Multiscale stochastic parameterization with deep Mori-Zwanzig formalism

Juan Nathaniel
Columbia University
jn2808@columbia.edu

Carla Roesch
Columbia University
cmr2293@columbia.edu

Jatan Buch
Aeolus Labs
jatan@aeolus.earth

Kaitlyn Loftus
Columbia University
kml2193@columbia.edu

Daniel Giles
University College London
d.giles@ucl.ac.uk

Derek DeSantis
Los Alamos National Laboratory
ddesantis@lanl.gov

Pierre Gentine
Columbia University
pg2328@columbia.edu

Abstract

We propose a deep learning implementation of the Mori–Zwanzig formalism, termed DMZ, as a principled multiscale, memory-aware, and stochastic modeling framework for developing subgrid-scale parameterization schemes in global circulation model. Here, robust multiscale representations are needed for stable and skillful long-term forecasts but remain limited by computational capacity. Coupled online tests with DMZ-based parameterization demonstrate stable multiyear simulation runs that maintain physically consistent interannual variability. When evaluated against satellite observations, DMZ significantly reduces precipitation bias, particularly over the Pacific Ocean where dominant modes of variability such as El Niño–Southern Oscillation emerge. These results highlight DMZ as a promising modeling paradigm that provides a principled stochastic, memory-based parameterization scheme to improve the fidelity of Earth system simulations.

1 Introduction

Physical systems often involve complex processes that interact across a wide range of spatial and temporal scales. Resolving all of these scales simultaneously still exceeds even today's most powerful computational capabilities [1, 2]. In fluid dynamics, for example, simulations typically span multiple levels of fidelity, from direct numerical simulation (DNS), to large-eddy simulation (LES), and even coarser models such as the general circulation model (GCM), each progressively approximating the effects of unresolved dynamics through subgrid-scale parameterization [3–5]. This creates a trade-off: whereas high-resolution emulators such as convection-resolving models (CRM) can capture multiscale processes more realistically, they remain too computationally expensive [6].

To address this limitation, stochastic parameterizations, inspired by early works of Hasselmann [7] and Leith [8], have been explored. Here, unresolved fast processes are approximated as random noise, effectively reducing the model dimensions while maintaining essential slow dynamics [9, 10]. More recently, hybrid approaches (e.g., [11]) combining machine learning (ML) with coarse-grained high-resolution simulations have emerged. These hybrid approaches appear capable of correcting systematic biases and thus show promise for improving precipitation and extreme statistics and

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Machine Learning and the Physical Sciences.

enabling long-term, high-fidelity simulations. However, ML-based parameterizations are often shown to be unstable, physically inconsistent, and typically rely on ad-hoc designs to represent memory and stochastic variability [12].

In this study, we propose a deep learning implementation of the Mori–Zwanzig (MZ) formalism to develop principled *stochastic*, *memory-based* parameterization schemes, termed **DMZ**. The formalism was originally developed by Mori [13] and Zwanzig [14] as projection-based methods to express the effect of both the resolved and unresolved variables in terms of the resolved ones. As a result, one can prescribe a set of governing equations for the resolved variables.

Recently, the MZ formalism has inspired a broad range of machine learning approaches that incorporate memory effects to enhance the solution of PDEs beyond purely Markovian models [15, 16]. These methods explicitly account for missing time delays [17] and have been successfully applied across disciplines, including materials science [18] and climate modeling [19].

Building on this extensive body of work, we extend MZ-inspired and memory-augmented modeling frameworks to subgrid-scale parameterization in climate models. Specifically, we apply DMZ to estimate subgrid temperature and specific humidity variability from coarse-grained outputs of high-resolution cloud-resolving models, following the approach of [11]. This complements previous studies leveraging the MZ formalism for stochastic parameterizations and non-Markovian representations in GCMs [20–23], including proposed applications in NOAA's next-generation global forecasting system [24].

Like [11], we achieve stable multiyear online simulations, demonstrating that such data-driven parameterizations can be robustly integrated within an interactive GCM. Our presented initial results demonstrate a promising proof-of-concept for incorporating data-driven memory terms into subgrid parameterizations, paving the way for future large-scale validation and comparison against existing stochastic schemes.

2 Deep Mori–Zwanzig formalism

Formalism We consider a discrete dynamical system of the form:

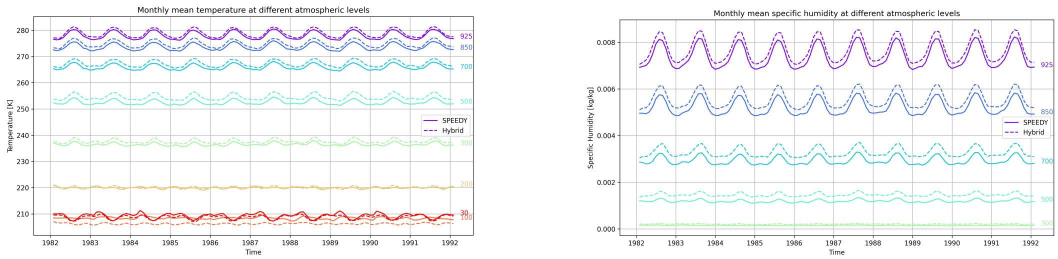
$$\boldsymbol{\phi}(t+1) = \mathbf{F}(\boldsymbol{\phi}(t)), \tag{1}$$

where $\boldsymbol{\phi} \in \mathbb{R}^N$ is the state variable and $\mathbf{F} : \mathbb{R}^N \to \mathbb{R}^N$ is the flow map. Following notation from [25], the MZ formalism describes the exact evolution of the states given an initial condition $\boldsymbol{\phi}(0)$ through the Generalized Langevin equation (GLE), where the nonlinear dynamics is decomposed into Markov, memory, and noise or orthogonal terms. The discretized GLE has the form:

$$\mathbf{g}_{n+1}(\boldsymbol{\phi}(0)) = \underbrace{\Omega^{(0)}\mathbf{g}_n(\boldsymbol{\phi}(0))}_{\text{Markov}} + \underbrace{\sum_{l=1}^{n} \Omega^{(l)}\mathbf{g}_{n-l}(\boldsymbol{\phi}(0))}_{\text{Memory}} + \underbrace{\mathbf{W_n}(\boldsymbol{\phi}(0))}_{\text{Orthogonal}}, \tag{2}$$

where $\mathbf{g}_n : \mathbb{R}^N \to \mathbb{R}^M$ is a feature vector, e.g., resolved variables or observables, such that $\mathbf{g}_n(\boldsymbol{\phi}(0)) \equiv \mathbf{g}(\mathbf{F}^n(\boldsymbol{\phi}(0)))$. The formalism stipulates that the dynamics of the resolved variables depend not only on their instantaneous Markovian contributions but also on their history through the memory term. The orthogonal dynamics are unobserved $\mathbf{W_n}$ and are commonly approximated as a stochastic process, such as Brownian motion, which we adopt in this work.

Approximation Although Equation 2 is exact, estimating the Markov and memory operators, i.e., $\{\Omega^{(0)}, \cdots, \Omega^{(n)}\}$, is non-trivial. One approach is to apply *Mori's linear projection* to approximate. However, the quality of the resulting predictions depends almost entirely on the choice of resolved variables $\mathbf{g}$ and their linear combinations. Alternatively, *Zwanzig's nonlinear projection* satisfies a particular optimality condition, but requires access to conditional expectations, which are often intractable in finite-data, high-dimensional settings. As such, we implement a data-driven implementation from [25], which shows that any regression analysis can be treated as projection operation in MZ. In short, the operators can be interpreted as regressors that map the independent quantities on the right-hand side (RHS; $[\mathbf{g}_1, \cdots, \mathbf{g}_n]$) to the quantities of interest on the left-hand side (LHS; $\mathbf{g}_{n+1}$).



(a) Temperature seasonality

(b) Specific humidity seasonality

Figure 1: Online DMZ-hybrid run seasonality checks for a 10-year run across pressure levels in hPa (different colors) shows stability and consistency similar to the SPEEDY control run. We remove top-of-atmosphere trends for specific humidity due to low water vapor mixing ratios.

Deep learning Mori–Zwanzig operators In this work, these regressors are implemented as deep learning models, parameterized by learnable weights and trained to minimize the mean-squared error (MSE). This yields a family of data-driven approximations, $\{\Omega^{(0)}_{\theta_0}, \cdots, \Omega^{(n)}_{\theta_n}\}$, which serve as tractable surrogates for the projection operators. We defer the complete treatment and derivation to [25]. This design enables the model to capture both short-term nonlinear interactions and longer-range memory effects within a unified framework. The resulting DMZ formulation can be viewed as a learned discrete convolution, where the kernel weights $\{\Omega^{(0)}_{\theta_0}, \cdots, \Omega^{(n)}_{\theta_n}\}$ encode the effective memory of the system.

To ensure physical relevance, the learned operators are trained on coarse-grained fields from high-resolution cloud-resolving simulations and evaluated in an online coupled setting with the SPEEDY GCM. In this context, $\Omega^{(0)}_{\theta_0}$ represents the instantaneous (Markovian) subgrid tendency, while $\Omega^{(l)}_{\theta_l}$ for $l > 0$ model the influence of prior states, effectively capturing unresolved variability due to time-lagged dynamic and thermodynamic processes. The residual term $\mathbf{W}_n$ acts as a stochastic closure for remaining variability and provides a measure of model uncertainty. This formulation preserves the interpretability of the MZ decomposition while allowing for purely data-driven learning.

3 Experiments

We present the experimental setup to demonstrate the utility of DMZ for developing memory-based, stochastic parameterization schemes. Note, the notation used here deviates slightly from earlier formalism in Equation 2: while GLE is expressed in terms of dynamical evolution, here we recast the problem in a regression form following [25]. Without loss of generality, we define the independent term (RHS) as $\mathbf{x}(t)$ and the dependent term (LHS) as $\mathbf{y}(t)$, following a canonical regression setup, i.e., $\mathbf{y}(t) = \mathrm{DMZ}[\mathbf{x}(t-s), \ldots, \mathbf{x}(t)]$, where s is the prescribed memory length.

Training We use high-resolution nested CRM simulations from the Met Office Unified Model (MetUM) [26, 27] run by [11] as training data. The MetUM set-up involves 80 nested limited-area models (LAM) that are forced by a global model. Each nested LAM has a grid resolution of 1.5 km. To generate training data, each LAM is divided into patches, each patch contains 224 $\times$ 224 grid points and is 336 km wide to approximately match the resolution of the SPEEDY grid (333 km). The nested simulations are run with a 6-hourly sampling frequency over 10 days. The subgrid variability of temperature ($\boldsymbol{\sigma}_T$) and specific humidity ($\boldsymbol{\sigma}_q$) are calculated over each patch to form the training data set. Using these CRM simulations, we estimate $\{\Omega^{(0)}_{\theta_0}, \cdots, \Omega^{(n)}_{\theta_n}\}$ by regressing $\boldsymbol{\sigma}_T$ and $\boldsymbol{\sigma}_q$, across $l = 8$ pressure levels:

$$\mathbf{y}(t) = [\boldsymbol{\sigma}_T(t), \quad \boldsymbol{\sigma}_q(t)]^\top \in \mathbb{R}^{2l}; \quad \boldsymbol{\sigma}_T(t), \boldsymbol{\sigma}_q(t) \in \mathbb{R}^l. \tag{3}$$

The independent variables, or predictors, are then given by:

$$\mathbf{x}(t) = \left[\boldsymbol{\mu}_T(t), \quad \boldsymbol{\mu}_q(t), \quad p_{\mathrm{sfc}}(t), \quad \mu_{\mathrm{orog}}, \quad \sigma_{\mathrm{orog}}\right]^\top \in \mathbb{R}^{2l+3}, \tag{4}$$

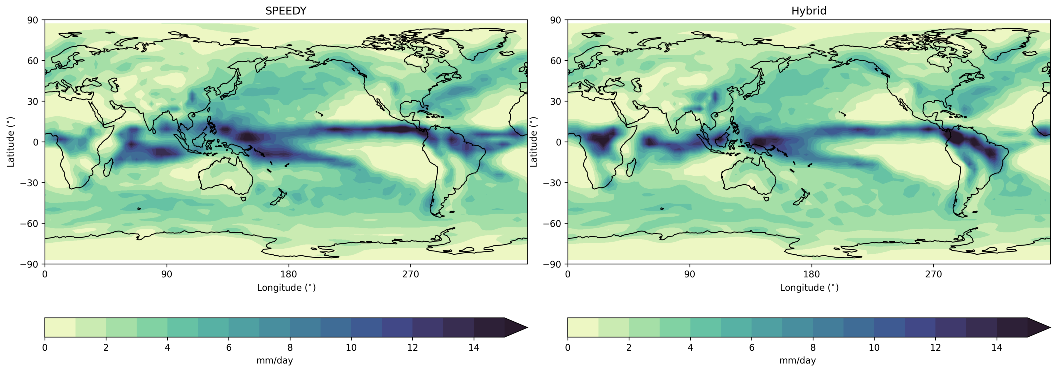


Figure 2: Comparing precipitation climatology for both the SPEEDY control and DMZ-hybrid online runs. The latter shows physical consistency similar to that of the control run.

where $\boldsymbol{\mu}_T(t), \boldsymbol{\mu}_q(t) \in \mathbb{R}^l$ are the subgrid means of temperature and specific humidity, $p_{\mathrm{sfc}}(t) \in \mathbb{R}$ is surface pressure, and $\mu_{\mathrm{orog}}, \sigma_{\mathrm{orog}} \in \mathbb{R}$ are the mean and variability of orography.

For each deep regressor, Ω_θ, we use a 2-layer multi-layer perceptron (MLP), with `SiLU` activation function [28], and 128 hidden dimensions. Training is performed with a learning rate of 10^{-4}, batch size of 32, over 200 epochs; the best performing model on the 80-20 validation split is saved. We prescribe the memory length, $s = 2$, as a proof-of-concept. Future ablation will examine the relative contribution of different memory length.

Inference We apply the learned operators on a sequence of inputs $[\mathbf{x}(t-s), \ldots, \mathbf{x}(t)]$, rather than a snapshot as was done in [11], to estimate $\mathbf{y}(t)$ in an online setting. We use SPEEDY (Simplified Parametrizations, primitivE-Equation DYnamics) as our host GCM [29], where DMZ corrects SPEEDY's resolved updates to temperature $\boldsymbol{\mu}_T^{\mathrm{SPEEDY}}(t)$ and specific humidity $\boldsymbol{\mu}_q^{\mathrm{SPEEDY}}(t)$ to account for subgrid effects every 6 hours. The correction applied at the end of each step then serves as input for subsequent integration, as given by:

$$\begin{bmatrix} \boldsymbol{\mu}_T(t) \\ \boldsymbol{\mu}_q(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_T^{\mathrm{SPEEDY}}(t) \\ \boldsymbol{\mu}_q^{\mathrm{SPEEDY}}(t) \end{bmatrix} + \begin{bmatrix} \hat{\boldsymbol{\sigma}}_T(t) \\ \hat{\boldsymbol{\sigma}}_q(t) \end{bmatrix} \odot \boldsymbol{\xi}(t), \tag{5}$$

where $\boldsymbol{\xi}(t) \sim \mathcal{N}(0, I)$ and $\odot$ is the element-wise dot operator. To ensure consistency, we ensure specific humidity has positive values.

Results We first present the online results from a 10-year, 6-hourly simulation run (~14k integration steps). As shown in Figure 1, the DMZ-hybrid run produces a stable 10-year simulation that successfully reproduces the annual cycle, comparable to the control run without subgrid variability correction. This stability is crucial, as it underpins the robustness of any data-driven parameterization.

Furthermore, we perform additional diagnostics focusing on precipitation, representing one of the more uncertain variables governed by complex multiscale processes [30]. As shown in Figure 2, the DMZ-hybrid online runs produce a physically consistent precipitation climatology over the 10-year simulation run. Finally, we evaluate whether the DMZ-based parameterization reduces precipitation bias by comparing against satellite-derived monthly precipitation from the Global Precipitation Climatology Project (GPCP) [31]. In Figure 3, we find substantial bias reductions over the Pacific Ocean, a key region where the El Niño–Southern Oscillation (ENSO), one of the dominant modes of variability in the Earth's circulation, emerges. Performance is poorer in tropical forest regions, likely due to the high precipitation variability that makes predictions challenging. Globally, online runs with DMZ-based parameterization achieve a lower climatological precipitation area-weighted RMSE of 1.55 mm/day compared to 1.69 mm/day in the SPEEDY control simulation.

4 Conclusion

In this work, we demonstrate the application of deep learning to approximate MZ projection operators for the development of a stochastic, memory-based parameterization scheme. Trained on high-

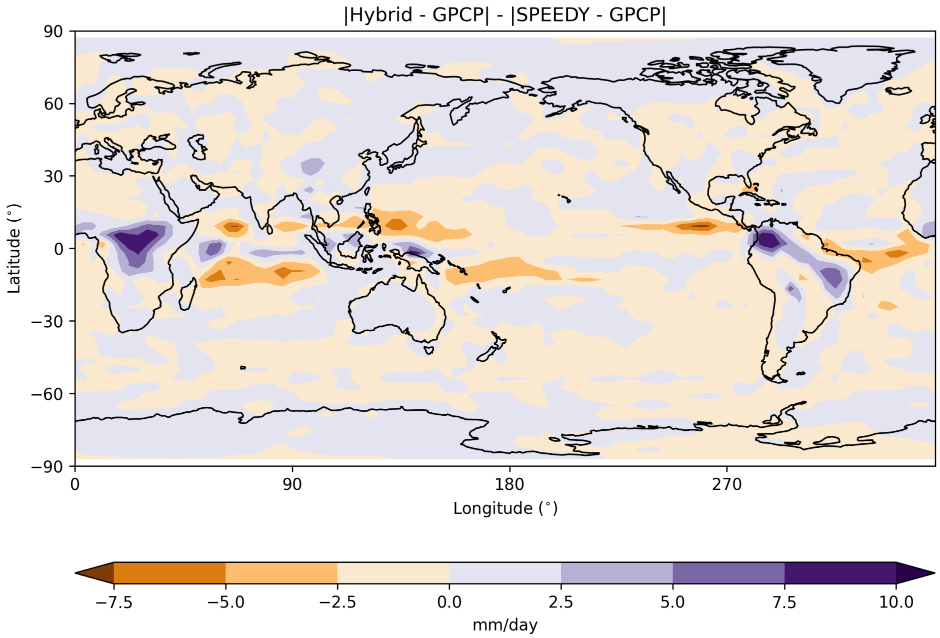


Figure 3: Comparing precipitation bias against GPCP observations (negative values indicate DMZ improvement over SPEEDY control run), we find significant reductions over the Pacific Ocean, an important region where ENSO, one of the main modes of variability, manifests.

resolution data and deployed in an online setting with interactive coupling to SPEEDY as the host GCM, DMZ produces stable and skillful multiyear online simulation runs.

At present, we have yet to perform extensive fine-tuning which results in only 8% improvement in precipitation RMSE relative to the SPEEDY control, below the $\sim$17% improvement reported for the memory-less approach of [11]. Nonetheless, these results demonstrate that data-driven memory terms can be stably integrated into an online coupled GCM. We view these findings as an important first step toward understanding how learned memory representations influence subgrid variability and long-term model behavior. Beyond their immediate performance, the results establish a transparent methodological framework for learning and evaluating MZ operators in coupled climate models, providing a solid foundation for systematic future improvements.

Future work on parameterization will explore the impact of increasing the memory length s and experimenting with alternative deep architectures on downstream task performance. Beyond parameterization, DMZ opens several promising directions, including applications in forecasting, where complex couplings between observed and unobserved variables can be naturally represented through the memory and orthogonal terms of the formalism [32]. Looking ahead, we envision DMZ as a general class of reduced-order models for emulating multiscale dynamics.

Limitations Fitting and maintaining a collection of regressors can be more challenging than training a single model as is typical in many ML regression tasks; essentially, the classical tradeoff is apparent here: what MZ offers in interpretability is partially lost in practicality. Addressing the challenge of making the data-driven estimation of MZ operators more tractable is an important focus of ongoing work.

Acknowledgment

The authors acknowledge funding, computing, and storage resources from the NSF Science and Technology Center (STC) Learning the Earth with Artificial Intelligence and Physics (LEAP) (Award #2019625). This work was supported in part by the U.S. Department of Energy, Office of Science (BER Regional and Global Model Analysis program, HiLAT-RASM project) and Scientific Computing Research (SciDAC Award 9233218CNA000001).

References

[1] Tapio Schneider, João Teixeira, Christopher S Bretherton, Florent Brient, Kyle G Pressel, Christoph Schär, and A Pier Siebesma. Climate goals and computing the future of clouds. *Nature Climate Change*, 7(1):3–5, 2017.

[2] Veronika Eyring, Pierre Gentine, Gustau Camps-Valls, David M Lawrence, and Markus Reichstein. AI-empowered next-generation multiscale climate modelling for mitigation and adaptation. *Nature Geoscience*, 17(10):963–971, 2024.

[3] Shinya Maeyama and Tomo-Hiko Watanabe. Extracting and modeling the effects of small-scale fluctuations on large-scale fluctuations by Mori–Zwanzig projection operator method. *Journal of the Physical Society of Japan*, 89(2):024401, 2020.

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

[5] Pierre Gentine, Mike Pritchard, Stephan Rasp, Gael Reinaudi, and Galen Yacalis. Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11):5742–5751, 2018.

[6] Tapio Schneider, L Ruby Leung, and Robert CJ Wills. Opinion: Optimizing climate models with process knowledge, resolution, and artificial intelligence. *Atmospheric Chemistry and Physics*, 24(12):7041–7062, 2024.

[7] Klaus Hasselmann. Stochastic climate models part I. Theory. *tellus*, 28(6):473–485, 1976.

[8] Cecil E Leith. Climate response and fluctuation dissipation. *Journal of Atmospheric Sciences*, 32(10):2022–2026, 1975.

[9] Georg Gottwald, Daan T Crommelin, Christian LE Franzke, et al. Stochastic climate theory. *Nonlinear and stochastic climate dynamics*, pages 209–240, 2017.

[10] Tim N Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 1(7):463–471, 2019.

[11] Daniel Giles, James Briant, Cyril J Morcrette, and Serge Guillas. Embedding machine-learnt sub-grid variability improves climate model precipitation patterns. *Communications Earth & Environment*, 5(1):712, 2024.

[12] Jerry Lin, Sungduk Yu, Liran Peng, Tom Beucler, Eliot Wong-Toi, Zeyuan Hu, Pierre Gentine, Margarita Geleta, and Mike Pritchard. Navigating the Noise: Bringing Clarity to ML Parameterization Design With $\mathcal{O}(100)$ Ensembles. *Journal of Advances in Modeling Earth Systems*, 17(4):e2024MS004551, 2025.

[13] Hazime Mori. Transport, collective motion, and Brownian motion. *Progress of theoretical physics*, 33(3):423–455, 1965.

[14] Robert Zwanzig. Nonlinear generalized Langevin equations. *Journal of Statistical Physics*, 9(3):215–220, 1973.

[15] Ricardo Buitrago Ruiz, Tanya Marwah, Albert Gu, and Andrej Risteski. On the benefits of memory for modeling time-dependent PDEs. *arXiv preprint arXiv:2409.02313*, 2024.

[16] Matthew Levine and Andrew Stuart. A framework for machine learning of model error in dynamical systems. *Communications of the American Mathematical Society*, 2(07):283–344, 2022.

[17] Abhinav Gupta and Pierre FJ Lermusiaux. Generalized neural closure models with interpretability. *Scientific Reports*, 13(1):10634, 2023.

[18] Burigede Liu, Eric Ocegueda, Margaret Trautner, Andrew M Stuart, and Kaushik Bhattacharya. Learning macroscopic internal variables and history dependence from microscopic models. *Journal of the Mechanics and Physics of Solids*, 178:105329, 2023.

[19] Judith Berner, Ulrich Achatz, Lauriane Batte, Lisa Bengtsson, Alvaro de la Cámara, Hannah M Christensen, Matteo Colangeli, Danielle RB Coleman, Daan Crommelin, Stamen I Dolaptchiev, et al. Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, 98(3):565–588, 2017.

[20] Christian LE Franzke, Terence J O'Kane, Judith Berner, Paul D Williams, and Valerio Lucarini. Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):63–78, 2015.

[21] Jeroen Wouters and Valerio Lucarini. Disentangling multi-level systems: averaging, correlations and memory. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(03):P03003, 2012.

[22] Jeroen Wouters and Valerio Lucarini. Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach. *Journal of Statistical Physics*, 151(5):850–860, 2013.

[23] Gabriele Vissio and Valerio Lucarini. A proof of concept for scale-adaptive parametrizations: the case of the Lorenz'96 model. *Quarterly Journal of the Royal Meteorological Society*, 144(710):63–75, 2018.

[24] Lisa Bengtsson, Jian-Wen Bao, Philip Pegion, Cecile Penland, Sara Michelson, and Jeffrey Whitaker. A model framework for stochastic representation of uncertainties associated with physical processes in NOAA's next generation global prediction system (NGGPS). *Monthly Weather Review*, 147(3):893–911, 2019.

[25] Yen Ting Lin, Yifeng Tian, Daniel Livescu, and Marian Anghel. Data-Driven Learning for the Mori–Zwanzig Formalism: A Generalization of the Koopman Learning Framework. *SIAM Journal on Applied Dynamical Systems*, 20(4):2558–2601, 2021.

[26] David Walters, Ian Boutle, Malcolm Brooks, Thomas Melvin, Rachel Stratton, Simon Vosper, Helen Wells, Keith Williams, Nigel Wood, Thomas Allen, et al. The Met Office unified model global atmosphere 6.0/6.1 and JULES global land 6.0/6.1 configurations. *Geoscientific Model Development*, 10(4):1487–1520, 2017.

[27] Mike Bush, Tom Allen, Caroline Bain, Ian Boutle, John Edwards, Anke Finnenkoetter, Charmaine Franklin, Kirsty Hanley, Humphrey Lean, Adrian Lock, et al. The first Met Office unified model–JULES regional atmosphere and land configuration, RAL1. *Geoscientific Model Development*, 13(4):1999–2029, 2020.

[28] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Swish: a self-gated activation function. *arXiv preprint arXiv:1710.05941*, 7(1):5, 2017.

[29] Franco Molteni. Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dynamics*, 20(2):175–191, 2003.

[30] Sara Shamekh, Kara D Lamb, Yu Huang, and Pierre Gentine. Implicit learning of convective organization explains precipitation stochasticity. *Proceedings of the National Academy of Sciences*, 120(20):e2216158120, 2023.

[31] Robert Adler, Jian-Jian Wang, Mathew Sapiano, George Huffman, David Bolvin, Eric Nelkin, and CDR NOAA. Global precipitation climatology project (GPCP) climate data record (CDR), version 1.3 (Daily). 2020.

[32] Fabrizio Falasca. Neural models of multiscale systems: conceptual limitations, stochastic parametrizations, and a climate application. *arXiv preprint arXiv:2506.22552*, 2025.