Discovering Interpretable Structural Model Errors in Climate Models

Rambod Mojgani
Department of Mechanical Engineering
Rice University
6100 Main St, Houston, TX 77005
rm99@rice.edu

Ashesh Chattopadhyay
Department of Mechanical Engineering and
Rice University
6100 Main St, Houston, TX 77005
akc6@rice.edu

Pedram Hassanzadeh
Department of Mechanical Engineering and
Department of Earth, Environmental and Planetary Sciences
Rice University
6100 Main St, Houston, TX 77005
pedram@rice.edu

Abstract

Inaccuracies in the models of the Earth system, i.e., structural and parametric model errors, lead to inaccurate climate change projections. Errors in the model can originate from unresolved phenomena due to a low numerical resolution, as well as misrepresentations of physical phenomena or boundaries (e.g., orography). Therefore, such models lead to inaccurate short–term forecasts of weather and extreme events, and more importantly, long term climate projections. While calibration methods have been introduced to address for parametric uncertainties, e.g., by better estimation of system parameters from observations, addressing structural uncertainties, especially in an interpretable manner, remains a major challenge. Therefore, with increases in both the amount and frequency of observations of the Earth system, algorithmic innovations are required to identify interpretable representations of the model errors from observations. We introduce a flexible, general-purpose framework to discover interpretable model errors, and show its performance on a canonical prototype of geophysical turbulence, the two–level quasi–geostrophic system. Accordingly, a Bayesian sparsity–promoting regression framework is proposed, that uses a library of kernels for discovery of model errors. As calculating the library from noisy and sparse data (e.g., from observations) using convectional techniques leads to interpolation errors, here we use a coordinate-based multi–layer embedding to impute the sparse observations. We demonstrate the importance of alleviating spectral bias, and propose a random Fourier feature layer to reduce it in the proposed embeddings, and subsequently enable an accurate discovery. Our framework is demonstrated to successfully identify structural model errors due to linear and nonlinear processes (e.g., radiation, surface friction, advection), as well as misrepresented orography.

1 Introduction

Numerical models are the core of modern weather and climate predictions. Despite their success in both short–term weather and long–term climate predictions, they are inundated with a slew of heuristics, and often suffer from high uncertainties. Such uncertainties lead to a range of low confidence estimates in the predicted state variables [1]. Current efforts in expanding the resolution

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

and frequency of the observations has provided us with a unique opportunity to learn the missing components of our models, i.e., model discrepancy, model inadequacy, or hereafter, model error. Although aritificial neural network (ANN)-based and statistical error correction schemes have shown great promise [2–5], they are not interpretable, and therefore cannot be reliably used to correct climate models.

Growing availability of observation data has provided us with a unique opportunity to learn closed–form expressions of model errors. Such discoveries are expected to lead to generalizable correction of operational models, which are the backbones of any actionable policies. Although algorithmic plausibility of such frameworks are demonstrated for canonical chaotic flows [6, 7], other challenges have to be addressed. In this paper, we focus on scaling such methods to multi–scale turbulent flows, and representing sparse observation.

In this work, we scale MEDIDA, Model Error Discovery with Interpretability and Data–assimilation [7], a Bayesian regression to discover model error from a domain interpretable library of kernels generated by learning an interpolant of observed states. Our proposed approach is based on 3 main components, i) constructing the model error given a library of kernels, i.e., a system identification problem, and ii) selecting a parsimonious combination of candidates for an expressive model, i.e., a regression problem, and iii) data imputation to account for the missing observations using a coordinate-based multi-layer perceptron (MLP), i.e., an MLP-based compressed sensing problem. Our main contributions are twofold, i) we demonstrate the use of Bayesian regression to discover parsimonious, closed–form and interpretable model error in two-layer quasi–geostrophic (QG) system, as a simple climate model [8], and ii) we delineate spectral bias as the principle challenge of using ANNs in modeling of turbulent flows, and propose the use of random Fourier Feature (RFF) layer to effectively embed a highly multi–scale turbulent flow for the purpose of equation discovery in climate applications.

2 Preliminaries

Discovery of Model Error Consider a nonlinear partial differential equation (PDE), $\partial_t w(x,t) = g\left(w(x,t)\right)$, in a continuous domain, where $w(x,t)$ is the state variable in a spatiotemporal domain. We assume the governing operator, $g\left(w(x,t)\right)$, is unknown. Moreover, we assume to have a model of the truth, i.e., $\partial_t w(x,t) = f\left(w(x,t)\right)$. Our goal is to find an interpretable model of error, $h\left(w(x,t)\right) := g\left(w(x,t)\right) - f\left(w(x,t)\right)$, such that the corrected model, i.e., $\partial_t w(x,t) = f\left(w(x,t)\right) + h\left(w(x,t)\right)$, can represent the true system more accurately.

By discretizing of the imperfect and perfect models in space and time, one can discover the model error given two constitutive observations at $t_i - \Delta t$ and t_i, i.e., $\{\boldsymbol{w}^o(t_i - \Delta t), \boldsymbol{w}^o(t_i)\}_{i=1}^n$. The model is discovered as the minimizer of a regression problem, i.e.,

$$\boldsymbol{c}^* = \underset{\boldsymbol{c}}{arg\min} \left\| \Delta \boldsymbol{w} - \boldsymbol{\Phi}\left(\boldsymbol{w}^o\right) \boldsymbol{c} \right\|_2 , \tag{1}$$

where $\boldsymbol{\Phi}\left(\boldsymbol{w^o}\left(t_i\right)\right)$ is a library of kernels representing possible and interpretable forms of model errors (comprised of polynomials and derivatives of the state, see appendix D), and $\Delta \boldsymbol{w} := \left(\boldsymbol{w}^o\left(t_i\right) - \boldsymbol{w}^m\left(t_i\right)\right)/\Delta t$ is the vector of model errors, stacked over many samples [7]. The regression problem is solved using relevance vector machines (RVMs) [9], a sparsity–promoting Bayesian approach [10, 11, 7]. The model error discovery of eq. (1) is carried over short time intervals of Δt, however, we expect short time horizon correction of model errors can lead to correction of biases in climate models [12, 13].

Spectral bias and turbulence We train a coordinate–based MLP as an interpolant to handle noisy and sparse observations (for details, see section 3). The use of ANNs in physical sciences is implicitly based on the assumption that the trained ANN represents the states as well as their higher order derivatives, or equivalently, the spectra of the states is fully represented. Here, we demonstrate that the multi–scale nature of turbulent flows can become a fundamental hindrance in MLP embedding of the turbulent flows, and therefore contradicts the assumption. While ANNs are highly expressive and can fit to any arbitrary input-output maps, in the machine learning (ML) community, it is also well known that they tend to be biased towards learning low–wavenumber features of the data, a phenomena called spectral bias [14]. At the same time turbulent fluid flows are, by definition, multi–scale phenomena and span over a wide range of wavenumbers. Therefore, naive MLP embeddings of

turbulent flows lead to poor representations of the high-wavenumber features. Such features do not significantly contribute to the most commonly used loss functions, e.g., $\mathcal{L}_2$-norms, and therefore are often dismissed when the output of a network is compared with the truth in the physical domain.

We emphasize that the discussed spectral bias has broader implications well beyond the problem that is focus of this paper, and the remedies in section 3 has wide–ranging applications, e.g., the use of ANN–based automatic differentiation (AD) in [6, 15]. In the context of equation discovery, that is of interest here, the error in high wavenumbers directly affects the accuracy of the vector of model errors and library of kernels, especially the high–order derivatives.

3 Our Proposed Method

Application of coordinate–based MLPs, i.e., $\boldsymbol{w}(\boldsymbol{x}_i, t_i) = \mathcal{N}(\boldsymbol{x}_i, t_i)$, to embed solutions of the PDEs [15] in ANNs has lead to many novel innovations in science and engineering. The high speed and low cost of inference, and possibility of AD have made such tools increasingly attractive for climate models. While the early developments focused on inverse problems to solve for parametric uncertainties given measurements/observations [16], more recently, similar MLP architecture are used in equation–discovery frameworks to uncover full governing equations [6, 17], or model error [17]. In this paper, an MLP is trained on observed states variables. Subsequently, the trained network is queried at the spatiotemporal location of the missing observations, to estimate the state variable at that location, effectively constructing an interpolant. Moreover, AD or other methods can be leveraged to build a library of kernels required for equation discovery (for a comparison, see appendix D). To alleviate the spectral bias in training of the MLP surrogate, we employ a non–trainable RFF layer [18]. This is especially crucial in equation–discovery of turbulent flows, which are multi–scale in nature.

4 Results

Consider the two-layer QG system of equations, a model of the mid–latitude baroclinic dynamics of the Earth's atmosphere [8], $\frac{\partial q_k}{\partial t} = -\alpha J(\psi_k, q_k) - \frac{(-1)^k}{\tau_{d_1}}(\psi_1 - \psi_2) + \frac{(-1)^k}{\tau_{d_2}}(\psi_R) - \delta_{k2} f$, where $k = 1$ and $k = 2$ represent upper and lower layers, respectively. The details of the model and each of the parameters are summarized in appendix A. We consider a vector of parameters to represent the system, i.e., $\boldsymbol{c}_s = [\alpha, 1/\tau_{d_1}, 1/\tau_{d_2}, 1/\tau_f, \beta]$, as well as the orography in fig. 1a, and a linear friction term, $f = \nabla^2\psi_2/\tau_f$. The imperfect model and corrected model errors are respectively measured as $\varepsilon_m = \|\boldsymbol{c}_s - \boldsymbol{c}_m\|_2/\|\boldsymbol{c}_s\|_2$ and $\varepsilon^* = \|\boldsymbol{c}_s - \boldsymbol{c}^*\|_2/\|\boldsymbol{c}_s\|_2$, where $\boldsymbol{c}_m$ and $\boldsymbol{c}^*$ are the vectors of coefficients in the imperfect model and the corrected model, respectively.

4.1 Discovery of Model Errors: Full data

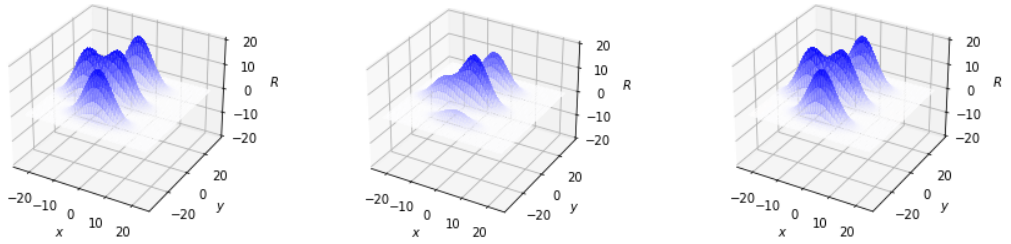
We compare 8 different cases of parametric and structural errors as shown in Table 1. Cases 1-3, and 6-8 represent model error in linear terms, cases 4 and 5 show error in the nonlinear terms. In all cases, the proposed approach has reduced the model error significantly. Figure 1b depicts an error in representation of orography. Using the proposed approach the orography is as shown in fig. 1c. Moreover, consider a case where the friction is modeled with quadratic drag, $f = [\partial_x(|\nabla\psi_2|\,\partial_x\psi_2) + \partial_y(|\nabla\psi_2|\,\partial_y\psi_2)]/\tau_f$. The proposed method successfully removes the incorrect term and discovers the correct linear form with less than 0.1% error in τ_f.

4.2 Discovery of Model Errors: Sparse observation

In fig. 2, we compare an MLP embedding without and with an RFF layer, which is employed to improve accuracy of embedding in a wider range of wavenumbers. The RFF layer consequently improves the accuracy of higher–order derivatives (figs. 2b to 2c), where the effect of spectral bias is clearly more significant. The effectiveness of the proposed approach in discovery of model error from sparse data is shown in table 2.

Table 1: The imperfect models of the QG system corrected using the proposed method. The true system is $\boldsymbol{c}_s = [1.00, 0.01, 0.01, 0.07, 0.196]$. The underscored values represent the model errors.

	Imperfect model		Corrected model	
	$[\alpha, 1/\tau_{d_1}, 1/\tau_{d_2}, 1/\tau_f, \beta]$	$\varepsilon_m(\%)$	$[\alpha, 1/\tau_{d_1}, 1/\tau_{d_2}, 1/\tau_f, \beta]$	$\varepsilon^*(\%)$
1	$[1.00, \underline{0.11}, \underline{0.11}, 0.07, 0.196]$	16.99	$[1.00, 0.01, 0.01, 0.07, 0.196]$	0.04
2	$[1.00, \underline{0.11}, \underline{0.11}, \underline{0.37}, 0.196]$	16.99	$[1.00, 0.010, 0.010, 0.072, 0.196]$	0.04
3	$[1.00, \underline{0}, \underline{0}, 0.07, 0.196]$	33.91	$[1.00, 0.01, 0.01, 0.072, 0.196]$	0.19
4	$[\underline{0.50}, 0.01, 0.01, 0.07, 0.196]$	49.06	$[1.00, 0.01, 0.01, 0.07, 0.196]$	0.13
5	$[\underline{0}, \underline{0.11}, \underline{0.11}, \underline{0.37}, 0.196]$	99.58	$[1.00, 0.010, 0.010, 0.070, 0.196]$	0.34
6	$[1.00, 0.01, 0.01, \underline{0.37}, 0.196]$	29.37	$[1.00, 0.01, 0.01, 0.068, 0.196]$	0.14
7	$[1.00, 0.01, 0.01, \underline{0}, 0.196]$	6.85	$[1.00, 0.01, 0.01, 0.069, 0.196]$	0.01
8	$[1.00, 0.01, \underline{0}, 0.07, 0.196]$	0.98	$[1.00, 0.01, 0.01, 0.070, 0.196]$	< 0.01
9	$[1.00, 0.01, 0.01, 0.07, \underline{0}]$	19.23	$[1.00, 0.01, 0.01, 0.070, 0.196]$	< 0.01

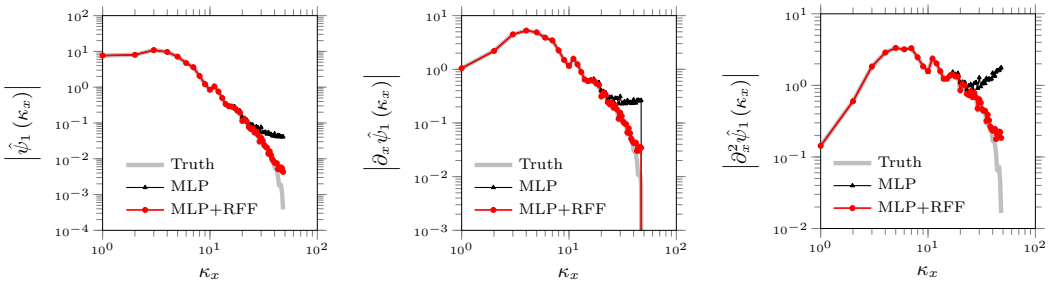


(a) True orography (b) Incorrect orography (c) Corrected orography

Figure 1: The orography of the true system (a), and the imperfect model (b). Our proposed approach successfully discovers the corrected orography as shown in (c).

Table 2: The imperfect model of case 2 (from Table 1) corrected using the proposed MLP-based interpolant in the presence of sparse observations. The sparsity, in percentage, shows the ratio of the missing observations to full domain. The RFF layer enables the discovery of the model error.

ANN-Sparsity	Imperfect model		Corrected model	
	$[1/\tau_{d_1}, 1/\tau_{d_2}, 1/\tau_f]$	$\varepsilon_m(\%)$	$[1/\tau_{d_1}, 1/\tau_{d_2}, 1/\tau_f]$	$\varepsilon^*(\%)$
MLP-0%	$[\underline{0.11}, \underline{0.11}, \underline{0.37}]$	10	$[0.11, 0.11, -7.03]$	> 100
MLP-RFF-0%	$[\underline{0.11}, \underline{0.11}, \underline{0.37}]$	10	$[0.01, 0.01, 0.07]$	< 0.01
MLP-RFF-5%	$[\underline{0.11}, \underline{0.11}, \underline{0.37}]$	10	$[0.010 \pm 0.002, 0.010 \pm 0.002, 0.072]$	< 0.01



(a) Fourier spectral of ψ_1 (b) Fourier spectral of $\partial_x \psi_1$ (c) Fourier spectral of $\partial_x^2 \psi_1$

Figure 2: Fourier spectra of the state variable (truth), classic MLP embedding of the truth, and the effect of the RFF layer. Similarly, for the first and second order spatial derivatives (zonal direction).

5 Conclusions and future work

In this paper, we proposed a framework to discover physics–based, interpretable, and closed–form structural and parametric model errors. We demonstrated and addressed the challenge of embedding of a turbulent flow in a coordinate-based MLP to impute the missing observations from sparse data. We emphasized on alleviating spectral bias by using an RFF layer in an equation discovery framework. Using the proposed method, we aim to leverage the observational data to correct existing biases in the climate models, and thus improve the confidence in the climate change projections. Future work will extend the method to transitory regimes, and the problem of noisy sparse observations.

References

[1] V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. YelekÇi, R. Yu, and B. Zhou, editors. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Technical Summary*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.

[2] Christopher S. Bretherton, Brian Henn, Anna Kwa, Noah D. Brenowitz, Oliver Watt-Meyer, Jeremy McGibbon, W. Andre Perkins, Spencer K. Clark, and Lucas Harris. Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. *Journal of Advances in Modeling Earth Systems*, 14(2):e2021MS002794, 2022.

[3] Alban Farchi, Patrick Laloyaux, Massimo Bonavita, and Marc Bocquet. Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, page qj.4116, July 2021. ISSN 0035-9009.

[4] Oliver Watt-Meyer, Noah D. Brenowitz, Spencer K. Clark, Brian Henn, Anna Kwa, Jeremy McGibbon, W. Andre Perkins, and Christopher S. Bretherton. Correcting weather and climate models by machine learning nudged historical simulations. *Geophysical Research Letters*, 48(15):e2021GL092555, 2021.

[5] Nan Chen and Yinling Zhang. A causality-based learning approach for discovering the underlying dynamics of complex systems from partial observations with stochastic parameterization. *preprint arXiv:2208.09104*, 2022.

[6] Zhao Chen, Yang Liu, and Hao Sun. Physics-informed learning of governing equations from scarce data. *Nature Communications*, 12(1):6136, December 2021. ISSN 2041-1723.

[7] Rambod Mojgani, Ashesh Chattopadhyay, and Pedram Hassanzadeh. Discovery of interpretable structural model errors by combining Bayesian sparse regression and data assimilation: A chaotic Kuramoto–Sivashinsky test case. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(6):061105, 2022.

[8] Norman A. Phillips. Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus*, 6(3):273–286, 1954.

[9] Michael E Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1(June):211–244, 2001.

[10] Sheng Zhang and Guang Lin. Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2217): 20180305, 2018.

[11] Laure Zanna and Thomas Bolton. Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, 47(17):e2020GL088376, 2020.

[12] T. N. Palmer and Antje Weisheimer. Diagnosing the causes of bias in climate models – why is it so hard? *Geophysical & Astrophysical Fluid Dynamics*, 105(2-3):351–365, 2011.

[13] M. J. Rodwell and T. N. Palmer. Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, 133(622):129–146, 2007.

[14] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5301–5310. PMLR, 09–15 Jun 2019.

[15] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics–informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019. ISSN 00219991.

[16] Maziar Raissi, Alireza Yazdani, and George Em Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

[17] Kadierdan Kaheman, Steven L. Brunton, and J. Nathan Kutz. Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *Machine Learning: Science and Technology*, 3(1):015031, March 2022. ISSN 2632-2153.

[18] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

[19] Basile Gallet and Raffaele Ferrari. The vortex gas scaling regime of baroclinic turbulence. *Proceedings of the National Academy of Sciences*, 117(9):4491–4497, 2020.

[20] Nicholas J. Lutsko, Isaac M. Held, and Pablo Zurita-Gotor. Applying the fluctuation–dissipation theorem to a two-layer model of quasigeostrophic turbulence. *Journal of the Atmospheric Sciences*, 72(8):3161 – 3177, 2015.

[21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *preprint arXiv:1412.6980*, 2014.

[22] Matthew E. Levine and Andrew M. Stuart. A framework for machine learning of model error in dynamical systems. *arXiv preprint arXiv:2107.06658*, 2021.

[23] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.

[24] Atılım Günes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: A survey. *Journal of Machine Learning Research*, 18(1): 5595–5637, January 2017. ISSN 1532-4435.

[25] Fourier features let networks learn high frequency functions in low dimensional domains, 2022. URL `https://bmild.github.io/fourfeat/`.

A Numerical solver

In this work, we consider the two–layer quasi–geostrophic (QG) equations [8],

$$\frac{\partial q_k}{\partial t} = -\alpha J\left(\psi_k, q_k\right) - \frac{(-1)^k}{\tau_{d_1}}\left(\psi_1 - \psi_2\right) + \frac{(-1)^k}{\tau_{d_2}}\left(\psi_R\right) - \delta_{k2} f \tag{2}$$

where $k = 1$ and $k = 2$, respectively, represent upper and lower layers of atmosphere, $q_k = \nabla^2\psi_k + (-1)^k\left(\psi_1 - \psi_2\right) + \beta y + \delta_{k2} R\left(x, y\right)$ is the potential vorticity, ψ_k is the streamfunction. The Jacobian term is $J(\psi_k, q_k) = \frac{\partial q_k}{\partial x}\frac{\partial \psi_k}{\partial y} - \frac{\partial q_k}{\partial y}\frac{\partial \psi_k}{\partial x}$ and α is tuning parameter for the sake of our numerical experiments. τ_{d_1} and τ_{d_2} are Newtonian relaxation time scale broken into two terms for the purpose of our experiments, f is a model for drag, e.g., linear drag, $f = \nabla^2\psi_2/\tau_f$, or quadratic drag [19], $f = \left[\partial_x\left(\left|\nabla\psi_2\right|\partial_x\psi_2\right) + \partial_y\left(\left|\nabla\psi_2\right|\partial_y\psi_2\right)\right]/\tau_f$, where τ_f is a Rayleigh friction time scale, and for $k = 2$, $\delta_{k2} = 1$, and is zero otherwise.

The topographic/orographic features are represented in $R\left(x, y\right)$ and acts on the lower layer. The orography is defined as summation over Gaussian distributions centered at $[x_i, y_i]$ with height of r_i, and variance of σ_i^2., i.e.,

$$R\left(x, y\right) = \sum_{i=1}^{N_r} r_i \mathcal{N}\left(\begin{bmatrix} x_i, \\ y_i \end{bmatrix}, \sigma_i^2\right). \tag{3}$$

The QG equations of (2) are solved using the implementation by Lutsko et al. [20]. All the parameters are the ones introduced in [20], unless otherwise stated. The non-dimensionalized zonal and meridional widths of size $L_x = 46$ and $L_y = 68$ on a Fourier grid of size 96×192, respectively. A baroclinically unstable jet is set in the upper layer such that $\partial\psi_R/\partial y = \mathrm{sech}^2\left(y/\sigma\right)$. A typical solution is shown in fig. 3.

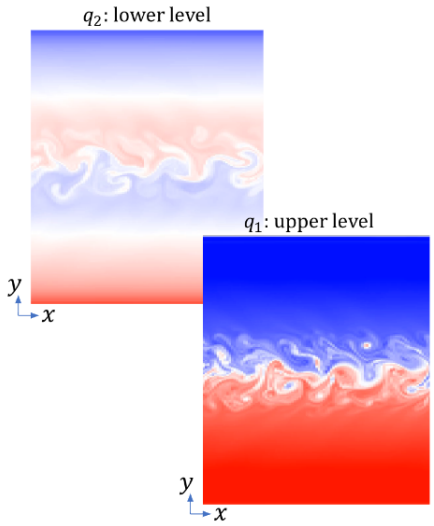


Figure 3: A typical snapshot of potential vorticity (q).

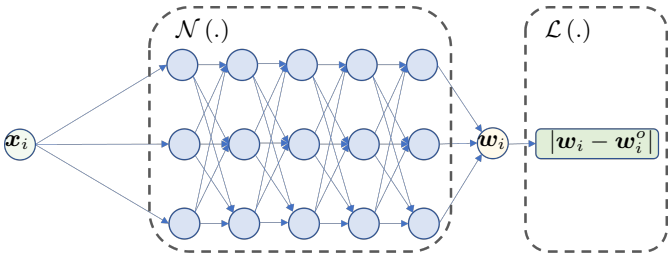


Figure 4: The architecture of coordinate–based MLP used as an interpolant to impute missing observations (baseline architecture).

B Data

The data for our experiments are reproducible from solution of QG equations discussed in appendix A. To ensure stationary turbulence, the spin–up period is dismissed (equivalent to 2500 Earth days).

C Details of the network (imputing the missing observations)

In this paper, a coordinated–based multi-layer perceptron (MLP) is proposed as an interpolant to impute the missing observations. The baseline architecture is described in appendix C.1. The random Fourier Feature (RFF) layer to alleviate the effect of spectral bias is described in appendix C.2. Both architectures are trained using the available observations, and are queried to impute the missing observations.

C.1 Coordinate-based multi layer perceptron

The output of the aritificial neural network (ANN) given i^{th} coordinate of a spatiotemporal grid, $\mathcal{N}(\boldsymbol{u})$, is the local and instantaneous state parameter, $\boldsymbol{w}(\boldsymbol{x}_i, t_i)$, i.e.,

$$\boldsymbol{w}(\boldsymbol{x}_i, t_i) = \mathcal{N}(\boldsymbol{u}) = \phi_m\left(\boldsymbol{W}_m \phi_{m-1}\left(\boldsymbol{W}_{m-1} \cdots \phi_1\left(\boldsymbol{W}_1 \boldsymbol{u} + \boldsymbol{b}_1\right) \cdots + \boldsymbol{b}_{m-1}\right) + \boldsymbol{b}_m\right), \tag{4}$$

where the input vector is the concatenation of the spatial and temporal location, i.e., $\boldsymbol{u} = [\boldsymbol{x}_i, t_i]^\top \in \mathbb{R}^{d+1}$, d is the dimension of physical space, $\phi_i(.)$ is the activation function at the i^{th}-layer, $\boldsymbol{W}_1 \in \mathbb{R}^{w \times (d+1)}$, $\boldsymbol{W}_i \in \mathbb{R}^{w \times w}, \forall i \in \{2, \cdots, m-1\}$, and $\boldsymbol{W}_w \in \mathbb{R}^{d \times w}$ are the weights, and $\boldsymbol{b}_i \in \mathbb{R}^w, \forall i \in \{1, \cdots, m-1\}$, and $\boldsymbol{b}_m \in \mathbb{R}^d$ are biases.

In this paper, the output vector is $\boldsymbol{w}(x_i, y_i, t_i) = [\psi_1(x_i, y_i, t_i), \psi_2(x_i, y_i, t_i)]$, the input vector is $\boldsymbol{u} = [x_i, y_i, t_i]^\top$. The network is of size $m = 5$, $w = 1000$, and $d = 2$. Adam optimizer is used in all the experiments [21].

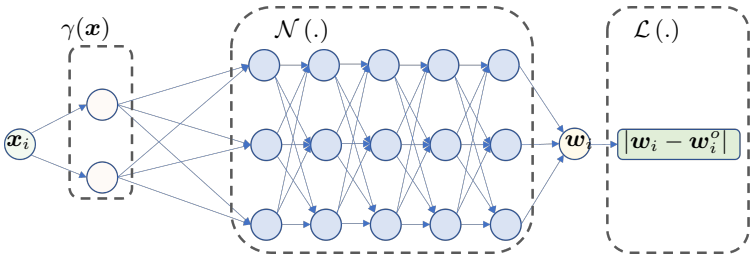


Figure 5: The proposed architecture to alleviate spectral bias in embedding of turbulent flows. A coordinate–based MLP with an RFF layer used as an interpolant to impute missing observations with features ranging a broad wavenumber spectrum (MLP+RFF).

C.2 Random Fourier Feature layer

The RFF layer alleviates the spectral bias while training by perturbing the coordinate-based input. The layer is constructed as

$$\gamma(\boldsymbol{u}) = \left[\cos\left(2\pi \mathbf{B}^\top \boldsymbol{u}\right), \sin\left(2\pi \mathbf{B}^\top \boldsymbol{u}\right)\right]^\top, \tag{5}$$

where each entry in $\mathbf{B} \in \mathbb{R}^{m \times d}$ is sampled from $\mathcal{N}\left(0, \sigma^2\right)$, and σ is a hyperparameter. We assume no prior on the frequency/wavenumber spectrum and draw samples from a Gaussian distribution. The output of the layer is then passed through the ANN architecture, i.e.,

$$\boldsymbol{w}(\boldsymbol{x}_i, t_i) = \mathcal{N}\left(\gamma\left(\boldsymbol{x}_i, t_i\right)\right). \tag{6}$$

The hyperpameter is chosen such that the statistics of the output matches the target flow, implicitly assuming that the flow has reached its statistical equilibrium, i.e., stationary turbulence. In our experiments, $\sigma = 1$ is shown to reach the lowest representation error.

D Library of kernels

While statistical and radial kernels are expressive and can provide accurate models of the error [22, 5], they often do not provide any interpretations of the missing physics.

The library consists of the candidate forms of the baroclinic-type jet, state variables and their first and second derivatives in both zonal and meridional directions, i.e.,

$$\left\{\psi_R, \psi_1, \psi_2, \frac{\partial \psi_1}{\partial x}, \frac{\partial \psi_1}{\partial y}, \frac{\partial \psi_2}{\partial x}, \frac{\partial \psi_2}{\partial y}, \nabla^2 \psi_1, \nabla^2 \psi_2\right\},$$

where $\nabla^2(.) := \nabla.\nabla(.)$. The derivative of the orography as expected to appear in (2), $R_i\left(x_i, y_i\right)$,

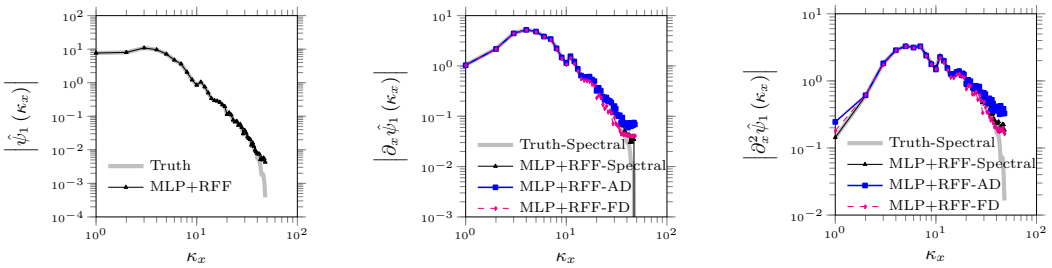
$$\left\{\frac{\partial \psi_2}{\partial y}\frac{\partial R_i}{\partial x}, \frac{\partial \psi_2}{\partial x}\frac{\partial R_i}{\partial y}\right\}.$$

Moreover, the following nonlinear terms are included,

$$\left\{\frac{\partial \psi_i}{\partial y}\frac{\partial \psi_j}{\partial x}, \frac{\partial \psi_i}{\partial y}\frac{\partial \nabla^2 \psi_j}{\partial x}, \frac{\partial \psi_i}{\partial x}\frac{\partial \nabla^2 \psi_j}{\partial y}\right\}, i, j \in \{1, 2\}.$$

Given the state variables, the derivatives in the library of kernels can be computed using any of the well–known arbitrary differentiation approaches, e.g., spectral methods, finite differences (FDs), and differentiation of a polynomial fit [23]. MLP embeddings of the state variables enable automatic differentiation (AD) [24] as another method of calculating the derivatives, where the output of the trained network can be differentiated with respect to the input coordinates.

AD is often regarded as analytical approximation of the derivatives [6], however, we emphasize that its accuracy depends on the accuracy of the embedding. Moreover, the computational cost of AD can become a bottleneck, especially in higher order derivatives [6] and large libraries. In fig. 6, we compare the accuracy of spectral derivative of the truth with derivatives of output of the trained MLP calculated using Fourier spectral, AD, and FD. It can be seen that AD, compared to the other two methods, is less accurate in high wavenumbers. Therefore, throughout this paper, we have used Fourier spectral derivatives to calculate the kernels.



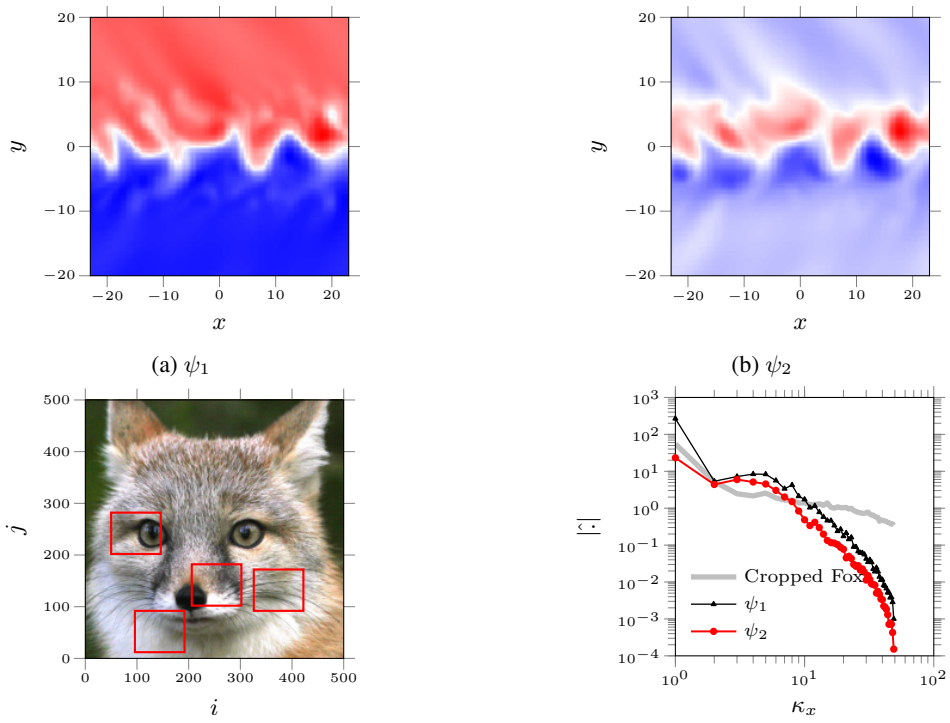
(a) Fourier spectral of ψ_1 (b) Fourier spectral of $\partial_x\psi_1$ (c) Fourier spectral of $\partial_x^2\psi_1$

Figure 6: Comparison of different methods of differentiation. A well–tuned RFF layer is used for theMLP embedding of a snapshot of ψ_1 as the truth. The first and second order spatial derivatives (zonal direction) are calculated using Fourier spectral derivatives, AD, and second order central finite difference FD.

E Fourier spectrum of turbulent flows

To demonstrate the significance of spectral bias in coordinate-based MLP embeddings of turbulent flows, the Fourier spectrum of a typical photo is compared to that of a solution of QG equations (fig. 7). The Fourier spectra is plotted as the mean of the magnitude of zonal Fourier transform of the states, i.e., the streamfunctions (ψ_1 and ψ_2 in fig. 7-(a-b)), and the cropped frames from the natural image (in fig. 7-(c)). The solution snapshots are of size 96×84 (in the jet region, see appendix A). Similarly, the natural image is cropped to the similar dimension of 96×84.

While the magnitude of the Fourier modes of the image decays $\mathcal{O}\left(10^2\right)$ over the range of the wavenumbers, the magnitude of the Fourier modes of ψ_1 and ψ_2 decay $\mathcal{O}\left(10^5\right)$ over the same range of the wavenumbers. Empirically, considering the spectral bias innate to training of MLPs, it is a significantly harder task to faithfully (in the full range of spectral domain) learn the highly multi–scale nature of turbulence compared to natural images.



(a) ψ_1

(b) ψ_2

(c) Natural image (from [25])

(d) Fourier spectral

Figure 7: Typical streamfunctions of the QG equations (a-b), and a typical natural image (c) are compared in the Fourier domain (d). The red boxes in (c) are randomly selected and are of the same dimensions of (a-b).