Banner appropriate to article type will appear here in typeset article

Online learning of subgrid-scale models for quasi-geostrophic turbulence in planetary interiors

Hugo Frezat[1] †, Thomas Gastine[1] and Alexandre Fournier[1]

[1]Université Paris Cité, Institut du Physique du Globe de Paris, UMR 7154 CNRS

(Received xx; revised xx; accepted xx)

The use of machine learning to represent subgrid-scale (SGS) dynamics is now well established in weather forecasting and climate modelling. Recent advances have demonstrated that SGS models trained via "online" end-to-end learning – where the dynamical solver operating on the filtered equations participates in the training – can outperform traditional physics-based approaches. Most studies, however, have focused on idealised periodic domains, neglecting the mechanical boundaries present e.g. in planetary interiors. To address this issue, we consider two-dimensional quasi-geostrophic turbulent flow in an axisymmetric bounded domain that we model using a pseudo-spectral differentiable solver, thereby enabling online learning. We examine three configurations, varying the geometry (between an exponential container and a spherical shell) and the rotation rate. Flow is driven by a prescribed analytical forcing, allowing for precise control over the energy injection scale and an exact estimate of the power input. We evaluate the accuracy of the online-trained SGS model against the reference direct numerical simulation using integral quantities and spectral diagnostics. In all configurations, we show that an SGS model trained on data spanning only one turnover time remains stable and accurate over integrations at least a hundred times longer than the training period. Moreover, we demonstrate the model's remarkable ability to reproduce slow processes occurring on time scales far exceeding the training duration, such as the inward drift of jets in the spherical shell. These results suggest a promising path towards developing SGS models for planetary and stellar interior dynamics, including dynamo processes.

Key words: quasi-geostrophic turbulence, subgrid-scale model, machine learning

---

MSC Codes *(Optional)* Please enter your MSC Codes here

1. Introduction

The interior of active planets and stars hosts turbulent flow operating over a wide range of time and length scales. In the Earth's fluid outer core for example, the convective flow of liquid iron sustains the Earth's magnetic field against ohmic decay. This process, referred to as the geodynamo, has been at work for over 3.5 billion years. Measurements of the present and past

† Email address for correspondence: hugo.frezat@gmail.com

variability of the geomagnetic field point to characteristic time scales ranging from a few years to millions of years (see the reviews by Landeau *et al.* 2022; Finlay *et al.* 2023, and references therein). Likewise, the flow in the outer core is expected to span a vast range of length scales; the thickness of viscous boundary layers is not expected to exceed one metre, while core-flow studies have evidenced a planetary-scale eccentric gyre whose size is commensurate with that of the core (Pais & Jault 2008). Observation of the geodynamo is limited and our understanding of its behaviour has benefited from numerical simulations, following the pioneering work of Glatzmaier & Roberts (1995). Standing open questions remain, however, such as the origin of geomagnetic polarity reversals and the reasons explaining the variability of occurrence of these extreme events over Earth's evolution. Addressing these questions using numerical models is hampered by the range of time and length scales necessary to achieve this task, as performing direct numerical simulations of the geodynamo with meter-scale resolution over geological time scales is unfeasible. In practice, numerical studies of reversals are performed at intermediate resolution (e.g. Wicht & Meduri 2016); they favor long-term, large-scale dynamics over fast, small-scale dynamics, under the tacit assumption that small-scale flow does not impact the long-term, large-scale geomagnetic variability.

In order to deal with the unresolved, subgrid-scale (SGS) flow, a popular approach is to resort to an eddy-diffusivity model, whereby small-scale transport of momentum, heat, or magnetic field, is parameterized in terms of an effective diffusivity. The simplest model prescribes enhanced scalar values for the diffusivities. The scalar model of Smagorinsky (1963) is a bit more sophisticated, as the value of the diffusivity rests upon an invariant of the strain rate tensor. Even if elaborate schemes that determine the eddy diffusivity from the large-scale flow have been explored for Navier-Stokes turbulence (Lesieur & Metais 1996; Meneveau & Katz 2000), geodynamo studies have resorted, for the most part, to a variant of eddy-diffusivity called hyperdiffusivity, which amounts to artificially increase the order of the diffusion operator acting on the field variables. Hyperdiffusivity is commonly used in atmospheric and oceanic fluid dynamics, especially to study quasi-geostrophic (QG) turbulence (see e.g. Maltrud & Vallis 1991). For three-dimensional spherical dynamo studies, which are based on a spherical harmonic (SH) expansion of the field variables in the horizontal direction, the traditional implementation of hyperdiffusivity consists of specifying an increase of diffusivities above some threshold SH degree ℓ_c. The functional form of the increase with SH degree ℓ varies among authors (see Glatzmaier & Roberts 1995; Kuang & Bloxham 1999; Aubert *et al.* 2017, for a few examples). Suffice it to say here that these hyperdiffusive models are by design anisotropic, since diffusion along the radial direction is unchanged. Also, from a spectral perspective, they can not account for non-local interactions.

In addition to enabling long-range interactions, a physically adequate parameterization of subgrid scales should capture the anisotropic nature of core turbulence, due to the combined action of magnetic field and planetary rotation (Braginsky & Meytlis 1990; Matsushima *et al.* 1999). In this respect, the similarity model proposed by Germano (1986) after earlier work by Leonard (1975) and Bardina *et al.* (1980) appears to be a good candidate, even if studies in nonmagnetic turbulent flows showed that it cannot alone provide adequate dissipation at small scales (Meneveau & Katz 2000). It was tested in the context of magneto-convection under rapid rotation in a triply periodic box by Buffett (2003), alongside a scalar eddy-diffusivity model, an hyperdiffusive model, and a Smagorinsky model. Comparisons of the SGS fluxes of momentum and heat calculated from a fully resolved direct numerical simulation showed the failure of the three scalar models to reproduce the anisotropy of the fluxes, in contrast to estimates based on the similarity model. Matsui & Buffett (2005) extended that study to planar dynamo models at moderate levels of turbulence (Reynolds number $Re \approx 120$). Investigating the full dynamo problem, they found that the similarity model can provide a valuable estimate of the SGS heat flux, Reynolds stress, Maxwell stress and electromotive force. Yet substantial commutation errors (e.g. Ghosal & Moin 1995) occurred near the top and bottom boundaries of the domain, in the boundary layers where the inhomogeneity of the grid is the strongest, leading to overly dissipative patterns. Matsui & Buffett (2007*a*) proposed correction terms to mitigate these errors, which they subsequently used in conjunction with the dynamic version of the similarity model of Germano *et al.* (1991) to improve on their plane layer dynamo SGS models (Matsui & Buffett 2007*b*). Comparison with a direct numerical simulation (DNS) demonstrated the increased fidelity of this approach, provided the time-and-space-dependent coefficient of each SGS flux be estimated independently. Further numerical experiments by Matsui & Buffett (2012, 2013) confirmed the relevance of such SGS models (that include commutation error correction terms) in spherical shell geometry, still in a moderately turbulent context ($Re \approx 120$). In spite of these rather encouraging results, to date, dynamic scale similarity models have neither been implemented nor tested in state-of-the-art spherical dynamo codes.

An alternative option, based on machine learning (ML) using DNS data), has recently come to the fore. This trending topic is reviewed by Brunton *et al.* (2020); Vinuesa & Brunton (2022). These ML components have shown their relevance in several applications, including atmospheric circulation (Yuval & O'Gorman 2020; Arcomano *et al.* 2023), oceanic circulation (Bolton & Zanna 2019) or sea-ice dynamics (Finn *et al.* 2023), but to our knowledge, applications to flow in planetary interiors remain to be seen. Beyond traditional SGS terms coming from coarse-graining, data-driven approaches have also been proposed to learn parameterizations for more complex or less directly observable subgrid processes. In particular, ML models have been trained to emulate the effects of cloud formation, precipitation, and radiative transfer (e.g. Rasp *et al.* 2018; Dueben & Bauer 2018; Yuval *et al.* 2021). In these works, the machine learning component has been trained "offline", by learning the SGS terms independently of their interaction with the solver dynamics. When these learned components are plugged back into a simulation, the lack of coupling with the governing equations during training can potentially cause instabilities and accumulation of small-scale energy. Using differentiable programming (Gelbrecht *et al.* 2023) allowed models to be trained end-to-end, also referred to as "online learning" (Rasp 2020). This new paradigm has been applied incrementally to more complex physical systems, including two-dimensional turbulence (Kochkov *et al.* 2021; Frezat *et al.* 2022) and, more recently, to a full weather and climate forecasting on the sphere (Kochkov *et al.* 2024). Note, however, that "online"-trained models that target SGS fluxes from a high-resolution reference simulation have only been applied to idealised domains without boundaries, thereby avoiding commutation errors that may arise inside, or in the vicinity of the boundary layers.

Our goal in this paper is to assess the potential of machine learning to describe SGS fluxes in fluid planetary interiors, in what can be regarded as the first step towards using ML-based SGS models for dynamo simulations. To that end, we consider a prototype consisting of the QG equations in a container with rigid boundaries. Flow is driven by a network of vorticity pumps, following the analytical forcing introduced by Lemasquerier *et al.* (2023), which makes it possible to control the scale of energy injection. This two-dimensional model can account for the formation and dynamics of zonal jets in the experiment by Lemasquerier *et al.* (2023).

We implement a differentiable solver of the governing equations that can be used to train "online" SGS model corrections. We show that the learned models are capable of producing stable and accurate simulations with respect to many metrics, using a relatively small amount of training data. In the current workflow, we use a negligible subset of data from a DNS to train a model that then extends the simulation over long timescales.

In §2, we describe the physical system and the spectral method, implemented within a differentiable programming framework. In §3, we introduce a Galerkin projection for generating coarse-grained data that comply with the boundary conditions. We then detail the implicit benefits of online learning for mitigating the commutation errors that arise when explicitly computing the SGS terms, and discuss the choice of loss function alongside subtleties related to the axisymmetric velocity in our spectral-space neural network architecture. In §4, we evaluate the learned models across three experiments representing different dynamical regimes, demonstrating their long-term stability and assessing their accuracy using statistical and spectral metrics. Finally, §5 examines the computational cost of the embedded neural network and outlines the challenges that must be addressed to apply this methodology to a model of Earth's geodynamo.

2. The physical model and its numerical approximation

2.1. *Governing equations*

We study the dynamics of a fluid of kinematic viscosity ν enclosed in a container rotating at constant angular velocity Ω. The direction of rotation defines the z-axis; the container is assumed to be symmetric about this axis, and about a plane orthogonal to z, which we shall refer to as the equatorial plane. In the following, we operate in cylindrical coordinates (s, φ, z) and consider the quasi-geostrophic (QG) approximation of the Navier-Stokes equations, which is expected to be valid in the limit of rapid rotation (e.g. Aubert *et al.* 2003; Guervilly & Cardin 2017; Lemasquerier *et al.* 2023). This approximation implies that the flow is invariant along the direction of rotation and that its complete characterisation can be performed in the equatorial domain, defined by the intersection of the volume of the container and the equatorial plane; consequently, field variables depend only upon s and φ. In this study, the equatorial domain is an annulus of inner radius s_i and outer radius s_o with a radius ratio $s_i/s_o = 0.35$. The shape of the container is further specified by the half-height of the container h, which is a function of the sole cylindrical radius s. In the following we shall consider two shapes: the exponential container and the spherical shell, see Fig. 1(a) and Fig. 1(b), respectively. In this paper, we adopt a dimensionless formulation of the QG equations using the annulus gap $d = s_o - s_i$ as the reference length scale and the viscous dissipation time d^2/ν as the reference time scale. Consequently, the non-dimensional Ekman number describing the ratio of viscous forces to Coriolis forces is given by $E = \nu/\Omega d^2$. The continuity equation under the QG approximation (Gillet & Jones 2006) then reads

$$\frac{1}{s}\frac{\partial(su_s)}{\partial s} + \frac{1}{s}\frac{\partial u_\varphi}{\partial \varphi} + \beta u_s = 0, \tag{2.1}$$

where the three-dimensional geometry of the fluid container is reflected by the β parameter which depends on the half-height h at the cylindrical radius s,

$$\beta = \frac{1}{h}\frac{\mathrm{d}h}{\mathrm{d}s}. \tag{2.2}$$

The two-dimensional nature of the problem makes it convenient to adopt a vorticity-streamfunction formulation, as described by Aubert *et al.* (2003); Gillet & Jones (2006) and implemented by Gastine (2019) in the reference open-source code `pizza`†. The cylindrically-radial and azimuthal velocity components are hence expressed using a streamfunction ψ such

† https://github.com/magic-sph/pizza

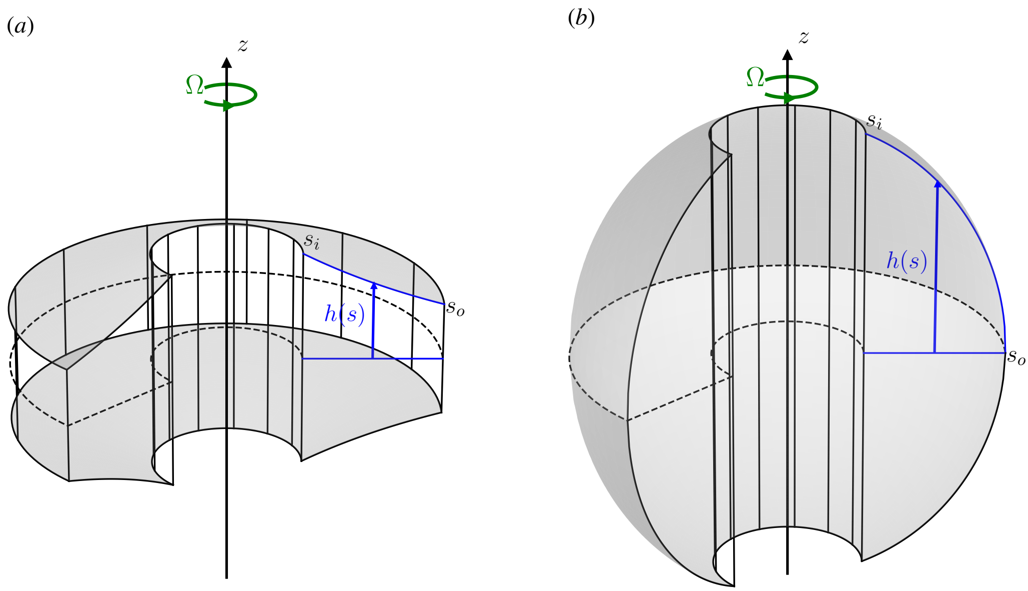

Figure 1: (*a*) Exponential and (*b*) spherical container geometries. The shape is prescribed through the half-height $h(s)$, where the cylindrical radius s varies between s_i and s_o. The container rotates at constant angular velocity Ω about the z-axis.

that

$$u_s = \frac{1}{s}\frac{\partial\psi}{\partial\varphi}, \tag{2.3}$$

$$u_\varphi = \langle u_\varphi\rangle_\varphi - \frac{\partial\psi}{\partial s} - \beta\psi. \tag{2.4}$$

The azimuthal velocity u_φ is explicitly decomposed into a mean zonal flow denoted by the angle brackets $\langle\cdots\rangle_\varphi$ and a non-zonal part, in order to ensure the periodicity of pressure (Plaut & Busse 2002). The axial vorticity ω is then expressed by

$$\omega = \frac{1}{s}\frac{\partial(s\langle u_\varphi\rangle_\varphi)}{\partial s} - \frac{1}{s}\frac{\partial(\beta s\psi)}{\partial s} - \nabla^2\psi. \tag{2.5}$$

Fluid motions are driven by a prescribed vorticity source term $\mathcal{F}$, such that the equations for the axial vorticity and the mean zonal velocity under the QG approximation read

$$\frac{\partial\omega}{\partial t} + \boldsymbol{\nabla\cdot}(\boldsymbol{u}\omega) = \frac{2}{E}\beta u_s + \mathcal{F} + \nabla^2\omega - \Upsilon\omega, \tag{2.6}$$

$$\frac{\partial\langle u_\varphi\rangle_\varphi}{\partial t} + \langle u_s\omega\rangle_\varphi = -\frac{\langle u_\varphi\rangle_\varphi}{s^2} + \nabla^2\langle u_\varphi\rangle_\varphi - \Upsilon\langle u_\varphi\rangle_\varphi, \tag{2.7}$$

supplemented with appropriate boundary and initial conditions to be defined below. The analytical expressions for half-height h, β and Ekman pumping Υ depend on the geometry of the three-dimensional container (here, exponential or spherical) and are given in Table 1. The contribution due to the Ekman pumping follows from the generic expression given by Greenspan (1968), while its derivation for the exponential container is provided in Appendix A; the spherical container case is detailed in Appendix A of Schaeffer & Cardin (2005). In both cases, the complete expression contains more than the term proportional to ω retained in this study, as it also includes terms linear in u_s and u_φ. In preliminary

Geometry	h	β	Υ
Exponential	$\exp(\beta s)$ $(\beta \in \mathbb{R}^-)$	β	$\dfrac{(1+\beta^2 h^2)^{1/4}}{E^{1/2} h}$
Spherical	$(s_o^2 - s^2)^{1/2}$	$-\dfrac{s}{h^2}$	$\dfrac{s_o^{1/2}}{E^{1/2} h^{3/2}}$

Table 1: Analytical expression of geometry-dependent quantities; half-height h, β and Ekman pumping contribution Υ for exponential and spherical containers.

calculations carried out with the reference `pizza` code (Gastine 2019), we found those to have a negligible impact on the statistical properties of the flow compared to the term in ω. Consequently, we decided for computational convenience to retain that latter term only.

The stationary source term $\mathcal{F}$ consists of N Gaussian sources of positive or negative vorticity, of characteristic half-width $\ell_{\mathcal{F}}$, that mimics the experimental setup of Lemasquerier *et al.* (2023), positive and negative Gaussian sources of vorticity of characteristic half-width $\ell_{\mathcal{F}}$, that mimic the experimental setup of Lemasquerier *et al.* (2023),

$$\mathcal{F}(x,y) = a_{\mathcal{F}} \sum_{i=1}^{N} (-1)^i \exp\left[-\left(\frac{x-x_i}{\ell_{\mathcal{F}}}\right)^2 - \left(\frac{y-y_i}{\ell_{\mathcal{F}}}\right)^2\right], \tag{2.8}$$

where (x, y) are the non-dimensional Cartesian coordinates of the center of each vortex, and $a_{\mathcal{F}}$ is the forcing amplitude. Throughout this study, we keep the same forcing structure and amplitude. The number of vortices N is controlled by the vortex spacing $\Delta_{\mathcal{F}} = 0.08$, while each individual vortex half-width $\ell_{\mathcal{F}}$ is set to 0.04. Figure 2 shows the corresponding distribution of vorticity sources over the annular domain.

Finally, we assume that the fluid is initially at rest and we prescribe no-slip mechanical boundary conditions at both boundaries, such that,

$$\psi = \frac{\partial \psi}{\partial s} = \langle u_\varphi \rangle_\varphi = 0 \text{ at } s = s_i, s_o. \tag{2.9}$$

2.2. *Numerical discretisation*

Since the boundary conditions are separable, any unknown field f can be expanded upon the tensor product of a periodic basis in the azimuthal direction and a bounded basis in the radial direction.

The expansion for the azimuthal direction is carried out by truncated (real) Fourier series on equally-spaced grid points $\varphi_k \in [0, 2\pi)$

$$f(s, \varphi_k) \approx {\sum_{m=0}^{N_m}}' \Re\left\{ f_m(s)\, e^{\mathrm{i} m \varphi_k} \right\}, \tag{2.10}$$

where the prime summation indicates that $\forall m > 0$, f_m coefficients are multiplied by two to account for the complex conjugate symmetry $f_{-m}^* = f_m$. Note that we set the number of points in longitude $N_\varphi = 3N_m$ to prevent aliasing from the quadratic terms (Boyd 2001; Canuto *et al.* 2006).

In the radial direction, we use a Chebyshev collocation method over a grid of N_s points

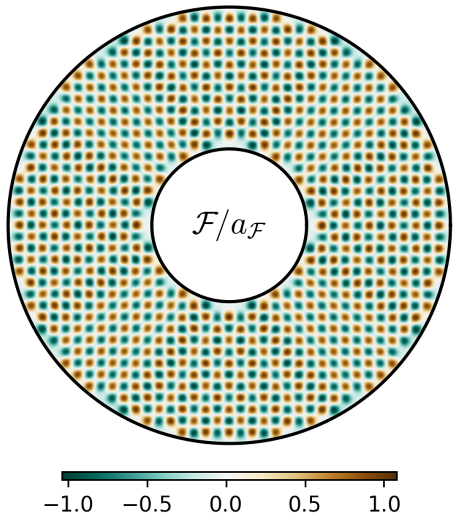

Figure 2: Cartesian forcing pattern $\mathcal{F}$ as described by Lemasquerier *et al.* (2023) producing inlet-outlet vorticity sources. Vortex spacing $\Delta_{\mathcal{F}} = 0.08$, radius $\ell_{\mathcal{F}} = 0.04$ and amplitude $a_{\mathcal{F}} = 2 \times 10^{10}$ are the same for all configurations studied. The magnitude of the pumps on this figure is normalised between 1 and -1 for cyclonic and anticyclonic vortices, respectively.

s_k; these points are the images of the Gauss-Lobatto points

$$x_k = \cos\left[\frac{\pi(k-1)}{N_s - 1}\right], \text{ with } 1 \leqslant k \leqslant N_s, \tag{2.11}$$

by the mapping

$$s_k = \frac{s_o - s_i}{2} x_k + \frac{s_o + s_i}{2}. \tag{2.12}$$

Accordingly, the Fourier coefficients f_m are expressed at each point s_k as

$$f_m(s_k) \approx \left(\frac{2}{N_s - 1}\right)^{1/2} \sum_{n=0}^{N_s-1}{}'' \hat{f}_{mn} T_n(x_k), \tag{2.13}$$

where the double prime summation indicates that both the first and last indices are multiplied by $1/2$ (Glatzmaier 1984). Here, the hat denotes Chebyshev coefficients and $T_n(x_k)$ is the first-kind Chebyshev polynomial of order n evaluated as x_k, defined by

$$T_n(x_k) \equiv \boldsymbol{T}_{kn} = \cos[n \arccos(x_k)]. \tag{2.14}$$

The set of equations (2.5), (2.6) and (2.7) are expanded in Fourier and Chebyshev series. Eqs. (2.6) and (2.5) yield the following set of coupled semi-discrete equations for $\hat{\omega}_{mn}$ and $\hat{\psi}_{mn}$ for any $m > 0$,

$$\sum_{n=0}^{N_s-1}{}'' \left[\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{T}_{kn} - \boldsymbol{O}_{kn}(m)\right] \hat{\omega}_{mn} = \mathcal{N}_\omega(m, s_k), \tag{2.15}$$

$$\sum_{n=0}^{N_s-1}{}'' \left[\boldsymbol{T}_{kn} \hat{\omega}_{mn} + \boldsymbol{P}_{kn}(m) \hat{\psi}_{mn}\right] = 0. \tag{2.16}$$

The real-valued collocation matrices entering Eqs. (2.15)-(2.16) are

$$\boldsymbol{O}_{kn}(m) = \boldsymbol{T}''_{kn} + \frac{1}{s_k}\boldsymbol{T}'_{kn} - \left(\frac{m^2}{s_k^2} + \Upsilon_k\right)\boldsymbol{T}_{kn}, \tag{2.17}$$

$$\boldsymbol{P}_{kn}(m) = \boldsymbol{T}''_{kn} + \left(\frac{1}{s_k} + \beta_k\right)\boldsymbol{T}'_{kn} + \left(\frac{\beta_k}{s_k} + \frac{\mathrm{d}\beta_k}{\mathrm{d}s} - \frac{m^2}{s_k^2}\right)\boldsymbol{T}_{kn}, \tag{2.18}$$

where $\boldsymbol{T}''_{kn}$ and $\boldsymbol{T}'_{kn}$ are the first and second derivative matrices of the n-th order Chebyshev polynomial at the collocation point x_k. The right-hand side term in Eq. (2.15) comprises the prescribed forcing, the Coriolis force, and the nonlinear term which is evaluated in physical space, then transformed back to spectral space:

$$\mathcal{N}_\omega(m, s_k) = \mathcal{F}(m, s_k) - \frac{2}{E}\beta_k u_s(m, s_k) - \boldsymbol{\nabla\cdot}\left(\frac{1}{N_\varphi}\sum_{j=0}^{N_\varphi}(\boldsymbol{u}\omega)\, e^{-\mathrm{i}m\varphi_j}\right). \tag{2.19}$$

Likewise, Eq. (2.7) for the mean azimuthal flow $\langle u_\varphi\rangle_\varphi$ is discretised as

$$\sum_{n=0}^{N_s-1}{}'' \left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{T}_{kn} - \boldsymbol{O}_{kn}(1)\right]\widehat{\langle u_\varphi\rangle_{\varphi}}_n = \mathcal{N}_{\langle u_\varphi\rangle_\varphi}(s_k), \tag{2.20}$$

where the non-linear term on the right-hand side contains the self interaction of the zonal wind (Aubert *et al.* 2003),

$$\mathcal{N}_{\langle u_\varphi\rangle_\varphi}(s_k) = -\frac{E}{2}\Upsilon_k u_{\varphi_0}\omega_0 - 2\sum_{m=1}^{N_m}\Re\left\{u_{s_m}\omega_m^*\right\}. \tag{2.21}$$

Boundary conditions are imposed on the so-called Tau lines, corresponding to their collocation grid boundaries. For the annulus, the first and last rows of the matrix include the constraint at $s = s_o$ and $s = s_i$, respectively. The Chebyshev coefficients $\hat{f}_{mn}$ of a function f subjected to homogeneous Dirichlet conditions must fulfill

$$\sum_{n=0}^{N_s-1}{}'' \hat{f}_{mn} = 0 \text{ at } s = s_o \text{ and } \sum_{n=0}^{N_s-1}{}''(-1)^n\hat{f}_{mn} = 0 \text{ at } s = s_i. \tag{2.22}$$

If f is instead subjected to homogeneous Neumann boundary conditions, its coefficients must satisfy

$$\sum_{n=0}^{N_s-1}{}'' n^2\hat{f}_{mn} = 0 \text{ at } s = s_o \text{ and } \sum_{n=0}^{N_s-1}{}''(-1)^{n+1}n^2\hat{f}_{mn} = 0 \text{ at } s = s_i. \tag{2.23}$$

The coupled system (2.15)-(2.16) gives rise to the following $(2N_s \times 2N_s)$ dense real matrix operator for each Fourier mode $m > 0$,

$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{T}_{kn} - \boldsymbol{O}_{kn} & 0 \\ \boldsymbol{T}_{kn} & \boldsymbol{P}_{kn} \end{bmatrix} \tag{2.24}$$

with Dirichlet and Neumann boundary conditions on ψ (2.9) imposed on the the first and last rows of the top-right and bottom-right quadrants of this matrix, respectively. The equation for the zonal flow $\langle u_\varphi\rangle_\varphi$ yields an additional $(N_s \times N_s)$ operator with Dirichlet boundary conditions (2.9) imposed on the first and last rows of this matrix.

The semi-discrete problem resulting from the Fourier-Chebyshev discretisation in space is advanced in time using the third-order BPR353 implicit-explicit (IMEX) scheme of Boscarino *et al.* (2013). This integrator was selected based on the comprehensive assessment of Gopinath *et al.* (2022), which evaluated 26 IMEX schemes in a related pseudo-spectral context. The implicit component is associated to the linear operator ($\mathcal{L}$, say) applied to the state variables, while the explicit one is related to the nonlinear terms $\mathcal{N}$.

3. Implicit subgrid-scale learning

To fully resolve the spatial scales in a certain configuration regime, we must run spectral simulations at DNS resolution, which in practice requires at least N_m truncated Fourier modes and N_s Chebyshev coefficients. We want to run "reduced" simulations that maintain this accuracy for coarser resolutions, i.e., $\overline{N_m} < N_m$ and/or $\overline{N_s} < N_s$. From now on, let us employ overbars to denote the unknown fields at coarser resolution. We know that without any correction, the reduced simulations will likely accumulate energy at the smallest resolved scale, or possibly lead to a numerical blow-up in the worst case scenario. To account for the missing energy transfers, corrections terms, or subgrid-scale (SGS) terms τ must be included in the reduced (or filtered) equations,

$$\frac{\partial \overline{\omega}}{\partial t} + \boldsymbol{\nabla} \cdot (\overline{\boldsymbol{u}}\,\overline{\omega}) = \frac{2}{E}\beta \overline{u_s} + \bar{\mathcal{F}} + \nabla^2 \overline{\omega} - \Upsilon \overline{\omega} + \tau_\omega, \tag{3.1}$$

$$\frac{\partial \langle \overline{u_\varphi} \rangle_\varphi}{\partial t} + \langle \overline{u_s}\,\overline{\omega} \rangle_\varphi = -\frac{\langle \overline{u_\varphi} \rangle_\varphi}{s^2} + \nabla^2 \langle \overline{u_\varphi} \rangle_\varphi - \Upsilon \langle \overline{u_\varphi} \rangle_\varphi + \tau_{\langle u_\varphi \rangle_\varphi}. \tag{3.2}$$

By construction, at the resolution of the DNS, the correction $\tau = [\tau_\omega, \tau_{\langle u_\varphi \rangle_\varphi}]$ vanishes and (3.1)-(3.2) is equivalent to the original set of equations (2.6)-(2.7). The purpose of this work is to use machine learning to learn a model $\mathcal{M}$ that produces an approximation of the SGS terms τ, given the knowledge of the reduced-resolution fields $\bar{f}$,

$$\mathcal{M}\left(\bar{f}\right) \approx \tau. \tag{3.3}$$

3.1. *Coarse-graining in Fourier-Chebyshev space*

Now, in order to generate a dataset to set-up the learning problem, we must design a composite transform $\mathcal{T}$ that maps direct-resolution fields to reduced-resolution fields such that

$$\mathcal{T}(f) = \left(\mathcal{T}_s \circ \mathcal{T}_\varphi\right)(f) \equiv \bar{f}. \tag{3.4}$$

In the azimuthal direction, it is sufficient to truncate the Fourier expansion to reduce the dimension of the grid while preserving the $[0, 2\pi)$ periodicity. The literature already contains many examples of machine learning models that have been trained on doubly-periodic (Maulik *et al.* 2019; Guan *et al.* 2022) or triply-periodic (Sirignano *et al.* 2020; Frezat *et al.* 2021) datasets. In addition to mode truncation –also called sharp spectral filter–, one can apply any filter that commutes with partial derivatives. Some popular examples include the Gaussian or the Box filter (Zhou *et al.* 2019). Here, we only retain $\overline{N_m} \leqslant N_m$ Fourier modes in the azimuthal direction,

$$\mathcal{T}_\varphi[f(s, \varphi_k)] = {\sum_{m=0}^{\overline{N_m}}}' \Re\left\{f_m(s)\, e^{\mathrm{i}m\varphi_k}\right\}. \tag{3.5}$$

In the radial direction, truncating the Chebyshev series to $\overline{N_s} \leqslant N_s$ polynomials is not sufficient, since the Chebyshev polynomials T_n are not boundary-preserving by construction. Instead, we truncate coefficients from a Galerkin basis $\chi_n(x_k)$ that satisfies the appropriate

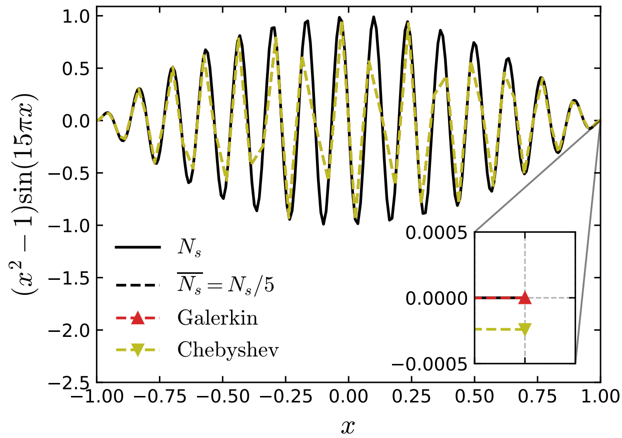

Figure 3: Illustration of the coarse-graining process for an analytical function $z(x) = (x^2 - 1)\sin(15\pi x)$ with $N_s = 300$ and $\overline{N_s} = 60$ using Chebyshev coefficients and using the boundary-preserving Dirichlet-Neumann Galerkin basis. The zoom inset clearly demonstrates that the Chebyshev truncation does not preserve the boundary conditions $z(\pm 1) = 0$.

boundary conditions,

$$\mathcal{T}_s[f_m(s_k)] = \sum_{n=0}^{\overline{N_s}-N_\chi} f^{\mathcal{G}}_{mn}\,\chi_n(x_k) \tag{3.6}$$

where $f^{\mathcal{G}}_{mn}$ are the Galerkin coefficients and N_χ is the number of boundary conditions fulfilled by the basis. In practice, the zonal flow $\langle\overline{u_\varphi}\rangle_\varphi$ is expanded onto the following Galerkin basis that satisfies Dirichlet boundary conditions,

$$\chi_n(x_k) = T_{n+2}(x_k) - T_n(x_k), \tag{3.7}$$
$$\chi(\pm 1) = 0, \tag{3.8}$$
$$N_\chi = 2. \tag{3.9}$$

This decomposition is not the only one that matches Dirichlet boundary conditions on both sides but it appears to be the most commonly employed (Shen 1995). For $\overline{\psi}$, we adopt the following Galerkin basis, which satisfies both Dirichlet and Neumann boundary conditions (e.g. Julien & Watson 2009),

$$\chi_n(x_k) = T_n(x_k) - \frac{2(n+2)}{n+3}T_{n+2}(x_k) + \frac{n+1}{n+3}T_{n+4}(x_k), \tag{3.10}$$
$$\chi(\pm 1) = \chi'(\pm 1) = 0, \tag{3.11}$$
$$N_\chi = 4. \tag{3.12}$$

To enforce the incompressibility constraint, we only coarse-grain the streamfunction $\overline{\psi}$ and the zonal velocity $\langle\overline{u_\varphi}\rangle_\varphi$ and then compute the corresponding reduced velocity $\overline{\boldsymbol{u}}$ using (2.3)-(2.4) and vorticity $\overline{\omega}$ using (2.5). We illustrate the coarse-graining process using Galerkin and Chebyshev coefficients truncation for an analytical function in Figure 3. We can observe in the bottom-right inset that, unlike the Chebyshev basis, the Galerkin basis preserves the boundary conditions $z(\pm 1) = 0$ during the coarse-graining process.

3.2. *Implicit subgrid terms using online learning*

From a machine learning point-of-view, the subgrid-scale problem is most of the time formulated as a supervised learning problem,

$$\arg\min_{\theta} L\left[\tau, \mathcal{M}(\overline{f};\theta)\right] \tag{3.13}$$

where θ are the trainable parameters of model $\mathcal{M}$, minimising a given loss function L. With this so-called "offline" approach (Sanderse *et al.* 2025), we need to build a dataset $\mathbb{D}$ of tuples (input, target) such that $\mathbb{D} := \{\overline{f}\} \to \{\tau\}$. The SGS term τ comes from the coarse-graining of the original equations (2.6)-(2.7). Within a learning framework, we aim to approximate the following SGS terms,

$$\tau = \begin{bmatrix} \tau_\omega \\ \tau_{\langle u_\varphi \rangle_\varphi} \end{bmatrix} = \begin{bmatrix} \nabla \cdot [\mathcal{T}(\boldsymbol{u})\mathcal{T}(\omega) - \mathcal{T}(\boldsymbol{u}\omega)] \\ \langle \mathcal{T}(u_s)\mathcal{T}(\omega) - \mathcal{T}(u_s\omega)\rangle_\varphi \end{bmatrix} \approx \mathcal{M}(\overline{f};\theta). \tag{3.14}$$

This definition, however, is only valid under the assumption that $\mathcal{T}$ commutes with partial derivatives (Sagaut 2006, chapter 2.2). For the annular geometry, the Gauss-Lobatto grid in the radial direction is not homogeneous and is thus responsible for the non-commutativity of the filtering and differentiation operators. This introduces a commutation term in the governing equations (Ghosal & Moin 1995). In practice, this term is however often ignored leading to a commutation error that is not well understood, in particular during the coarse-graining process (see the investigation by Yalla *et al.* 2021). Recently, Jakhar *et al.* (2024) trained a model for Rayleigh-Bénard convection based on a similar Fourier-Chebyshev spatial discretisation; they applied filtering in the periodic direction only, thereby avoiding the commutation error. With regard to the QG turbulence problem of interest here, let us briefly state that experiments based on the "offline" approach were systematically unsuccessful, in the sense that the trained SGS models always led to unstable simulations. We will consequently not discuss the "offline" approach further. The main objective of model $\mathcal{M}$ is to lead to an accurate evolution of the coarse variables $\overline{f}(t)$ with respect to that of the direct variables $f(t)$. A more realistic supervised learning problem would write,

$$\arg\min_{\theta} L\left[f(t), \overline{f}(t)\right] \tag{3.15}$$

where $\overline{f}(t)$ is bound to the subgrid model $\mathcal{M}$ by the flow operator γ_θ,

$$\overline{f}(t) = \gamma_\theta(\overline{f}, t_0, t) = \overline{f}(t_0) + \int_{t_0}^{t} \left[\mathcal{L}\overline{f}\,(t') + \mathcal{N}\left(\overline{f}, t'\right) + \mathcal{M}\left(\overline{f}(t');\theta\right)\right] \mathrm{d}t' \tag{3.16}$$

that advances the system (3.1)-(3.2) from time t_0 to time t, while optimising for the parameters θ such that the loss function L in (3.15) is minimised. In its discrete form, the evolution operator $\gamma_\theta(\overline{f}, t_0, t)$ corresponds to solving multiple sub-stages of the IMEX integrator. As the model $\mathcal{M}$ contains nonlinearities, it is treated explicitly as an additive source term, which results in practice in 3 calls per time step within the BPR353 IMEX scheme. It is clear that the flow operator (3.16) in the supervised learning problem (3.15) does not explicitly requires knowledge of the subgrid term τ. This "online" formulation (3.15) is often pointed out as more stable (Frezat *et al.* 2022; Sanderse *et al.* 2025) compared to the "offline" formulation (3.13). In addition, we propose here that the model $\mathcal{M}$ can implicitly learn to correct other error sources such as, e.g., the commutation error due to the radial coarse-graining. This formulation also allows some flexibility in prescription of the loss function L. Here, we define a kinetic energy mismatch metric,

$$L\left[f(t), \overline{f}(t)\right] = \pi \sum_{m=0}^{\overline{N_m}}{}' \int_{s_i}^{s_o} \left(|\overline{u_{s_m}}(t) - u_{s_m}(t)|^2 + |\overline{u_{\varphi_m}}(t) - u_{\varphi_m}(t)|^2 \right) s \, \mathrm{d}s. \tag{3.17}$$

Numerically, solving (3.15) is performed using a flavor of stochastic gradient descent and thus requires the gradient of the loss function $\boldsymbol{\nabla}_\theta L$. Expanding using the flow operator (3.16) gives,

$$\frac{\partial L}{\partial \theta} = \left(\frac{\partial L}{\partial \gamma_\theta}\right)^{\mathsf{T}} \left[\int_{t_0}^{t} \left(\frac{\partial \mathcal{L}\overline{f}(t')}{\partial \theta} + \frac{\partial \mathcal{N}(\overline{f}, t')}{\partial \theta} + \frac{\partial \mathcal{M}(\overline{f}(t'); \theta)}{\partial \theta} \right) \mathrm{d}t' \right]. \tag{3.18}$$

It clearly appears that optimising for the parameters θ using the "online" formulation requires the adjoint of the model $\mathcal{M}$ and the implicit $\mathcal{L}$ and explicit $\mathcal{N}$ forward solver terms. There exists many strategies to obtain these quantities, either exact or approximate (Ouala *et al.* 2024), which are not without drawbacks. Obtaining the exact adjoint can be done either by an explicit derivation, or by rewriting the solver using a differentiable framework, which can represent a substantial undertaking. On the other hand, approximate approaches are often computationally prohibitive, and can introduce sources of error in the optimisation problem (Frezat *et al.* 2023). In our case, the numerical method is simple enough for us to implement a solver from scratch using the JAX library (Bradbury *et al.* 2018). The library algorithmically provides us with exact derivatives using automatic differentiation (Baydin *et al.* 2018).

3.3. *Learning setup*

In a classical "offline" setup, data is generated such that samples are not correlated. For the "online" setup, we need continuous sub-trajectories from the DNS, spaced at time intervals matching those of the reduced resolution. In practice, it is easier to work with constant time steps. Let us denote stable direct-resolution and reduced-resolution time steps δt and $\overline{\delta t}$, respectively. The reduced simulation is defined on a coarser grid, and thus we have $\overline{\delta t} > \delta t$. Moreover, for simplicity, we assume that $\overline{\delta t}$ is related to δt by an integer factor n_{coarse} such that $\overline{\delta t} = n_{\mathrm{coarse}} \delta t$. Finally, we want to generate a dataset mapping initial sub-trajectory conditions $\overline{f}(t_0)$ to high-resolution (or coarse-grained high-resolution) variables that align with the reduced-resolution time step $\overline{\delta t}$,

$$\mathbb{D} \coloneqq \{\overline{f}(t_0)\} \to \{f(t_0 + s\overline{\delta t})\}_{s \in [1, N_{\mathrm{steps}}]} \tag{3.19}$$

where N_{steps} is the number of discrete time steps required to integrate from t_0 to t in (3.16). The value of N_{steps} is often limited by the available memory of the device, since automatic differentiation keeps track of the gradient graph built from any operation applied in the loss function L. Although it is possible to increase this limit with checkpointing (Sapienza *et al.* 2024), the width of the time interval over which adjoint optimisation is performed is constrained by the intrinsic limit of predictability of the system due to turbulence (see the discussion on training timescales in §4.1 below). For consistency, the size of the dataset is determined by the dynamics of a given configuration, defined in this study by the geometry of the container and a value of the Ekman number E. In practice, we build datasets that span one eddy turnover time, for each of the three configurations that will be detailed in the next section. Models are trained using the AdamW algorithm (Loshchilov & Hutter 2019) with a fixed learning rate of 2×10^{-5}.

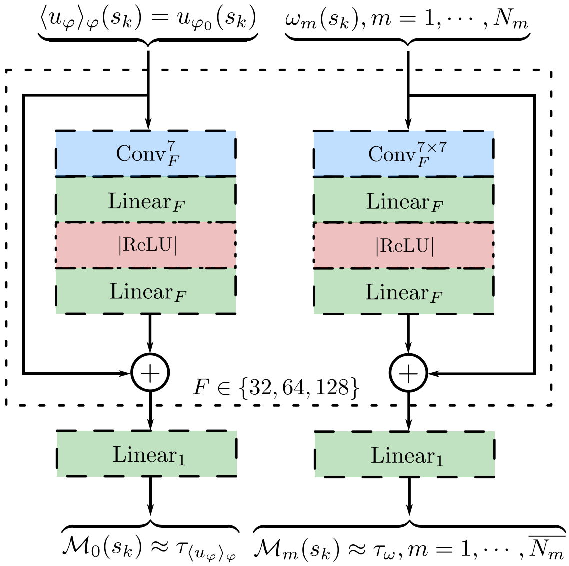

Figure 4: Illustration of the spectral-space architecture. The architecture is split into two paths for the axisymmetric velocity and the vorticity equations. The model applies three (residual) blocks of increasing number of features $F \in \{32, 64, 128\}$ containing each a Conv layer, a Linear layer and a non-linear mod ReLU, or $|\mathrm{ReLU}| = \mathrm{ReLU}(z + b)z/|z|$ activation (Arjovsky *et al.* 2016) followed by a Linear layer. The last Linear layer projects the high-dimensional features F into a single feature. The total number of trainable parameters for this architecture is approximately 666k.

3.4. *Spectral-space architecture*

Model $\mathcal{M}$ should operate in spectral space, since it is meant to integrate smoothly in an existing pseudo-spectral solver. Operating in spectral space avoids potential back and forth transforms between spectral space and grid space, but demands complex-valued neural networks (Hirose 2006; Trabelsi *et al.* 2018). In addition, model $\mathcal{M}$ should have an inference time that does not exceed that of the calculation of the linear and non-linear terms of (3.1)-(3.2). We take inspiration from modern convolutional blocks, also called ConvNext, which have been shown to compete favorably with Transformers in terms of accuracy and scalability (Liu *et al.* 2022). We propose a low complexity variant of the described ConvNext architectures in order to reduce the computational cost at inference. The complexity of this model, i.e., size of kernels, number of layers and features has been selected by trial and error. The overall model $\mathcal{M}$ is also divided into two paths that produce subgrid corrections to (3.1) and (3.2), respectively. An illustration with the inner details of the spectral-space architecture is shown in Figure 4. The left path is applied to $\langle u_\varphi \rangle_\varphi$, namely the axisymmmetric $m = 0$ mode of the azimuthal velocity, while the right path is applied to ω for all non-axisymmetric modes $m > 0$. In practice, the left path takes a one-dimensional input of size $\overline{N_s}$, while the right path takes a two-dimensional input of size $[\overline{N_m} - 1, \overline{N_s}]$. Internally, each trainable layer is real-valued operating on complex-valued input, with backpropagation supported by Wirtinger derivation rules (e.g. Hjørungnes & Gesbert 2007), i.e., real and imaginary parts are treated separately.

	Geometry	E	$a_{\mathcal{F}}$	δt	t_L	λ	dim($\mathbb{D}$)
(i)	Exp. ($\beta = -1/s_o$)	2×10^{-7}	2×10^{10}	4×10^{-8}	1.70×10^{-4}	2.1×10^4	850
(ii)	Spherical	3×10^{-7}	2×10^{10}	4×10^{-8}	1.55×10^{-4}	3.1×10^4	775
(iii)	Spherical	1×10^{-6}	2×10^{10}	4×10^{-8}	1.30×10^{-4}	3.2×10^4	650

Table 2: Numerical model parameters: Ekman number E, forcing amplitude $a_{\mathcal{F}}$ and steady-state fixed time step δt. Note that a smaller time step δt might be required to ensure stability during the transient. The turnover time t_L and the growth rate λ provide us with statistical information about the dynamics of the simulation. These time-related statistics are used to estimate the size of the dataset dim($\mathbb{D}$).

4. Results

To evaluate the capabilities of the learned subgrid models, we explore three different configurations that yield a turbulent quasi-geostrophic regime. Numerically, the maximum grid resolution that can fit in GPU memory is limited by the imprint of the dense collocation matrices (2.16); consequently, we use $(N_m, N_s) = (640, 321)$ for the DNS of each configuration. For the coarse-grained resolution, we truncate our original grid to $(\overline{N_m}, \overline{N_s}) = (128, 65)$, which corresponds to a 5-fold ratio compared to the DNS grid. At this resolution, the Ekman layers are barely resolved since they only contain approximately 5 grid points (see Shishkina *et al.* 2010). The subgrid-scale problem becomes significantly more difficult with a 10-fold ratio grid $(\overline{N_m}, \overline{N_s}) = (64, 33)$; in that case, the Ekman layers are definitively not resolved, and we did not explore the 10-times coarser resolutions further. For each physical configuration, the transient regime is only computed using DNS. Dataset generation and evaluation start when a statistically steady state has been reached, i.e., when the rms velocity fluctuations settle to a uniform constant value. The three configurations share a common forcing amplitude $a_{\mathcal{F}} = 2 \times 10^{10}$. To investigate the influence of rotation and shape of the three-dimensional container on the turbulent flow, we vary both the Ekman number and the $\beta(s)$ profile.

In the remainder of this study, we will refer to the three setups summarised in Table 2 and illustrated in Figure 5 using the following convention:

(i) **Exponential container with $\beta = -1/s_o$ and $E = 2 \times 10^{-7}$** (Figure 5, top row). Due to the dominant role of rotation, we observe the formation of zonal jets of alternated directions, which yield concentric rings of vorticity. As β is constant, a similar dynamics is observed throughout the domain. Fluctuations about the mean flow have a magnitude comparable to that of zonal jets. This configuration leads to multistable solutions, as the number and position of jets vary across different realisations (as in e.g. Fig. 6 of Lemasquerier *et al.* 2023).

(ii) **Spherical container with $E = 3 \times 10^{-7}$** (Figure 5, middle row). This configuration exhibits two dynamical regions. In the inner part of the annulus, $|\beta|$ is minimum (Table 1), and the dynamics is dominated by turbulence with a predominance of vortices, while the external part of the annulus shows spiralling structures typical of Rossby waves. This segregated dynamics is typical of the flow observed in spherical containers, in which $|\beta|$ increases quickly outwards (see, e.g. Fig. 5 in Barrois *et al.* 2022). Segregation follows from to the local competition between advection-dominated processes, whose timescale is the local turnover time, and propagation-dominated processes, whose timescale is the Rossby wave period. The transition between the two regions occurs when those two timescales are comparable (Guervilly & Cardin 2017).

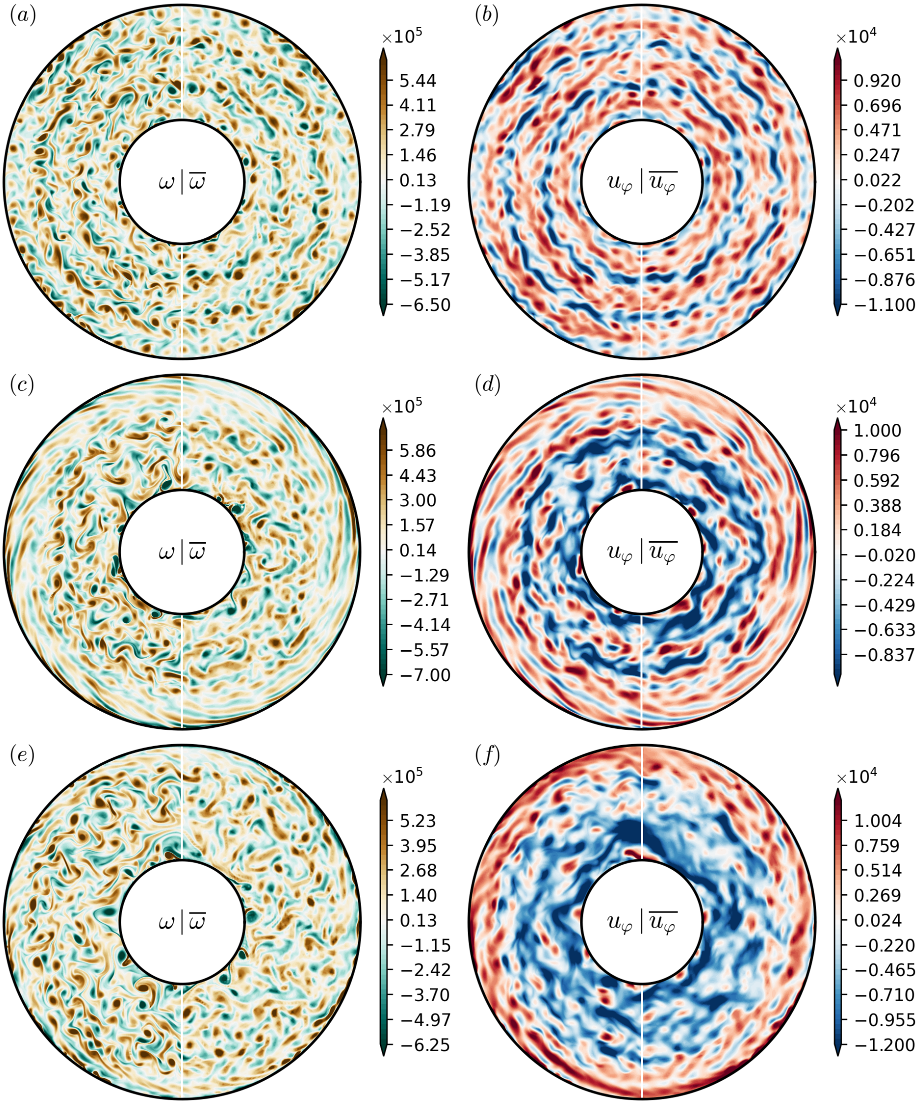

Figure 5: Vorticity ω (left) and azimuthal velocity u_φ (right) from DNS for the three configurations. Top: configuration (i), exponential container with $E = 2 \times 10^{-7}$ and $\beta = -1/s_o$. Middle: configuration (ii), spherical shell with $E = 3 \times 10^{-7}$. Bottom: configuration (iii), spherical shell with $E = 10^{-6}$. In each panel, the left and right halves represent direct-resolution and coarse-grained fields, respectively. See text for details.

(iii) **Spherical container with** $E = 10^{-6}$ (Figure 5, bottom row). Configuration (iii) is configuration (ii) with an increased Ekman number. Advective processes therefore prevail in most of the fluid volume, confining Rossby waves to a thin fluid region close to the outer boundary. The pair of prograde and retrograde jets are stronger in magnitude compared with that of configuration (ii).

We adopt the following averaging definitions to compute diagnostics and evaluation metrics,

$$\langle f\rangle_\varphi = \frac{1}{2\pi}\int_0^{2\pi} f\,\mathrm{d}\varphi, \quad \langle f\rangle_s = \frac{2}{s_o^2 - s_i^2}\int_{s_i}^{s_o} f\,s\,\mathrm{d}s, \quad \langle f\rangle = \langle f\rangle_{\varphi,s} \tag{4.1}$$

where $\langle\cdot\rangle_\varphi$ corresponds to azimuthal averaging, $\langle\cdot\rangle_s$ to radial averaging and $\langle\cdot\rangle$ to domain (surface) average, i.e., in both azimuthal and radial directions.

The total surface kinetic energy density E_T can be decomposed into a zonal component E_Z for the axisymmetric flow and a residual component E_R,

$$E_T = \frac{1}{2}\langle \boldsymbol{u}^2\rangle = E_Z + E_R = \frac{1}{2}\langle u_{\varphi_0}^2\rangle_s + \sum_{m=1}^{N_m}\langle u_{s_m}^2\rangle_s + \langle u_{\varphi_m}^2\rangle_s\,. \tag{4.2}$$

The Reynolds and Rossby numbers are accordingly defined by

$$Re = \sqrt{2E_T}, \quad Ro = ReE. \tag{4.3}$$

We are also interested in quantities that are more sensitive to the small scales of the flow, closer to where subgrid-scale transfers occur, such as the total enstrophy,

$$\mathcal{Z} = \frac{1}{2}\langle(\boldsymbol{\nabla}\times\boldsymbol{u})^2\rangle = \frac{1}{2}\langle\omega^2\rangle. \tag{4.4}$$

From the ratio of kinetic energy and enstrophy, one introduces the viscous dissipation lengthscale

$$\ell_\nu = \sqrt{\frac{E_T}{\mathcal{Z}}}. \tag{4.5}$$

The numerical simulations are also evaluated in terms of their power balance, which is obtained by the inner product of the Navier-Stokes equations with $\boldsymbol{u}$, followed by spatial averaging. In our setups, turbulence is driven by the power input

$$\mathcal{P}_\mathcal{F} = \langle \boldsymbol{f}\cdot\boldsymbol{u}\rangle, \tag{4.6}$$

where $\boldsymbol{f}$ stands for the driving force that sustains the forcing pattern shown in Figure 2. An analytical expression of $\mathcal{P}_\mathcal{F}$ is derived in Appendix B. Kinetic energy is dissipated both by Ekman friction and bulk viscous dissipation. Even if we do not have an analytical expression for the power loss by Ekman friction, denoted by $\mathcal{D}_\Upsilon$, we can use the time average energy budget to retrieve its time average value. The instantaneous power budget indeed reads

$$\frac{\mathrm{d}\langle E_T\rangle}{\mathrm{d}t} = \mathcal{P}_\mathcal{F} - \langle\omega^2\rangle - \mathcal{D}_\Upsilon, \tag{4.7}$$

where we understand that the second-term on the right-hand side corresponds to the bulk viscous dissipation. Upon time averaging, denoted by $\langle\cdot\rangle_t$, we get

$$0 = \langle\mathcal{P}_\mathcal{F}\rangle_t - \langle\omega^2\rangle_t - \langle\mathcal{D}_\Upsilon\rangle_t, \tag{4.8}$$

which allows us to characterise the relative fraction of energy dissipated through Ekman friction,

$$\alpha_\nu^\Upsilon \equiv \frac{\langle\mathcal{D}_\Upsilon\rangle_t}{\langle\mathcal{D}_\Upsilon\rangle_t + \langle\omega^2\rangle_t}. \tag{4.9}$$

4.1. *Training timescales*

Our goal in this study is not to evaluate the ability of a given model to generalise over different initial conditions, parameters, or grid resolutions. Instead, we explore the possibility

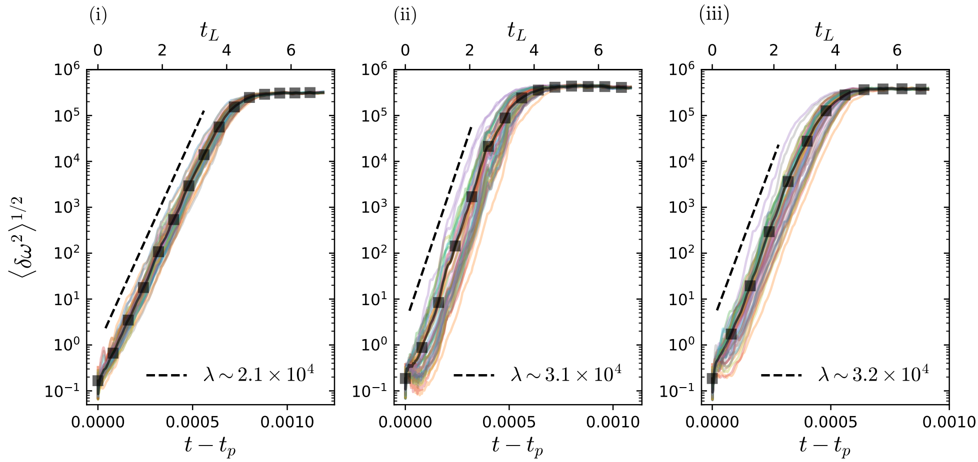

Figure 6: Evolution of the deviation (based on the root mean squared vorticity error) for ensembles of perturbed simulations for the configurations (i), (ii) and (iii). The perturbation is set at $t = t_p$ as a single randomly localised Gaussian source similar to those appearing in (2.8), activated with an amplitude $10^{-10} a_{\mathcal{F}}$. In each panel, the dashed line shows exponential growth for a growth rate λ obtained by least-squares fit.

of training a model to extend the initial integration of a DNS at reduced numerical cost over a long temporal horizon. As exposed in §3.3, we build datasets from the DNS that span one turnover time $t_L = 1/Re$ for each configuration and proceed with reduced simulations for 100 turnover times. The number of DNS samples used for online learning is thus $\dim(\mathbb{D}) \approx t_L/\overline{\delta t}$, where the coarse time step size $\overline{\delta t}$ is adjusted considering grid ratios between the direct and coarse-grained resolutions. Here, since the time step is limited by the azimuthal part of the advection term, we adopt $\overline{\delta t} = 5\delta t$, due to the five-fold coarsening of the azimuthal grid. Table 2 indicates that the three configurations have similar values for t_L, configuration (iii) being the most turbulent; it has the largest Ekman number, while all configurations share the same forcing amplitude $a_{\mathcal{F}}$. The size of the datasets ranges from 675 samples for configuration (iii) to 850 samples for the least turbulent configuration (i). The number of consecutive samples N_{steps} used for the online learning strategy is set by a practical constraint: accumulating gradients in the loss computation (3.18) takes up GPU memory, which, in practice, limits us to $N_{\text{steps}} = 25$ time steps separated by a constant $\overline{\delta t}$. In any event, since we study deterministically chaotic dynamics, we must check that this hardware-imposed training timescale is small compared with the decorrelation time of each configuration, that is governed by the nonlinear flow operator in (3.16). In order to quantify the divergence of the three configurations over a sub-trajectory timespan, we consider for each configuration a reference vorticity ω_{ref} which has reached statistical equilibrium, and that we perturb by a tiny amount at some arbitrary time t_p. If ω_p refers to the subsequent evolution of the perturbed solution, the growth of the difference $\delta\omega = \omega_p - \omega_{\text{ref}}$ is initially exponential, with a growth rate λ such that

$$\langle \delta\omega^2 \rangle^{1/2} \propto \exp \lambda (t - t_p). \tag{4.10}$$

For each configuration, we define an ensemble of perturbations at t_p by activating, for each ensemble member, a random pump in the collection given in (2.8), at a modest amplitude of $10^{-10} a_{\mathcal{F}} = 2$. We show in Figure 6 the evolution of the norm of the difference between the reference solution and the ensemble of perturbed trajectories. On each panel, a phase of exponential growth if followed by a saturated phase. Dashed lines show exponential

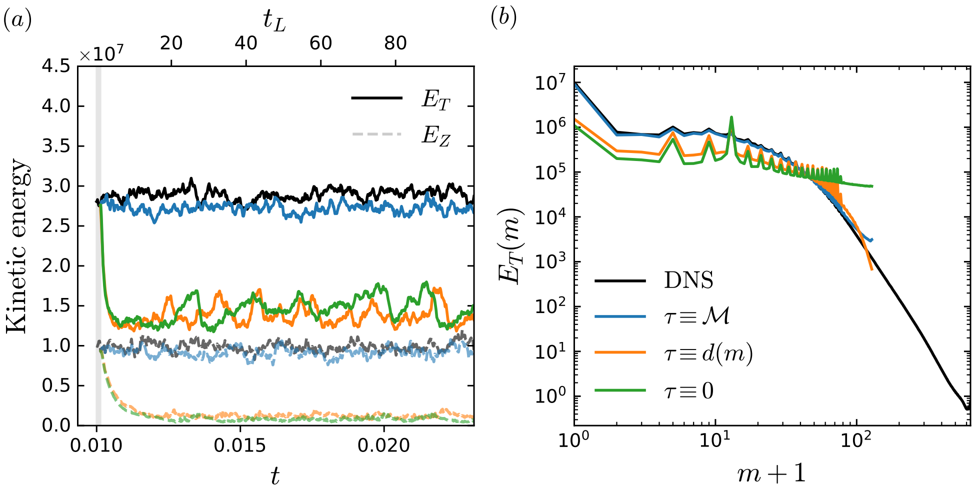

Figure 7: (*a*) Evolution of the total E_T (solid lines) and zonal E_Z (dashed lines) kinetic energy and (*b*) time-averaged energy spectra as a function of the wavenumber m (*b*) for configuration (iii). The highlighted region at the beginning of the kinetic energy evolution corresponds to the training dataset time span, which is equivalent to one turnover time. Simulations with the different models were started at this point and integrated for for 100 turnover time t_L.

growths for values of λ obtained by least-squares, which are listed in Table 2 and correspond to timescales λ^{-1} in the range $0.2 - 0.3\ t_L$. Figure 6 shows that each configuration is completely decorrelated at $t - t_p \approx 4t_L \approx 5 \times 10^{-4}$, which is around 100 times larger than the sub-trajectory timespan for learning of $25\,\overline{\delta t} = 5 \times 10^{-6}$. In other words, the timespan chosen for online learning represents $\mathcal{O}(1)$% of the physical decorrelation time for all three configurations, which ensures that the gradient-based optimisation is meaningful.

4.2. *Alternative models*

We henceforth denote $\tau \equiv \mathcal{M}$ the model trained with the implicit "online" scheme described above and $\tau \equiv 0$ the absence of any correction, i.e., a simulation at reduced resolution. We assess the performance of $\tau \equiv \mathcal{M}$ against a standard hyperdiffusive model, denoted as $\tau \equiv d(m)$. In practice here, hyperdiffusion only acts along the azimuthal direction such that the viscous terms in (3.1)-(3.2) are multiplied in spectral space by a factor

$$d(m) = a_d^{m-m_d},\ m_d \leqslant m < \overline{N_m}, \tag{4.11}$$

where a_d is the hyperdiffusivity amplitude and m_d is the wavenumber beyond which hyperdiffusivity operates (Nataf & Schaeffer 2015). Following the prescriptions by Davy *et al.* (2023), m_d is set to three times the wavenumber of the peak of the radial velocity spectrum $|u_{s_m}|^2$. This corresponds to $m_d \approx 36$ for configurations (ii-iii) and $m_d \approx 96$ for configuration (i). The amplitude a_d is calibrated by trial and error such that a_d suppresses energy pile-up at small-scales, with values ranging from $a_d = 1.04$ for configurations (ii-iii) to $a_d = 1.12$ for configuration (i).

The impact of a chosen model on the energetic properties of a solution is illustrated in Figure 7 for configuration (iii). Figure 7(*a*) shows the time series of E_T and E_Z for the DNS, $\tau \equiv \mathcal{M}$, $\tau \equiv d(m)$, and $\tau \equiv 0$. Once the steady state has been reached, the energy contained in the zonal jets corresponds to about 30% of the kinetic energy content. We observe that the under-resolved simulation $\tau \equiv 0$ and the hyperdiffusivity model $\tau \equiv d(m)$ lose up to 50%

(i)

	$\langle Re\rangle_t$	$\langle E_Z/E_T\rangle_t$	$\langle \mathcal{Z}\rangle_t$	$\langle \ell_\nu\rangle_t$	$\langle \mathcal{P}_\mathcal{F}\rangle_t$	α_ν^Υ
DNS	5887	0.273	3.82×10^{10}	2.13×10^{-2}	2.34×10^{11}	0.674
$\tau\equiv\mathcal{M}$	**5772**	**0.270**	$\mathbf{3.60\times10^{10}}$	$\mathbf{2.15\times10^{-2}}$	$\mathbf{2.35\times10^{11}}$	**0.694**
$\tau\equiv d(m)$	4176	0.046	5.06×10^{10}	1.31×10^{-2}	2.32×10^{11}	0.563
$\tau\equiv 0$	4198	0.034	6.93×10^{10}	1.13×10^{-2}	2.22×10^{11}	0.375

(ii)

	$\langle Re\rangle_t$	$\langle E_Z/E_T\rangle_t$	$\langle \mathcal{Z}\rangle_t$	$\langle \ell_\nu\rangle_t$	$\langle \mathcal{P}_\mathcal{F}\rangle_t$	α_ν^Υ
DNS	6495	0.243	5.05×10^{10}	2.05×10^{-2}	1.61×10^{11}	0.374
$\tau\equiv\mathcal{M}$	**6339**	**0.239**	$\mathbf{4.73\times10^{10}}$	$\mathbf{2.07\times10^{-2}}$	$\mathbf{1.60\times10^{11}}$	**0.410**
$\tau\equiv d(m)$	4479	0.125	4.18×10^{10}	1.55×10^{-2}	1.88×10^{11}	0.555
$\tau\equiv 0$	4251	0.052	7.40×10^{10}	1.10×10^{-2}	1.72×10^{11}	0.142

(iii)

	$\langle Re\rangle_t$	$\langle E_Z/E_T\rangle_t$	$\langle \mathcal{Z}\rangle_t$	$\langle \ell_\nu\rangle_t$	$\langle \mathcal{P}_\mathcal{F}\rangle_t$	α_ν^Υ
DNS	7603	0.342	4.21×10^{10}	2.62×10^{-2}	1.63×10^{11}	0.485
$\tau\equiv\mathcal{M}$	**7376**	**0.337**	$\mathbf{4.02\times10^{10}}$	$\mathbf{2.60\times10^{-2}}$	$\mathbf{1.66\times10^{11}}$	**0.516**
$\tau\equiv d(m)$	5267	0.108	5.05×10^{10}	1.66×10^{-2}	2.68×10^{11}	0.623
$\tau\equiv 0$	5390	0.074	1.04×10^{11}	1.18×10^{-2}	2.46×10^{11}	0.155

Table 3: Statistical evaluation of the proposed models for each configuration explored in this study. Reynolds number Re, zonal energy ratio E_Z/E_T, enstrophy $\mathcal{Z}$, dissipation lengthscale ℓ_ν, injected power $\mathcal{P}_\mathcal{F}$ and ratio of dissipative processes α_ν^Υ are time-averaged (denoted $\langle\cdot\rangle_t$) over 100 turnover times.

of kinetic energy compared to the DNS in a few turnover times t_L. This loss can be mostly ascribed to the loss of mean zonal energy. Figure 7(b) shows the corresponding time-averaged kinetic energy spectra as a function of azimuthal order m. The under-resolved simulation with $\tau\equiv 0$ presents an accumulation of kinetic energy at the tail of the spectrum. This accumulation is not observed with the hyperdiffusivity model $\tau\equiv d(m)$, which was precisely designed to prevent such pile-up. Both $\tau\equiv 0$ and $\tau\equiv d(m)$ models underestimate the zonal energy ($m=0$) by almost an order of magnitude, but also, to a lesser extent, the energy contained in the intermediate scales $m\sim 1-30$. In contrast, the implicit "online" model $\tau\equiv\mathcal{M}$ matches the kinetic energy spectra of the DNS up to $\overline{N_m}$, in spite of underestimating the total energy content by about 5%, as can be seen in the time series of E_T in Figure 7(a).

4.3. *Integrated quantities*

For each configuration, Table 3 shows the time average of the Reynolds number, zonal to total energy ratio, enstrophy, dissipation lengthscale, injected power, and the fraction of dissipation due to Ekman friction. Time averaging $\langle\cdot\rangle_t$ is performed over 100 turnover times, for the direct, $\tau\equiv\mathcal{M}$, $\tau\equiv d(m)$, and $\tau\equiv 0$ numerical simulations. We observe a 20 − 30% drop of the Reynolds number $\langle Re\rangle_t$ for all configurations with respect to the DNS, except for $\tau\equiv\mathcal{M}$ that remains within 5% of the reference value. The zonal energy ratio E_Z/E_T is not well reproduced by the under-resolved and hyperdiffusive simulations. This indicates that most of the zonal energy is lost in these models, likely because they do not properly account for the subgrid transfers involved in the production of the axisymmetric velocity

$\langle u_\varphi \rangle_\varphi$. This effect is particularly dramatic in the exponential container where most of the zonal energy is lost. The accumulation of energy at small scales, is well reflected in the total enstrophy $\mathcal{Z}$, which is always larger in the under-resolved simulations compared to their DNS counterparts. The models with hyperdiffusivity $\tau \equiv d(m)$ perform better on this metric, since they are effectively dissipating energy at small-scale (recall Fig. 7*b*), but they still fail to capture backward energy injection and as such severely underestimate the zonal energy content. Finally, the $\tau \equiv \mathcal{M}$ models are the only ones that successfully match the dissipation lengthscales of the DNS, which indicates a correct balance of energy and enstrophy production. Regarding the fraction of energy dissipated through Ekman friction, α_ν^Υ, let us stress again that it is exactly computed via the balance (4.9), as we have numerical control over bulk viscous dissipation and injected power. Lemasquerier *et al.* (2023) could not precisely measure the latter in their experiments; they produced estimates of the input power based either on the spectral distribution of the residual kinetic energy or on the assumption that energy was entirely dissipated by Ekman friction, i.e. $\alpha_\nu^\Upsilon = 1$ in our parlance. The relationship between these two estimates showed deviation from pure proportionality (see their Fig. 8(*a*)), which suggests that α_ν^Υ was markedly below unity for some experiments. Here we find that it is about 67% for configuration (i), 37% for configuration (ii) and 48% for configuration (iii), see Table 3. The friction ratio is in all cases reproduced within 5% by the learned model $\tau \equiv \mathcal{M}$. The energy budget, and hence α_ν^Υ, is significantly altered for the under-resolved simulations and hyperdiffusivity models. For the spherical configurations (ii) and (iii), $\tau \equiv d(m)$ models dissipate more through Ekman friction compared to the reference DNS, while the $\tau \equiv 0$ simulations dissipate over 80% of their energy through bulk viscous dissipation.

4.4. *Zonal structures*

A hallmark of β-plane turbulence is the formation of zonal jets. The width of these jets is often found to be closely related to the so-called Rhines scale (Rhines 1975), here defined using the Rossby number, such that

$$\ell_R = 2\pi \sqrt{\frac{Ro}{|\beta|}}. \tag{4.12}$$

In the exponential container considered in configuration (i), β is constant which makes the definition of ℓ_R well-posed. Figure. 8(*a*) shows the time-averaged radial profiles of the zonal velocity $\langle u_\varphi \rangle_\varphi$ for this configuration. Seven jets of alternated directions of similar width are obtained. The segment in the bottom right corner indicates the good agreement between the jet widths and half the Rhines scale (e.g. Heimpel & Aurnou 2007). The small standard deviations –shaded area– highlight the stability of the jets over time in this configuration. The simulation with $\tau \equiv \mathcal{M}$ accurately reproduces the jets pattern of the DNS, in stark contrast to the other two models which lose the structure of the jets, mostly at the center of the domain.

Zonal jets in quasi-geostrophic flows are directly related to the distribution of potential vorticity (PV) (McIntyre 2003) here defined by

$$q = \frac{\langle \omega \rangle_\varphi + 2E^{-1}}{h}. \tag{4.13}$$

Regions of weak PV gradients hence correspond to retrograde zonal jets, while prograde jets coincide with regions of strong PV gradients. In the limit of quasi-geostrophic turbulence with $Re \gg 1$ and $Ro \ll 1$, the mixing of PV can yield the formation of staircases, i.e. successive regions of homogeneous PV separated by sharp interfaces (e.g. Scott & Dritschel 2012; Verhoeven & Stellmach 2014). Figure 8(*b*) shows that in our simulation the potential

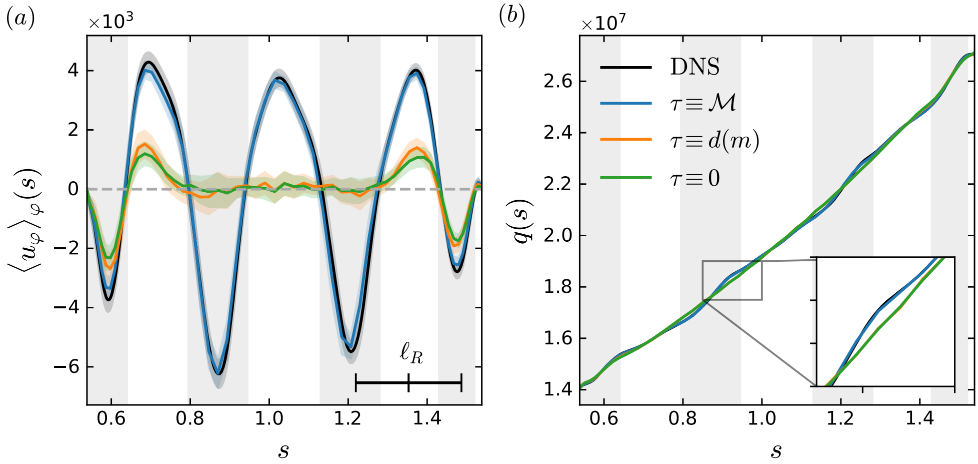

Figure 8: (*a*) Time-averaged radial profile of the azimuthal velocity $\langle u_\varphi \rangle_\varphi$ and (*b*) potential vorticity q for configuration (i) which has constant β. The alternating zonal jets are almost entirely lost at coarse resolution $\tau \equiv 0$, but also with hyperdiffusivity $\tau \equiv d(m)$. The horizontal segment in panel (*a*) corresponds to the Rhines scale ℓ_R (Eq. 4.12) in this configuration. The inset in panel (*b*) highlights the gradient of potential vorticity in the strongest retrograde jet.

vorticity is almost linear with moderate steepening in the regions where the jets are retrograde. Given the moderate parameters considered here –relatively large Ekman number and weak forcing–, the jets are still subjected to a strong role played by viscous friction and hence only carry a limited fraction of the total energy with $E_Z \approx 0.3 E_T$. It hence comes to no surprise that the level of staircasing remains mild (see e.g. Lemasquerier *et al.* 2023, for similar results at comparable parameters). This modest PV mixing is however sufficient to yield an asymmetry between the rounded prograde jets in regions of weaker PV gradient and the sharper retrograde jets when the gradient is stronger. This effect is well accounted for by the $\tau \equiv \mathcal{M}$ model but mostly absent from the alternative models $\tau \equiv 0$ and $\tau \equiv d(m)$.

4.5. *Spectral analysis*

In the presence of a β effect, two-dimensional rotating turbulent flows are expected to be anisotropic. In the so-called zonostrophic turbulence regime (e.g. Sukoriansky *et al.* 2002; Galperin *et al.* 2010), the kinetic energy spectrum can be conveniently split into a residual component E_R which accounts for non-axisymmetric motions and a zonal component E_Z, following

$$E_R(k) = C_R \epsilon^{2/3} k^{-5/3}, \quad E_Z(k_s) = C_Z \tilde{\beta}^2 k_s^{-5}, \tag{4.14}$$

where $\tilde{\beta} = 2\beta/E$, k_s is the wavenumber in the direction orthogonal to the zonal flow and k is the total wavenumber. The residual spectrum is expected to adhere to the classical Kolmogorov-Batchelor-Kraichnan (KBK) theory of two-dimensional turbulence, where $C_R \approx 5-6$ is a universal constant and ϵ is the energy transfer rate (Boffetta & Ecke 2012). In the above equation, C_Z is also a universal constant of order unity, in practice found to vary between 0.3 and 3 (see e.g. Sukoriansky *et al.* 2007; Galperin *et al.* 2014; Cabanes *et al.* 2017; Lemasquerier *et al.* 2023).

To assess the relevance of the zonostrophic theory to our numerical simulations, we adopt a spectral transform appropriate to the cylindrical annulus, in a similar fashion to Cabanes *et al.* (2024). To do so, we decompose the velocity field components (u_s, u_φ) using the

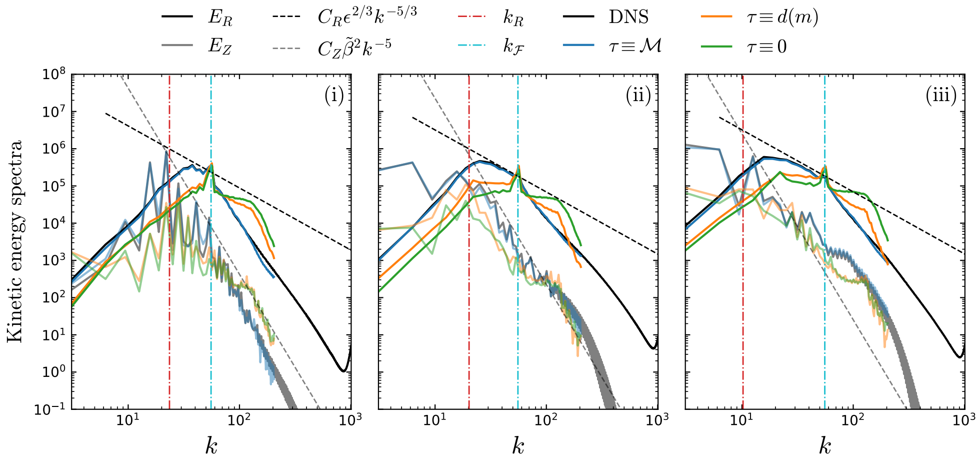

Figure 9: Time-averaged kinetic energy spectra of the residual (E_R) and zonal (E_Z) flow for each model and the three configurations studied. Dashed lines indicate the respective theoretical slopes for E_R and E_Z. Vertical dot-dashed lines mark the forcing and Rhines scales, $k_{\mathcal{F}}$ and k_R, in cyan and red, respectively. The theoretical prediction for the residual spectra (dashed black lines) assumes $C_R = 5$. For the zonal spectra, we adopt $C_Z = 0.1$ for cases (i) and (iii) and $C_Z = 0.08$ for case (ii). For the spherical cases (ii) and (iii), $\tilde{\beta}$ is evaluated at mid-depth.

so-called Weber-Orr transform, which can be expressed by (see e.g. Wordsworth *et al.* 2008),

$$u_{nm} = \int_{s_i}^{s_o} \int_0^{2\pi} u(s,\varphi)\Psi_{nm}(s)e^{-\mathrm{i}m\varphi} s\,\mathrm{d}s\,\mathrm{d}\varphi, \tag{4.15}$$

where n and m respectively denote the radial and azimuthal dimensionless wavenumbers. In the annulus, the support function Ψ_{nm} is a linear combination of Bessel functions of the first and second kind J_m and Y_m such that (e.g. MacRobert 1932)

$$\Psi_{nm}(s) = Y_m(a_{nm}s_o)J_m(a_{nm}s) - J_m(a_{nm}s_o)Y_m(a_{nm}s). \tag{4.16}$$

To enforce Dirichlet boundary conditions at both ends in the radial direction, the wavenumbers a_{nm} have to fulfill the following transcendental equation (Cinelli 1965)

$$Y_m(a_{nm}s_i)J_m(a_{nm}s_o) - J_m(a_{nm}s_i)Y_m(a_{nm}s_o) = 0. \tag{4.17}$$

From Eq. (4.15), we then compute the kinetic energy spectra using Parseval's identity

$$E_T = \sum_n \sum_{m=-N_m}^{N_m} E_{nm}, \tag{4.18}$$

where

$$E_{nm} = \frac{1}{2}\frac{\pi^2}{s_o^2 - s_i^2}\frac{a_{nm}^2 J_m^2(a_{nm}s_o)}{J_m^2(a_{nm}s_i) - Y_m^2(a_{nm}s_o)}\left(\left|u_{s_{nm}}\right|^2 + \left|u_{\varphi_{nm}}\right|^2\right). \tag{4.19}$$

Since a_{nm} depends on the azimuthal order m, the wavenumbers k are defined using the zonal wavenumber $k_n \equiv a_{n0}$ (Lemasquerier *et al.* 2023). The spectra of the non-axisymmetric residuals with $m \neq 0$ are then constructed by summation of the 2D spectral contribution E_{nm} into the bins defined by a_{n0}. To that end, we introduce $\mathcal{R}(p)$ as the ensemble of wavenumber indices (n, m) which satisfy $a_{n0} - \delta k_n/2 \leqslant a_{nm} \leqslant a_{n0} + \delta k_n/2$, where $\delta k_n = k_{n+1} - k_n \approx \pi$ is the almost equidistant bin spacing. Energy conservation then demands that $\sum_p E(k_p)\delta k_p = \sum_n \sum_{m=-N_m}^{N_m} E_{nm}$ such that $E(k_p) = 1/\delta k_p \sum_{(n,m)\in\mathcal{R}(p)} E_{nm}$. The dimension of $E(k_p)$ in the above expressions is a cubed length scale per squared time unit.

Figure 9 shows the time-averaged zonal spectra $E_Z(k)$ and residual spectra $E_R(k)$ respectively computed over 5000 and 100 statistically-independent snapshots. For each configuration, vorticity is injected at wavenumber $k_\mathcal{F}$ which is inversely proportional to the vortex spacing $\Delta_\mathcal{F}$ with $k_\mathcal{F} = \pi\sqrt{2}/\Delta_\mathcal{F} \approx 55$ (dash-dotted cyan lines). This corresponds to the localised peaks of residual energy visible in the three configurations for $k \approx 50-60$. The forcing hence mostly occurs at a much larger scale than the truncation of the coarse-grained models, to ensure that the model correction only includes energy transfers related to subgrid processes. The energy transfer rate can be estimated by $\epsilon = \mathcal{P}_\mathcal{F}$ once the statistically-steady state has been reached.

For the three considered configurations, we observe a limited range of scales –slightly broader in configurations (ii) and (iii) compared to (i)– over which the residual spectra adhere to the theoretical scaling $C_R\epsilon^{2/3}k^{-5/3}$ (dashed black lines), taking $C_R = 5$. At scales smaller than the forcing scale, the residual spectra drop with a slope close to -5, steeper than the expected -3 scaling due to the direct cascade of enstrophy in two-dimensional turbulence (Boffetta & Ecke 2012). Such steeper scaling behaviors were also reported by Lemasquerier *et al.* (2023) and are attributed to the role played by the large-scale drag (Boffetta 2007). While the $\tau \equiv \mathcal{M}$ model accurately captures the residual spectra up to the scale of the numerical truncation, both the under-resolved simulation and hyperdiffusivity model fail to capture the $-5/3$ slope and lack energetic content at large scale. Regarding the zonal spectra, we recover in configurations (i) and (ii) a range of scales where the spectra also conform to the theoretical scaling $C_Z\tilde{\beta}^2k^{-5}$ with $C_Z \approx 0.1$ (dashed grey lines). The -5 slope is less obvious in configuration (iii) and limited to a narrower range of wavenumbers. One possible explanation for this difference is that in configuration (ii) the jets are mostly located in the lower half of the domain, whereas in configuration (iii) they are developed throughout the entire fluid volume (recall Fig. 5). This implies that the variability of β in the spherical container will have a greater impact on the zonal spectrum in configuration (iii) than in configuration (ii). The value obtained for C_Z is in the low range of published values, consistent with those obtained by Lemasquerier *et al.* (2023) for a configuration with a similar topographic β effect. The agreement with the -5 slope approximately stops at the Rhines scale $k_R = 2\pi/\ell_R$ (dash-dotted red lines), which is also close to the most energetic scale in the three configurations. We confirm that our configurations are not in a zonostrophic regime by observing a zonostrophy index $R_\beta \approx 1$. This index is defined as $R_\beta = k_\beta/k_R$, where k_β is the spectral transitional wavenumber at the intersection of the theoretical $-5/3$ and -5 slopes. While the -5 slope is observed in each model, the value of C_Z is underestimated in both the under-resolved simulation and the hyperdiffusivity model. In contrast, the learned model $\tau \equiv \mathcal{M}$ correctly reproduces the entire zonal spectrum.

4.6. *Temporal patterns: jets and waves*

Finally, we explore the temporal dependence of large-scale jets and Rossby waves in the vicinity of the outer boundary for configuration (ii). In the case of thermal convection in a rapidly rotating thick spherical shell, a quasi-periodic inward drift of zonal jets was reported by Rotvig (2007), using both the three-dimensional description of the system and its quasi-geostrophic approximation. Such a migration can be attributed to the variations of $\beta(s)$ over one single jet, that promotes a better energy transfer on the inward flank of retrograde jets (Chemke & Kaspi 2015). Figure 10 shows the time-evolution of the mean zonal velocity $\langle u_\varphi\rangle_\varphi(s,t)$ over 100 turnover times t_L. Figure 10(a) reveals that jets in the DNS indeed drift inward in a quasi-periodic manner, with a recurrence time scale of about 10 t_L. The only

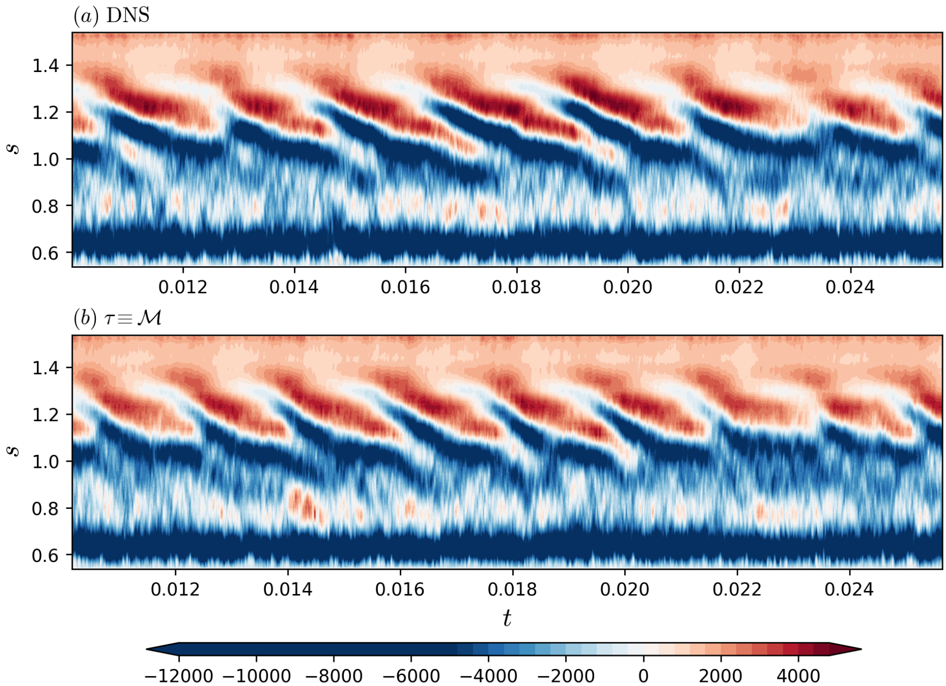

Figure 10: (a) Time evolution of the zonal velocity $\langle u_\varphi \rangle_\varphi(s,t)$ for configuration (ii) computed with the DNS. (b) Same quantity computed using the learned model $\mathcal{M}$. The time interval corresponds to 100 turnover times.

model capable of reproducing those drifts is $\tau \equiv \mathcal{M}$, and it does so with the same typical recurrence time, as illustrated by Fig. 10 (b).

Close to the outer boundary, as seen previously in Fig. 5(c), Rossby waves prevail. Figure 11 shows a longitude-time map of the cylindrically radial velocity u_s at $s = 1.32$ over one single turnover time. Observed patterns evolve on time scales much shorter than the jet drift time scale. The evolution of these non-axisymmetric patterns results from the superposition of two effects: advection by the average background zonal flow at this radius, indicated by the black line in Fig. 11, and retrograde transport by Rossby waves, since patterns appear steeper than would be expected from pure advection by the zonal flow. The model $\tau \equiv \mathcal{M}$ also successfully predicts these rapidly evolving small-scale structures. Its success stems from an accurate representation of advection by the zonal flow -in agreement with the reference DNS- which preserves the quasi-linear conditions necessary for Rossby wave propagation in this region.

5. Discussion

We proposed a subgrid-scale model correction that is stable and accurate with respect to a variety of metrics in the context of QG turbulence in a cylindrical geometry with no-slip boundary conditions. The model is designed as a neural network that leverages differentiable programming to optimise an “online” loss function that is able to backpropagate the adjoint of the numerical model. We first described a Galerkin projection that allows us to generate a dataset that maintains boundary conditions when a spectral Chebyshev discretisation is used. Then, we discussed how the presence of commutation error arising in the product

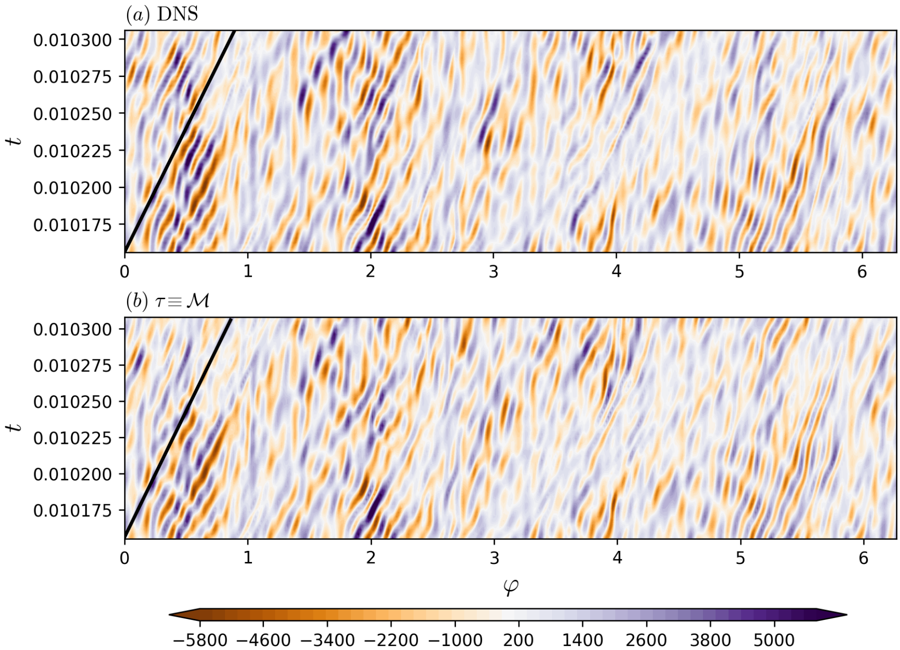

Figure 11: (*a*) Hovmöller diagrams (azimuth-time) of the radial velocity u_s at a fixed radius $s = 1.32$ inside the Rossby waves region for configuration (ii) computed with the DNS. (*b*) Same quantity computed using the learned model $\mathcal{M}$. The time interval approximately corresponds to one single turnover time. The solid black line shows the drift due to the advection by zonal velocity, with $\langle u_\varphi \rangle_{\varphi,t}(s) = 1320$ for the DNS and $\langle u_\varphi \rangle_{\varphi,t}(s) = 1268$ for the learned model $\mathcal{M}$.

of filtered quantities in the non-homogeneous (radial) direction and propagating into the definition of the subgrid-scale term makes "offline" learning difficult. In practice (although not shown in the paper), we found that a model trained using the "offline" strategy does not lead to stable simulations. However, we did not investigate whether this comes from already known instabilities from "offline" learning (Frezat *et al.* 2022; Yan *et al.* 2025) or from the commutation error itself. Instead, the "online" learning does not involve any explicit calculation of the subgrid term, and only computes a mismatch between quantities that have been projected to a coarse-grained grid. In the result section, we arbitrarily set the training timespan to one turnover time, which corresponds to relatively small datasets that contain less than 1000 samples for our configurations. While exploring the importance of this timescale, we found that reducing the size of the dataset further did not significantly impact the accuracy of the model. We have observed that the model is able to predict fast processes, such as Rossby waves, but also slow processes such as zonal jets drifting in the spherical container. Since the model has only seen a fraction of this phenomenon, we argue that the energy transfers produced by the model are mostly small-scale and statistically uniform over small timespans. During inference, embarking the NNs results in additional computations that increases the prediction time. On a single NVIDIA A-100 GPU, an integration over 100 turnover times for configuration (i) takes 30.5 minutes using the learned model $\tau \equiv \mathcal{M}$ while the coarse resolution $\tau \equiv 0$ and hyperdiffusivity $\tau \equiv d(m)$ simulations necessitate 5 minutes. Yet, this overhead allows for an adequate description of the subgrid-scale energy transfers, while enabling a 6 fold reduction compared to the full DNS simulation, whose completion requires 181 minutes. Currently, the configurations studied in this work are limited to a relatively moderate Reynolds number $Re \sim 10^4$ due to the computational cost of the Chebyshev collocation method, which produces dense matrices. Using an efficient implementation based on the Chebyshev integration method (Marti *et al.* 2016), it is possible to reach more turbulent regimes, up to $Re \sim 10^6$. In this paper, we used the same code to generate data and train the model, so that we are sure that there is no mismatch between the numerical method used to generate data and the numerical method used to integrate the system during the learning phase. In practice, it should be possible to generate the dataset separately, using a distributed high-performance code such as (Gastine 2019) and use the differentiable collocation implementation to perform training and model evaluation, since these only live on the coarse-grained grid. This work represents a first step towards the implementation of stable and accurate subgrid-scale models for Earth's dynamo. Here we lifted multiple constraints;

First, the artificial forcing allows us to reduce the number of equations to close to one, i.e., the evolution of vorticity, plus the axisymmetric part that has been treated separately in the model. But in practice, these quantities are connected and we demonstrated that minimising the kinetic energy mismatch is enough to learn a correction for both equations. Now, changing the equations so that the system is forced by convection, we have a new equation for the temperature (perturbation) evolution $\vartheta(t)$. It is still an open question whether it is possible to train a single model with a weighted loss function $L[f(t), \overline{f}(t)] = \lambda_{\boldsymbol{u}} L[\boldsymbol{u}(t), \overline{\boldsymbol{u}}(t)] + \lambda_\vartheta L[\vartheta(t), \overline{\vartheta}(t)]$ that combines a mismatch between velocities and temperature or if we need to train two separate models that produce a single subgrid term each for a specific equation. In the latter case, the "online" strategy would probably require two integration passes that optimise for the parameters of each model sequentially.

Second, the QG system represents a two-dimensional, computationally efficient approximation to the full three-dimensional system in a spherical shell. Here, the complexity gap is expected to be important because of the additional dimension. Moreover, the dynamo system also solves for the magnetic field evolution $\boldsymbol{B}(t)$, which means that the number of subgrid-scale terms to close is then 5 when considering a poloidal-toroidal decomposition of the velocity and magnetic fields (e.g. Glatzmaier 1984).

Acknowledgements. Numerical computations were performed on the S-CAPAD/DANTE platform, IPGP, France.

Funding. This work has been supported by the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101093038 (ChEESE-2P). As part of the "France 2030" initiative, this work has benefited from a national grant managed by the French National Research Agency (Agence Nationale de la Recherche) attributed to the Exa-DI project of the NumPEx PEPR program, under the reference ANR-22-EXNU-0006.

Declaration of interests. The authors report no conflict of interest.

Data availability statement. The code used to generate data, train and evaluate models has been made openly available in https://github.com/hrkz/online-sgs-qg-planetary as a collection of scripts and notebooks, and archived on Software Heritage with SWHID swh:1:snp:ee512615bc7ef0bf8d1e0f4bb99c617526b96b14.

Author ORCIDs. H. Frezat, https://orcid.org/0000-0002-0291-4020; T. Gastine, https://orcid.org/0000-0003-4438-7203; A. Fournier, https://orcid.org/0000-0003-3276-0496

Author contributions. Authors may include details of the contributions made by each author to the manuscript'

Appendix A. Ekman pumping contribution

In this appendix, we recall the Ekman pumping contribution outside the tangent cylinder in a spherical container and compute the contribution in an exponential container, in cylindrical coordinates. The Ekman pumping flow is given by Greenspan (1968),

$$
\begin{aligned}
\boldsymbol{u}\cdot\boldsymbol{n}|_{\pm h} &= -\frac{E^{1/2}}{2}\boldsymbol{n}\cdot\boldsymbol{\nabla}\times\left(\frac{\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u}}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}\right)\\
&= \boldsymbol{n}\cdot\left[\frac{\boldsymbol{\nabla}\times(\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u})}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}+\boldsymbol{\nabla}\left(\frac{1}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}\right)\times(\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u})\right]
\end{aligned}
\tag{A 1}
$$

where $\boldsymbol{n}$ is the outward unit vector normal to the surface. From now on, we assume that the vertical velocity u_z increases linearly with z such that

$$
\frac{\partial u_z}{\partial z} = \beta u_z = \frac{1}{h}\frac{\mathrm{d}h}{\mathrm{d}s}u_s \tag{A 2}
$$

In a spherical container with $h = (s_o^2 - s^2)^{1/2}$, Ekman pumping is well known, e.g. (Schaeffer & Cardin 2005) and can be expressed as,

$$
\boldsymbol{u}\cdot\boldsymbol{n}|_{\pm h} = \frac{s}{s_o}u_s \pm \frac{h}{s_o}u_z = -\frac{E^{1/2}}{2}\frac{h^{1/2}}{s_o^{1/2}}\left[\omega_z + \frac{s}{2h^2}u_\varphi + \frac{5}{2}\frac{s_o s}{h^3}u_s - \frac{s}{h^2}\frac{\partial u_s}{\partial\varphi}\right]. \tag{A 3}
$$

As discussed in §2.1, we only retain the term proportional to vorticity in our implementation. The stretching term used in (2.6)-(2.7) for the spherical container then reads

$$
\frac{2}{E}\frac{\partial u_z}{\partial z} = -\frac{2}{E}\frac{s}{h^2}u_s + \underbrace{\frac{s_o^{1/2}}{E^{1/2}h^{3/2}}}_{\Upsilon}\omega_z. \tag{A 4}
$$

Let us now derive (A 1) for an exponential container with $h = \exp(\beta s)$. Let us expand the first part using the following identity,

$$
\boldsymbol{\nabla}\times(\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u}) = \boldsymbol{n}(\boldsymbol{\nabla}\cdot\boldsymbol{u}) - \boldsymbol{u}(\boldsymbol{\nabla}\cdot\boldsymbol{n}) + \boldsymbol{\nabla}\boldsymbol{n}\cdot\boldsymbol{u} - \boldsymbol{\nabla}\boldsymbol{u}\cdot\boldsymbol{n} \pm \boldsymbol{\omega}, \tag{A 5}
$$

where $\boldsymbol{\omega} = \boldsymbol{\nabla}\times\boldsymbol{u}$ is the vorticity. The outward unit normal to the exponential container surface is expressed by

$$
\boldsymbol{n} = \frac{1}{(1+\beta^2h^2)^{1/2}}\begin{pmatrix}-\beta h & 0 & \pm 1\end{pmatrix}^{\mathsf{T}}, \tag{A 6}
$$

where $+1$ (resp. -1) is to be used in the $z > 0$ (resp. $z < 0$) half-space. The corresponding gradient and divergence read

$$
\boldsymbol{\nabla}\boldsymbol{n} = \begin{pmatrix}
\dfrac{\beta^2 h}{(1+\beta^2h^2)^{3/2}} & 0 & 0\\
0 & -\dfrac{\beta h}{s(1+\beta^2h^2)^{1/2}} & 0\\
\pm\dfrac{\beta^3h^2}{(1+\beta^2h^2)^{3/2}} & 0 & 0
\end{pmatrix}, \quad \boldsymbol{\nabla}\cdot\boldsymbol{n} = -\frac{\beta^2 h}{(1+\beta^2h^2)^{3/2}} - \frac{\beta h}{s(1+\beta^2h^2)^{1/2}}.
$$

In cylindrical coordinates, the gradient of the velocity $\boldsymbol{\nabla u}$ reads, assuming that $\partial u_s/\partial z = \partial u_\varphi/\partial z = 0$,

$$\boldsymbol{\nabla u} = \begin{pmatrix} \dfrac{\partial u_s}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_s}{\partial u_\varphi} - \dfrac{u_\varphi}{s} & \dfrac{\partial u_z}{\partial z} \\ \dfrac{\partial u_\varphi}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_\varphi}{\partial \varphi} + \dfrac{u_s}{s} & \dfrac{\partial u_\varphi}{\partial z} \\ \dfrac{\partial u_z}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_z}{\partial \varphi} & \dfrac{\partial u_z}{\partial z} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial u_s}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_s}{\partial u_\varphi} - \dfrac{u_\varphi}{s} & 0 \\ \dfrac{\partial u_\varphi}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_\varphi}{\partial \varphi} + \dfrac{u_s}{s} & 0 \\ \dfrac{\partial u_z}{\partial s} & \dfrac{1}{s}\dfrac{\partial u_z}{\partial \varphi} & \pm\dfrac{1}{h}\dfrac{\mathrm{d}h}{\mathrm{d}s}u_s \end{pmatrix}.$$

Substituting these quantities in the identity (A 5), omitting the calculation in φ since $n_\varphi = 0$, we get the following

$$\boldsymbol{\nabla} \times (\boldsymbol{n} \times \boldsymbol{u} \pm \boldsymbol{u}) = \begin{pmatrix} \dfrac{\beta h}{s(1+\beta^2h^2)^{1/2}}u_s + \dfrac{\beta h}{(1+\beta^2h^2)^{1/2}}\dfrac{\partial u_s}{\partial s} + \dfrac{\beta h}{s}\dfrac{\partial u_s}{\partial \varphi} \\ \cdots \\ \pm\dfrac{\beta^2h^2}{s(1+\beta^2h^2)^{1/2}}u_s \pm \dfrac{\beta}{(1+\beta^2h^2)^{1/2}}u_s \pm \dfrac{\beta^2h^2}{(1+\beta^2h^2)^{1/2}}\dfrac{\partial u_s}{\partial s} \pm \omega_z \end{pmatrix}.$$

Then, the first part of the Ekman contribution (A 1) simplifies to,

$$\frac{1}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}\boldsymbol{n}\cdot\boldsymbol{\nabla}\times(\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u}) = \frac{1}{(1+\beta^2h^2)^{1/4}}\left[\omega_z - \frac{\beta}{(1+\beta^2h^2)^{1/2}}u_s - \frac{\beta^2h^2}{s}\frac{\partial u_s}{\partial \varphi}\right]. \quad \text{(A 7)}$$

For the second part of the Ekman contribution, let us first observe that relationship with planetary vorticity f,

$$\boldsymbol{\nabla}\left(\frac{1}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}\right) = f\boldsymbol{e}_s = \frac{\beta^3h^2}{2(1+\beta^2h^2)^{3/4}}\boldsymbol{e}_s, \quad \text{(A 8)}$$

and thus

$$\boldsymbol{n}\cdot f\boldsymbol{e}_s \times (\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u}) = \frac{1}{\sqrt{1+\beta^2h^2}}\left[f\left(1+\beta^2h^2\right)^{1/2}u_s + fu_\varphi\right].$$

Replacing f from (A 8), we obtain the second part of the Ekman contribution (A 1),

$$\boldsymbol{n}\cdot\boldsymbol{\nabla}\left(\frac{1}{\sqrt{|\boldsymbol{n}\cdot\boldsymbol{e}_z|}}\right)\times(\boldsymbol{n}\times\boldsymbol{u}\pm\boldsymbol{u}) = \frac{1}{(1+\beta^2h^2)^{1/4}}\left[\frac{\beta^3h^2}{2(1+\beta^2h^2)^{1/2}}u_s + \frac{\beta^3h^2}{2(1+\beta^2h^2)}u_\varphi\right]. \quad \text{(A 9)}$$

Finally, combining (A 7) and (A 9) gives,

$$\boldsymbol{u}\cdot\boldsymbol{n}|_{\pm h} = -\frac{E^{1/2}}{2}\frac{1}{(1+\beta^2h^2)^{1/4}}\left[\omega_z - \frac{2\beta-\beta^3h^2}{2(1+\beta^2h^2)^{1/2}}u_s - \frac{\beta^2h^2}{s}\frac{\partial u_s}{\partial \varphi} + \frac{\beta^3h^2}{2(1+\beta^2h^2)}u_\varphi\right]. \quad \text{(A 10)}$$

Now, observe that expanding $\boldsymbol{u}\cdot\boldsymbol{n}|_{\pm h}$ using (A 6) gives an expression for u_z as a function of the Ekman pumping. Using a difference formula $\partial u_z/\partial z = (u_z(h) - u_z(-h))/2h$, the stretching term used in (2.6)-(2.7) for the exponential container can be derived similarly than for the spherical container by substituting u_z using (A 10) and neglecting the non-dominant term,

$$\frac{2}{E}\frac{\partial u_z}{\partial z} = \frac{2}{E}\beta u_s + \underbrace{\frac{(1+\beta^2h^2)^{1/4}}{E^{1/2}h}}_{\Upsilon}\omega_z. \quad \text{(A 11)}$$

Appendix B. Power input

In order to derive the expression of the input power $\mathcal{P}_{\mathcal{F}}$, we have to uncurl the expression of the source of vorticity $\mathcal{F}$ in Eq. (2.8). To that end, we begin by considering a single Gaussian pump located at the origin of the 2D plane. In cylindrical coordinates (s, φ, z), we seek a vector $\mathbf{f}$ whose curl reads

$$\boldsymbol{\nabla} \times \mathbf{f} = \exp\left[-\left(\frac{s}{\ell_{\mathcal{F}}}\right)^2\right]\hat{\boldsymbol{z}}.$$

Specifically we require $\mathbf{f}$ to be of the form $\mathbf{f} = f(s)\hat{\boldsymbol{\varphi}}$. In addition, we ask f to be regular at $s = 0$. It is straightforward to obtain

$$f(s) = \frac{\ell_{\mathcal{F}}^2}{2s}\left\{1 - \exp\left[-\left(\frac{s}{\ell_{\mathcal{F}}}\right)^2\right]\right\}.$$

To facilitate subsequent superposition, we switch from cylindrical to Cartesian coordinates, and note that

$$\hat{\boldsymbol{\varphi}} = \frac{-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}}{s}.$$

Consequently, for a single unitary Gaussian source located at the origin, we have

$$\mathbf{f} = f(s)\hat{\boldsymbol{\varphi}} = \frac{\ell_{\mathcal{F}}^2}{2s^2}\left\{1 - \left[\exp -\left(\frac{s}{\ell_{\mathcal{F}}}\right)^2\right]\right\}(-y\hat{\boldsymbol{x}} + x\hat{\boldsymbol{y}}).$$

If the center of the source is now placed at (x_i, y_i) the previous expression becomes

$$\mathbf{f} = \frac{\ell_{\mathcal{F}}^2}{2s_i^2}\left\{1 - \exp\left[-\left(\frac{s_i}{\ell_{\mathcal{F}}}\right)^2\right]\right\}\left[-(y - y_i)\hat{\boldsymbol{x}} + (x - x_i)\hat{\boldsymbol{y}}\right], \tag{B 1}$$

where $s_i = s_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}$. Let $p_{\mathcal{F}}$ denote the input power density. Using Eq. (B 1) in conjunction with Eq. (2.8) leads by superposition to

$$p_{\mathcal{F}} = \frac{a_{\mathcal{F}}\ell_{\mathcal{F}}^2}{2}\sum_{i=1}^{N}\frac{(-1)^i}{s_i^2}\left\{1 - \exp\left[-\left(\frac{s_i}{\ell_{\mathcal{F}}}\right)^2\right]\right\}\left[(x - x_i)u_y - (y - y_i)u_x\right], \tag{B 2}$$

where $u_x = u_s\cos\varphi - u_\varphi\sin\varphi$ and $u_y = u_s\sin\varphi + u_\varphi\cos\varphi$. The input power at any time t averaged over the domain is then $\mathcal{P}_{\mathcal{F}} = \langle p_{\mathcal{F}}\rangle$, with the domain average operator defined in Eq. (4.1).

REFERENCES

Arcomano, Troy, Szunyogh, Istvan, Wikner, Alexander, Hunt, Brian R & Ott, Edward 2023 A hybrid atmospheric model incorporating machine learning can capture dynamical processes not captured by its physics-based component. *Geophysical Research Letters* **50** (8), e2022GL102649.

Arjovsky, Martin, Shah, Amar & Bengio, Yoshua 2016 Unitary evolution recurrent neural networks. In *International Conference on Machine Learning*, pp. 1120–1128. PMLR.

Aubert, Julien, Gastine, Thomas & Fournier, Alexandre 2017 Spherical convective dynamos in the rapidly rotating asymptotic regime. *Journal of Fluid Mechanics* **813**, 558–593.

Aubert, Julien, Gillet, Nicolas & Cardin, Philippe 2003 Quasigeostrophic models of convection in rotating spherical shells. *Geochemistry, Geophysics, Geosystems* **4** (7).

Bardina, Jorge, Ferziger, J & Reynolds, WC 1980 Improved subgrid-scale models for large-eddy simulation. In *13th Fluid and PlasmaDynamics Conference*, p. 1357. American Institute of Aeronautics and Astronautics.

Barrois, Olivier, Gastine, Thomas & Finlay, Christopher C 2022 Comparison of quasi-geostrophic, hybrid and 3-D models of planetary core convection. *Geophysical Journal International* **231** (1), 129–158.

Baydin, Atilim Gunes, Pearlmutter, Barak A, Radul, Alexey Andreyevich & Siskind, Jeffrey Mark 2018 Automatic differentiation in machine learning: a survey. *Journal of machine learning research* **18** (153), 1–43.

Boffetta, Guido 2007 Energy and enstrophy fluxes in the double cascade of two-dimensional turbulence. *Journal of Fluid Mechanics* **589**, 253–260.

Boffetta, Guido & Ecke, Robert E 2012 Two-dimensional turbulence. *Annual Review of Fluid Mechanics* **44** (1), 427–451.

Bolton, Thomas & Zanna, Laure 2019 Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems* **11** (1), 376–399.

Boscarino, Sebastiano, Pareschi, Lorenzo & Russo, Giovanni 2013 Implicit-explicit Runge–Kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit. *SIAM Journal on Scientific Computing* **35** (1), A22–A51.

Boyd, John P 2001 *Chebyshev and Fourier spectral methods*, 2nd edn. Dover.

Bradbury, James, Frostig, Roy, Hawkins, Peter, Johnson, Matthew James, Leary, Chris, Maclaurin, Dougal, Necula, George, Paszke, Adam, VanderPlas, Jake, Wanderman-Milne, Skye & Zhang, Qiao 2018 JAX: composable transformations of Python+NumPy programs.

Braginsky, SI & Meytlis, VP 1990 Local turbulence in the earth's core. *Geophysical & Astrophysical Fluid Dynamics* **55** (2), 71–87.

Brunton, Steven L, Noack, Bernd R & Koumoutsakos, Petros 2020 Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* **52** (1), 477–508.

Buffett, Bruce A 2003 A comparison of subgrid-scale models for large-eddy simulations of convection in the Earth's core. *Geophysical Journal International* **153** (3), 753–765.

Cabanes, Simon, Aurnou, Jonathan, Favier, Benjamin & Le Bars, Michael 2017 A laboratory model for deep-seated jets on the gas giants. *Nature Physics* **13** (4), 387–390.

Cabanes, Simon, Gastine, Thomas & Fournier, Alexandre 2024 Zonostrophic turbulence in the subsurface oceans of the Jovian and Saturnian moons. *Icarus* **415**, 116047.

Canuto, Claudio, Hussaini, M Youssuff, Quarteroni, Alfio & Zang, Thomas A 2006 *Spectral methods*, , vol. 285. Springer.

Chemke, Rei & Kaspi, Yohai 2015 Poleward migration of eddy-driven jets. *Journal of Advances in Modeling Earth Systems* **7** (3), 1457–1471.

Cinelli, G1 1965 An extension of the finite Hankel transform and applications. *International Journal of Engineering Science* **3** (5), 539–559.

Davy, Bradley, Mound, Jonathan E, Davies, Christopher & Tobias, Steven 2023 The effects of hyperdiffusion on rotating Rayleigh-Bénard convection: Insights and computational benefits. In *AGU Fall Meeting Abstracts*, , vol. 2023, pp. NG42A–01.

Dueben, Peter D & Bauer, Peter 2018 Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development* **11** (10), 3999–4009.

Finlay, Christopher C, Gillet, Nicolas, Aubert, Julien, Livermore, Philip W & Jault, Dominique 2023 Gyres, jets and waves in the Earth's core. *Nature Reviews Earth & Environment* **4** (6), 377–392.

Finn, Tobias Sebastian, Durand, Charlotte, Farchi, Alban, Bocquet, Marc, Chen, Yumeng, Carrassi, Alberto & Dansereau, Véronique 2023 Deep learning subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell elasto-brittle rheology. *The Cryosphere* **17** (7), 2965–2991.

Frezat, Hugo, Balarac, Guillaume, Le Sommer, Julien, Fablet, Ronan & Lguensat, Redouane 2021 Physical invariance in neural networks for subgrid-scale scalar flux modeling. *Physical Review Fluids* **6** (2), 024607.

Frezat, Hugo, Fablet, Ronan, Balarac, Guillaume & Le Sommer, Julien 2023 Gradient-free online learning of subgrid-scale dynamics with neural emulators. *arXiv preprint* .

Frezat, Hugo, Le Sommer, Julien, Fablet, Ronan, Balarac, Guillaume & Lguensat, Redouane 2022 A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems* **14** (11), e2022MS003124.

Galperin, Boris, Sukoriansky, Semion & Dikovskaya, Nadejda 2010 Geophysical flows with anisotropic turbulence and dispersive waves: flows with a β-effect. *Ocean Dynamics* **60**, 427–441.

Galperin, Boris, Young, Roland MB, Sukoriansky, Semion, Dikovskaya, Nadejda, Read, Peter L, Lancaster, Andrew J & Armstrong, David 2014 Cassini observations reveal a regime of zonostrophic macroturbulence on Jupiter. *Icarus* **229**, 295–320.

Gastine, Thomas 2019 pizza: an open-source pseudo-spectral code for spherical quasi-geostrophic convection. *Geophysical Journal International* **217** (3), 1558–1576.

Gelbrecht, Maximilian, White, Alistair, Bathiany, Sebastian & Boers, Niklas 2023 Differentiable programming for Earth system modeling. *Geoscientific Model Development* **16** (11), 3123–3135.

Germano, Massimo 1986 A proposal for a redefinition of the turbulent stresses in the filtered Navier-Stokes equations. *Physics of Fluids* **29** (7), 2323–2324.

Germano, Massimo, Piomelli, Ugo, Moin, Parviz & Cabot, William H 1991 A dynamic subgrid-scale eddy viscosity model. *Physics of Fluids A: Fluid Dynamics* **3** (7), 1760–1765.

Ghosal, Sandip & Moin, Parviz 1995 The basic equations for the large eddy simulation of turbulent flows in complex geometry. *Journal of Computational Physics* **118** (1), 24–37.

Gillet, N & Jones, CA 2006 The quasi-geostrophic model for rapidly rotating spherical convection outside the tangent cylinder. *Journal of Fluid Mechanics* **554**, 343–369.

Glatzmaier, Gary A 1984 Numerical simulations of stellar convective dynamos. I. the model and method. *Journal of Computational Physics* **55** (3), 461–484.

Glatzmaier, Gary A & Roberts, Paul H 1995 A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* **377** (6546), 203–209.

Gopinath, Venkatesh, Fournier, Alexandre & Gastine, Thomas 2022 An assessment of implicit-explicit time integrators for the pseudo-spectral approximation of Boussinesq thermal convection in an annulus. *Journal of Computational Physics* **460**, 110965.

Greenspan, Harvey Philip 1968 *The theory of rotating fluids*. Cambridge University Press.

Guan, Yifei, Chattopadhyay, Ashesh, Subel, Adam & Hassanzadeh, Pedram 2022 Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *Journal of Computational Physics* **458**, 111090.

Guervilly, Céline & Cardin, Philippe 2017 Multiple zonal jets and convective heat transport barriers in a quasi-geostrophic model of planetary cores. *Geophysical Journal International* **211** (1), 455–471.

Heimpel, Moritz & Aurnou, Jonathan 2007 Turbulent convection in rapidly rotating spherical shells: A model for equatorial and high latitude jets on Jupiter and Saturn. *Icarus* **187** (2), 540–557.

Hirose, Akira 2006 *Complex-valued neural networks*, 1st edn. Springer.

Hjørungnes, Are & Gesbert, David 2007 Complex-valued matrix differentiation: Techniques and key results. *IEEE Transactions on Signal Processing* **55** (6), 2740–2746.

Jakhar, Karan, Guan, Yifei, Mojgani, Rambod, Chattopadhyay, Ashesh & Hassanzadeh, Pedram 2024 Learning closed-form equations for subgrid-scale closures from high-fidelity data: Promises and challenges. *Journal of Advances in Modeling Earth Systems* **16** (7), e2023MS003874.

Julien, Keith & Watson, Mike 2009 Efficient multi-dimensional solution of PDEs using Chebyshev spectral methods. *Journal of Computational Physics* **228** (5), 1480–1503.

Kochkov, Dmitrii, Smith, Jamie A, Alieva, Ayya, Wang, Qing, Brenner, Michael P & Hoyer, Stephan 2021 Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences* **118** (21), e2101784118.

Kochkov, Dmitrii, Yuval, Janni, Langmore, Ian, Norgaard, Peter, Smith, Jamie, Mooers, Griffin, Klöwer, Milan, Lottes, James, Rasp, Stephan, Düben, Peter & others 2024 Neural general circulation models for weather and climate. *Nature* **632** (8027), 1060–1066.

Kuang, Weijia & Bloxham, Jeremy 1999 Numerical modeling of magnetohydrodynamic convection in a rapidly rotating spherical shell: weak and strong field dynamo action. *Journal of Computational Physics* **153** (1), 51–81.

Landeau, Maylis, Fournier, Alexandre, Nataf, Henri-Claude, Cébron, David & Schaeffer, Nathanaël 2022 Sustaining Earth's magnetic dynamo. *Nature Reviews Earth & Environment* **3** (4), 255–269.

Lemasquerier, Daphné, Favier, Benjamin & Le Bars, Michael 2023 Zonal jets experiments in the gas giants' zonostrophic regime. *Icarus* **390**, 115292.

Leonard, Athony 1975 Energy cascade in large-eddy simulations of turbulent fluid flows. In *Advances in Geophysics*, , vol. 18, pp. 237–248. Elsevier.

Lesieur, Marcel & Metais, Olivier 1996 New trends in large-eddy simulations of turbulence. *Annual Review of Fluid Mechanics* **28** (1), 45–82.

Liu, Zhuang, Mao, Hanzi, Wu, Chao-Yuan, Feichtenhofer, Christoph, Darrell, Trevor & Xie, Saining 2022 A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 11976–11986.

Loshchilov, Ilya & Hutter, Frank 2019 Decoupled weight decay regularization. In *International Conference on Learning Representations*.

MacRobert, TM 1932 XVI.—Fourier integrals. *Proceedings of the Royal Society of Edinburgh* **51**, 116–126.

Maltrud, ME & Vallis, GK 1991 Energy spectra and coherent structures in forced two-dimensional and beta-plane turbulence. *Journal of Fluid Mechanics* **228**, 321–342.

Marti, P, Calkins, MA & Julien, K 2016 A computationally efficient spectral method for modeling core dynamics. *Geochemistry, Geophysics, Geosystems* **17** (8), 3031–3053.

Matsui, Hiroaki & Buffett, Bruce A 2005 Sub-grid scale model for convection-driven dynamos in a rotating plane layer. *Physics of the Earth and Planetary Interiors* **153** (1-3), 108–123.

Matsui, Hiroaki & Buffett, Bruce A 2007*a* Commutation error correction for large eddy simulations of convection driven dynamos. *Geophysical and Astrophysical Fluid Dynamics* **101** (5-6), 429–449.

Matsui, Hiroaki & Buffett, Bruce A 2007*b* A dynamic scale-similarity model for dynamo simulations in a rotating plane layer. *Geophysical and Astrophysical Fluid Dynamics* **101** (5-6), 451–468.

Matsui, Hiroaki & Buffett, Bruce A 2012 Large-eddy simulations of convection-driven dynamos using a dynamic scale-similarity model. *Geophysical & Astrophysical Fluid Dynamics* **106** (3), 250–276.

Matsui, Hiroaki & Buffett, Bruce A 2013 Characterization of subgrid-scale terms in a numerical geodynamo simulation. *Physics of the Earth and Planetary Interiors* **223**, 77–85.

Matsushima, Masaki, Nakajima, Takahiro & Roberts, Paul H 1999 The anisotropy of local turbulence in the Earth's core. *Earth, Planets and Space* **51**, 277–286.

Maulik, Romit, San, Omer, Rasheed, Adil & Vedula, Prakash 2019 Subgrid modelling for two-dimensional turbulence using neural networks. *Journal of Fluid Mechanics* **858**, 122–144.

McIntyre, Michael E 2003 Potential vorticity. *Encyclopedia of Atmospheric Sciences* **2**, 685–694.

Meneveau, Charles & Katz, Joseph 2000 Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics* **32** (1), 1–32.

Nataf, H-C & Schaeffer, Nathanaël 2015 Turbulence in the core. In *Treatise on Geophysics*, 2nd edn., pp. 161–181. Elsevier.

Ouala, Said, Chapron, Bertrand, Collard, Fabrice, Gaultier, Lucile & Fablet, Ronan 2024 Online calibration of deep learning sub-models for hybrid numerical modeling systems. *Communications Physics* **7** (1), 402.

Pais, MA & Jault, Dominique 2008 Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field. *Geophysical Journal International* **173** (2), 421–443.

Plaut, Emmanuel & Busse, Friedrich H 2002 Low-Prandtl-number convection in a rotating cylindrical annulus. *Journal of Fluid Mechanics* **464**, 345–363.

Rasp, Stephan 2020 Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: General algorithms and Lorenz 96 case study (v1. 0). *Geoscientific Model Development* **13** (5), 2185–2196.

Rasp, Stephan, Pritchard, Michael S & Gentine, Pierre 2018 Deep learning to represent subgrid processes in climate models. *Proceedings of the national academy of sciences* **115** (39), 9684–9689.

Rhines, Peter B 1975 Waves and turbulence on a beta-plane. *Journal of Fluid Mechanics* **69** (3), 417–443.

Rotvig, Jon 2007 Multiple zonal jets and drifting: thermal convection in a rapidly rotating spherical shell compared to a quasigeostrophic model. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **76** (4), 046306.

Sagaut, Pierre 2006 *Large eddy simulation for incompressible flows*. Springer Science & Business Media.

Sanderse, Benjamin, Stinis, Panos, Maulik, Romit & Ahmed, Shady E 2025 Scientific machine learning for closure models in multiscale problems: A review. *Foundations of Data Science* **7** (1), 298–337.

Sapienza, Facundo, Bolibar, Jordi, Schäfer, Frank, Groenke, Brian, Pal, Avik, Boussange, Victor, Heimbach, Patrick, Hooker, Giles, Pérez, Fernando, Persson, Per-Olof & others 2024 Differentiable programming for differential equations: A review. *arXiv preprint* .

Schaeffer, Nathanaël & Cardin, Philippe 2005 Quasigeostrophic model of the instabilities of the Stewartson layer in flat and depth-varying containers. *Physics of Fluids* **17** (10).

Scott, Richard K & Dritschel, David G 2012 The structure of zonal jets in geostrophic turbulence. *Journal of Fluid Mechanics* **711**, 576–598.

Shen, Jie 1995 Efficient spectral-Galerkin method II. direct solvers of second-and fourth-order equations using Chebyshev polynomials. *SIAM Journal on Scientific Computing* **16** (1), 74–87.

Shishkina, Olga, Stevens, Richard JAM, Grossmann, Siegfried & Lohse, Detlef 2010 Boundary layer structure in turbulent thermal convection and its consequences for the required numerical resolution. *New journal of Physics* **12** (7), 075022.

Sirignano, Justin, MacArt, Jonathan F & Freund, Jonathan B 2020 DPM: A deep learning PDE augmentation method with application to large-eddy simulation. *Journal of Computational Physics* **423**, 109811.

Smagorinsky, Joseph 1963 General circulation experiments with the primitive equations: I. the basic experiment. *Monthly Weather Review* **91** (3), 99–164.

Sukoriansky, Semion, Dikovskaya, Nadejda & Galperin, Boris 2007 On the arrest of inverse energy cascade and the Rhines scale. *Journal of the Atmospheric Sciences* **64** (9), 3312–3327.

Sukoriansky, Semion, Galperin, Boris & Dikovskaya, Nadejda 2002 Universal spectrum of two-dimensional turbulence on a rotating sphere and some basic features of atmospheric circulation on giant planets. *Physical Review Letters* **89** (12), 124501.

Trabelsi, Chiheb, Bilaniuk, Olexa, Zhang, Ying, Serdyuk, Dmitriy, Subramanian, Sandeep, Santos, Joao Felipe, Mehri, Soroush, Rostamzadeh, Negar, Bengio, Yoshua & Pal, Christopher J 2018 Deep complex networks. In *International Conference on Learning Representations*.

Verhoeven, Jan & Stellmach, Stephan 2014 The compressional beta effect: A source of zonal winds in planets? *Icarus* **237**, 143–158.

Vinuesa, Ricardo & Brunton, Steven L 2022 Enhancing computational fluid dynamics with machine learning. *Nature Computational Science* **2** (6), 358–366.

Wicht, Johannes & Meduri, Domenico G 2016 A gaussian model for simulated geomagnetic field reversals. *Physics of the Earth and Planetary Interiors* **259**, 45–60.

Wordsworth, RD, Read, PL & Yamazaki, YH 2008 Turbulence, waves, and jets in a differentially heated rotating annulus experiment. *Physics of Fluids* **20** (12).

Yalla, Gopal R, Oliver, Todd A, Haering, Sigfried W, Engquist, Björn & Moser, Robert D 2021 Effects of resolution inhomogeneity in large-eddy simulation. *Physical Review Fluids* **6** (7), 074604.

Yan, Fei Er, Frezat, Hugo, Sommer, Julien Le, Mak, Julian & Otness, Karl 2025 Adjoint-based online learning of two-layer quasi-geostrophic baroclinic turbulence. *Journal of Advances in Modeling Earth Systems* **17** (7), e2024MS004857.

Yuval, Janni, O'Gorman, Paul A & Hill, Chris N 2021 Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters* **48** (6), e2020GL091363.

Yuval, Janni & O'Gorman, Paul A 2020 Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications* **11** (1), 3295.

Zhou, Zhideng, He, Guowei, Wang, Shizhao & Jin, Guodong 2019 Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Computers & Fluids* **195**, 104319.