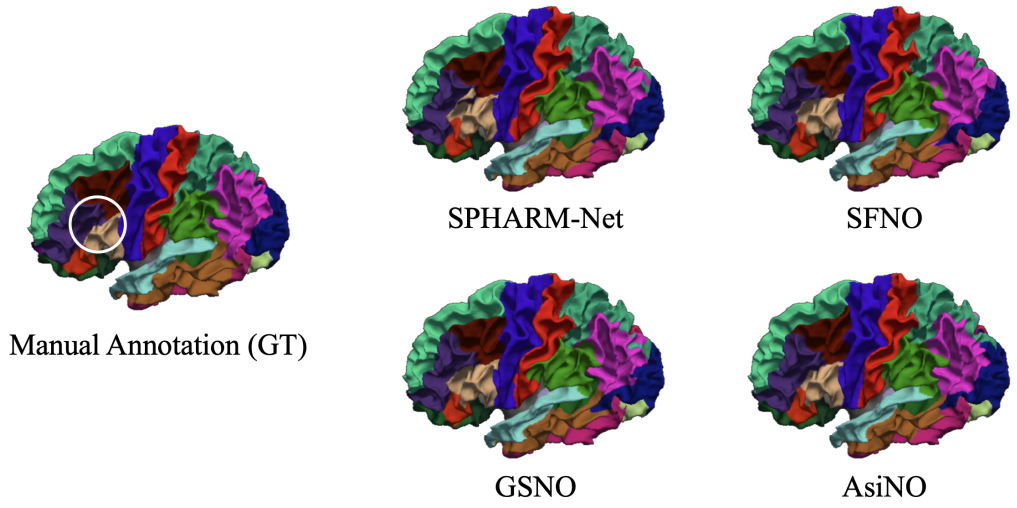

Figure 10: **Cortical surface parcellation in native space.** Additional visualizations of cortical parcellation results on NAMIC and Mindboggle, comparing the predictions of SFNO, GSNO, and AsiNO with the ground truth.

B.4 Cortical Surface Modelling

Datasets and preprocessing. We evaluate cortical surface parcellation on the NAMIC (Desikan et al., 2006) and Mindboggle (Klein & Tourville, 2012) datasets following the data preprocessing and spherical surface representation of SPHARM-Net (Desikan et al., 2006; Ha & Lyu, 2022) (GitHub). Both datasets provide cortical surface representations with anatomical parcellation labels, allowing the evaluation of spherical learning methods on biomedical geometric data. The cortical surfaces and their corresponding parcellation labels are represented in spherical space for model training and evaluation. Further details regarding the datasets, spherical parameterization, and preprocessing procedures can be found in the original SPHARM-Net study (Ha & Lyu, 2022).

Training and evaluation. Each dataset is partitioned into training, validation, and test sets with a fixed ratio of 60%/20%/20%. We perform five independent runs with different random seeds while maintaining identical data partitions. All compared methods use the same preprocessing, training strategy, and evaluation protocol. In particular, SFNO, GSNO, and AsiNO adopt the same network backbone as the resnet-type architecture (Bonev et al., 2023), including the number of operator blocks, embedding dimensions, point-wise MLPs, and residual connections, with the spectral operator being the primary architectural difference.

For each independent run, we compute the mean Dice score across all test subjects and anatomical regions (Ha & Lyu, 2022). The final results are reported as the mean and standard deviation of the five run-level Dice scores, providing an assessment of prediction accuracy and variability across independent training runs.

Evaluation metric. We evaluate cortical parcellation using the Dice similarity coefficient (DSC) (Ha & Lyu, 2022), which measures the spatial overlap between predicted and reference anatomical regions. For a predicted region $\widehat{\Omega}_r$ and its corresponding ground-truth region Ω_r, the Dice score is defined as

$$\mathrm{Dice}_r = \frac{2|\widehat{\Omega}_r \cap \Omega_r|}{|\widehat{\Omega}_r| + |\Omega_r|}, \tag{56}$$

where $|\cdot|$ denotes the number of spherical surface vertices belonging to the corresponding region. The overall Dice score is obtained by averaging across anatomical regions and test subjects. Higher Dice scores indicate greater agreement between the predicted and reference cortical parcellations.

We report the number of trainable parameters and training throughput for all compared methods. Inference throughput is measured in samples per second under the same evaluation protocol. As reported in Table 4, AsiNO requires approximately 0.85M parameters on both datasets, substantially fewer than the compared spherical operator models, while achieving higher inference throughput. **Additional visualization in native space:** Figure 10 presents additional cortical parcellation results in native space, complementing the spherical-space visualizations in Figure 4. The native-space representations provide an anatomical view of the predicted cortical regions, allowing further qualitative comparison of the parcellation patterns produced by different spherical operators.