

Auditing how an AI weather model uses lower-stratospheric and moisture information: structure-destroying interventions in GraphCast

Yi Zou

yizou505@gmail.com

Independent Researcher <https://orcid.org/0009-0000-6493-7255>

Research Article

Keywords: GraphCast, machine-learning weather prediction, stratosphere–troposphere coupling, moisture, information-use audit, intervention experiments

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11530492/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Auditing how an AI weather model uses lower-stratospheric and moisture information: structure-destroying interventions in GraphCast

Yi Zou^{1*}

^{1*}Independent Researcher, China.

Corresponding author(s). E-mail(s): yizou505@gmail.com;

Abstract

We audit how GraphCast (1°, 13 levels, top at 50 hPa) uses lower-stratospheric and moisture information via input interventions in 72 ten-day ERA5 forecasts (2016–2021, 18 per season). Replacing the atmosphere above 150 hPa with its zonal mean changes Northern Hemisphere Z500 by 55 m at 120 h, 1.49 [1.40, 1.59] times the response to incoherent random-phase noise with identical per-latitude spectra. Zonal-mean and climatological replacements are indistinguishable over all dates. Per unit normalized input, mid-tropospheric perturbations act 4.6 times more strongly than the stratospheric knockout at 24 h (6.6–6.9 under alternative normalization) and 1.8 times at 120 h; the stratospheric share is largest in boreal winter. Global humidity scaling changes 30–70°N Z500 eddy variance monotonically, down to 0.66 of control for complete drying, unlike a conservative norm-matched temperature-noise control. Moisture deficits recover with an e-folding time of about 126 h; surpluses e-fold within 30 h, but one-third persists.

Keywords: GraphCast, machine-learning weather prediction, stratosphere–troposphere coupling, moisture, information-use audit, intervention experiments

1 Introduction

Machine-learning weather prediction (MLWP) models such as GraphCast [Lam et al. \(2023\)](#) rival physics-based systems on medium-range scores [Rasp et al. \(2024\)](#). Using them as virtual laboratories requires knowing what information they actually use.

Counterfactual input experiments on frozen MLWP models include idealized dynamical tests [Hakim and Masanam \(2024\)](#); [Elms \(2026\)](#), gradient sensitivities [Baño-Medina et al. \(2025\)](#), rotated-Earth tests [Höyer et al. \(2026\)](#), out-of-distribution initial climates [Rackow et al. \(2024\)](#); [Amiramjadi et al. \(2026\)](#), and optimized or localized perturbations [Hakim and Agrawal \(2026\)](#); [Huang et al. \(2026\)](#). We contribute intervention designs and controls for two information channels: the lower stratosphere/tropopause region and moisture.

The stratosphere influences tropospheric jets, storm tracks, and annular modes [Baldwin and Dunkerton \(2001\)](#); [Kidston et al. \(2015\)](#). On subseasonal timescales, SNAPSI [Hitchcock et al. \(2022\)](#) quantifies this influence by nudging the stratosphere of forecast systems toward observed or climatological states. Relaxation toward the truth has been applied to the stratosphere of a neural weather model [Perkan et al. \(2025\)](#), and SNAPSI-style nudging has been applied to ACE2 for the 2018 and 2019 sudden stratospheric warmings [Jones \(2026\)](#). [Wu \(2026\)](#) found that GraphCast does not capture subseasonal stratosphere–troposphere coupling. [Subich \(2025\)](#) computed level-wise input sensitivities of a 37-level GraphCast over 1–5 days. We ask how strongly GraphCast forecasts depend on lower-stratospheric information, whether that dependence rests on eddy structure or only on amplitude, and how it compares with the troposphere at equal input norm.

Latent heating intensifies extratropical cyclones. Zeroing the moisture of tropical-cyclone seeds suppresses their growth in MLWP models [Hakim and Masanam \(2024\)](#); [Elms \(2026\)](#), and globally uniform perturbations can expose non-local shortcuts in ML emulators [Mahesh et al. \(2026\)](#). We extend this to global, multi-dose humidity scaling, measuring extratropical eddy variance and the model’s implied moisture-recovery timescale.

Norm-matched random perturbations are an established reference [Hakim and Agrawal \(2026\)](#); what is new is the combination: structure-destroying stratospheric knockouts paired with a same-spectrum random-phase control and a norm-matched tropospheric perturbation, and a global humidity dose–response paired with a norm-matched temperature-noise control, evaluated over 72 initial dates.

2 Data and Methods

2.1 Model and initial conditions

We use the public GraphCast-small checkpoint [Lam et al. \(2023\)](#). It has 1° resolution, 13 pressure levels from 50 to 1000 hPa, and the 2to5 mesh, uses precipitation as input and output, and was trained on ERA5 1979–2015. We ran it with `graphcast` v0.2 in bfloat16 for 40 autoregressive 6-h steps (240 h). Initial conditions at $t - 6$ h and t , and the verifying states, are ERA5 [Hersbach et al. \(2020\)](#). We took them from the WeatherBench 2 [Rasp et al. \(2024\)](#) 1.5° conservatively regridded archive and interpolated them bilinearly to 1°. Control forecasts have global Z500 RMSE of 4.7, 14.3, 30.3, and 79.3 m at 24, 72, 120, and 240 h. We use 72 initial dates at 00 UTC on the 1st or 15th of every month of 2016–2021, giving 18 per season. Each date has one control and 20 perturbed forecasts. Given the 50-hPa model top and 10-day horizon,

the study concerns how the model uses lower-stratosphere/tropopause information. It does not address the downward influence of sudden stratospheric warmings.

2.2 Experiment A: lower-stratosphere knockouts

Each knockout modifies all six 3-D variables (T, z, u, v, ω, q) at both input times. The modified levels are $\{50\}$, $\{50, 100\}$, or $\{50, 100, 150\}$ hPa, which we call “above” 50, 100, or 150 hPa. The knockouts are as follows.

- **zm**: replace each field by its zonal mean, which removes the eddies.
- **sh**: shift each field by 90° in longitude.
- **clim_150**: replace 50–150 hPa with the 2010–2014 mean for the same calendar date.
- **oy_150**: replace 50–150 hPa with the same date in an adjacent year.

Two controls accompany them. The **rand_150** control starts from the **zm_150** increment δ , which equals minus the eddy field. For each variable, level, latitude circle, and input time, we randomized the phases of all zonal wavenumbers $k \geq 1$ independently. The control therefore has the same norm as δ on every latitude circle, variable, and level (realized ratio 1.000000). It differs from δ only in the loss of phase structure, inter-variable balance, and vertical and temporal coherence. The **tropo_matched** control is the eddy component at 400–600 hPa with randomized phases. It is scaled per date to the same normalized RMS as δ , using GraphCast’s `stddev.by_level` without area weighting. Finally, **zm_150** at amplitudes 0.25 and 0.5 tests linearity. Full definitions are given in Online Resource 1 (Text S1).

2.3 Experiment B: global humidity scaling

We multiplied specific humidity at all levels and both input times by $a \in \{0, 0.25, 0.5, 0.75, 1.25, 1.5\}$ and scaled the 6-h precipitation input by the same factor. Doses with $a > 1$ are out of distribution: about 40% of tropospheric mass is supersaturated at $a = 1.5$ (Online Resource 1, Table S2). As a bad-input control, we added a random-phase eddy temperature perturbation at all levels. It is scaled to the normalized norm of the humidity change for $a = 0$ and $a = 0.5$ ($T\text{-noise}_0$, $T\text{-noise}_{0.5}$). Because σ_q is small at 50–100 hPa, the matched temperature noise is 8.5 ($a = 0$) and 4.3 ($a = 0.5$) times the natural eddy-temperature amplitude. Comparisons against it are therefore conservative tests.

2.4 Diagnostics and statistics

Experiment A. The primary metric is the area-weighted RMS difference in Z500 from the control over 20–90°N (ΔZ_{500}). Secondary diagnostics are zonal-mean fractions, 30–70°N wavenumber spectra, vertical profiles, and 6-h imbalance indicators.

Experiment B. The primary metric is the per-date ratio of Z500 eddy variance (variance of the deviation from the zonal mean) to the control, over 30–70°N or 30–70°S. We also track recovery of total column water (TCW) through the fraction $(TCW_{\text{ctl}} - TCW_a)/[(1 - a)TCW_{\text{ctl}}(6\text{h})]$. Its e-folding time is the first 6-h lead at which the mean fraction falls below e^{-1} .

Statistics. Confidence intervals (95%) come from a paired bootstrap over dates with 2000 resamples, and ratios are ratios of date means. Seven pre-registered contrasts are Holm-corrected: five ratios at 120 h and two eddy-variance differences at 240 h. Seasonal differences are tested with permutation tests (10 000 permutations of season labels) and Holm-corrected across nine seasonal tests (Online Resource 1, Table S3).

3 Results

3.1 GraphCast uses coherent lower-stratospheric eddy structure

Replacing the atmosphere above 150 hPa by its zonal mean changes NH Z500 by 16.4 m [14.9, 17.9] at 24 h, growing to 35.9, 55.2, and 101.5 m at 72, 120, and 240 h (Fig. 1a; Online Resource 1, Table S1). At 120 h the response grows with knockout depth: 16.8, 31.2, and 55.2 m above 50, 100, and 150 hPa. The response spreads downward from the perturbed levels (Fig. 1b; at 6 h, 101.5 m at 50 hPa and 6.0 m at 500 hPa). By 120 h the 250-hPa response (79.9 m) exceeds that at 50 hPa (75.2 m). The response is approximately linear in amplitude through 120 h. Relative to the full-amplitude response, the 0.5 and 0.25 knockouts give 1.06 and 0.98 per unit amplitude at 120 h, and 0.87 and 0.77 at 24 h.

At 24 h the random-phase control produces nearly the same response as the knockout (zm/rand 1.04 [1.00, 1.07]). From 72 h onward the knockout acts substantially more strongly: 1.57 [1.50, 1.66] at 72 h and 1.49 [1.40, 1.59] at 120 h (Holm $p < 0.001$; Table 1). Because the two perturbations have identical norms on every latitude circle, this ratio measures the effect of coherent, balanced eddy structure relative to incoherent noise with the same spectrum. It is not an initial shock: at 6 h, the random-phase control perturbs 500-hPa vertical velocity more (0.075 vs 0.049 Pa s⁻¹). A coherent but misplaced eddy field acts more strongly still (sh/zm 1.54 [1.49, 1.58] at 24 h and 1.29 [1.25, 1.33] at 120 h).

The Z500 response is overwhelmingly zonally asymmetric. The zonal-mean fraction of the NH ΔZ_{500} variance is 0.094, 0.072, 0.070, and 0.104 at 24, 72, 120, and 240 h (Fig. 1c). From 72 h onward, about half of the 30–70°N power lies in wavenumbers 1–3 (0.50 at 72 h and 0.49 at 120 h). The model thus carries tropopause-region information into synoptic eddies. The zonal-mean, annular-mode pathway [Baldwin and Dunkerton \(2001\)](#); [Kidston et al. \(2015\)](#) is not expected to act from these levels within 10 days.

3.2 Zonal-mean, climatological, and other-year stratospheres

Averaged over all dates, the zonal-mean and 2010–2014 climatological replacements are indistinguishable. The ratio zm/clim is 0.99 [0.95, 1.03] at 72 h and 0.97 [0.93, 1.02] at 120 h (Holm $p = 0.22$), and the two give nearly identical 6-h imbalance (500-hPa vertical velocity 0.049 vs 0.050 Pa s⁻¹). A real stratosphere from an adjacent year is more harmful than the climatology, with oy/clim 1.30 [1.23, 1.37] at 24 h and 1.14 [1.09, 1.20] at 120 h (Holm $p < 0.001$). The two converge by 240 h (1.03 [0.99, 1.08]). As with the shift, a realistic but wrong eddy field degrades the forecast more than a smooth one does.

3.3 Stratospheric versus tropospheric sensitivity

At equal normalized input norm, a random-phase eddy perturbation at 400–600 hPa changes NH Z500 more than zm.150 does. The ratio is 4.58 [4.42, 4.77] at 24 h, 2.61 at 72 h, and 1.84 [1.76, 1.92] at 120 h (Holm $p < 0.001$). Under residual-based normalization it would be 6.6–6.9 at 24 h and 2.7–2.8 at 120 h (Online Resource 1, Table S5). This complements the level-wise profiles of Subich (2025) with structure-defined perturbations and a 10-day horizon. By 240 h both responses approach error saturation (110.7 and 101.5 m, compared with a control RMSE of 79.3 m), and the ratio falls to 1.09 [1.05, 1.13]. We therefore focus on 24–120 h.

The relative influence of the stratosphere is largest in boreal winter (Online Resource 1, Fig. S1). The tropo/zm ratio at 120 h is 1.71 [1.58, 1.84] in DJF and 2.21 [1.99, 2.47] in JJA (Holm-corrected permutation $p = 0.005$). The absolute zm.150 response at 120 h is 77.9 m [72.6, 83.5] in DJF and 32.4 m [28.6, 36.6] in JJA. It remains largest in DJF per unit input norm (104 vs 57 m; permutation $p = 0.0005$; Online Resource 1, Table S4).

3.4 Global humidity dose–response

Humidity scaling changes NH Z500 eddy variance monotonically for a from 0 to 1.25 (Fig. 3; Table 2). At 120 h the ratio is 0.661 [0.636, 0.685] for $a = 0$, 0.761 for $a = 0.25$, 0.833 [0.811, 0.854] for $a = 0.5$, 0.902 for $a = 0.75$, and 1.093 [1.075, 1.112] for $a = 1.25$. Drying weakens and moistening strengthens extratropical eddies, as expected from latent-heat feedback. The SH responds similarly to drying (0.607 [0.578, 0.637] at $a = 0$ and 120 h).

The temperature-noise control behaves differently (Fig. 2a). It first increases eddy variance, so the NH difference ($a = 0$ minus $T\text{-noise}_0$) is -0.360 at 24 h. It then collapses the eddy variance to 0.215 at 120 h, far below the drying experiment. At 240 h, drying retains more NH eddy variance than the matched noise, by 0.391 [0.339, 0.442] for $a = 0$ and 0.350 [0.308, 0.392] for $a = 0.5$ (both Holm $p < 0.001$). In the SH the $a = 0$ margin is 0.125 [0.083, 0.171]. The noise also degrades skill far more: global Z500 RMSE at 120 h is 112.0 m, compared with 82.5 m for $a = 0$ and 30.3 m for the control. It even drives global-mean precipitation negative (-0.90 mm per 6 h at 24 h). Because this control is oversized, these margins are conservative; unlike generic corruption, the humidity response is graded and sign-reversing.

The NH drying effect varies with season (Holm-corrected permutation $p = 0.002$). At 120 h the $a = 0$ ratio is 0.741 in DJF and 0.602 in SON. The excess of the drying response over the temperature-noise control at 240 h is positive in all seasons, largest in DJF (0.56) and smallest in JJA (0.14 [0.08, 0.20]; Holm-corrected $p < 0.001$).

3.5 Asymmetric moisture recovery

After drying, the model restores TCW over several days (Fig. 2c). For $a = 0$, the fraction of the initial deficit remaining is 0.802, 0.529, 0.379, and 0.218 at 24, 72, 120, and 240 h. The e-folding time is 126 h for $a = 0$ and $a = 0.25$, 108 h for $a = 0.5$, and 90 h for $a = 0.75$. Added moisture is removed much faster, with e-folding times of 42 h for $a = 1.25$ and 30 h for $a = 1.5$. Removal happens mainly through a precipitation

burst: for $a = 1.5$, global-mean precipitation at 6 h is 2.46 mm per 6 h, against 0.74 in the control. After that, about one-third of the excess persists (0.327 at 240 h). The drying timescale lies between the median (4–5 days) and mean (8–10 days) water vapor lifetimes estimated by [Sodemann \(2020\)](#), although the two quantities are defined differently.

4 Discussion and Conclusions

Structure-destroying interventions with matched controls show that GraphCast uses the coherent, balanced eddy structure of the lowermost stratosphere and tropopause region. Removing that structure changes NH Z500 by 55 m at 120 h, about 1.5 times the effect of incoherent noise with identical per-latitude norms. Per unit normalized input, the mid-troposphere remains 4.6 (24 h) to 1.8 (120 h) times more influential, and the relative stratospheric influence is largest in boreal winter. Zonal-mean and climatological lower stratospheres are interchangeable when averaged over all dates, whereas realistic but wrong ones are more harmful. Complementing SNAPSI-type nudging in physics-based [Hitchcock et al. \(2022\)](#) and ML models [Jones \(2026\)](#); [Perkan et al. \(2025\)](#), which measures how much predictability the stratosphere carries, our approach diagnoses how the model uses that information.

In the moisture channel, global humidity scaling yields a graded, sign-reversing eddy response that is clearly distinct from a conservative norm-matched temperature-noise control. Moisture recovery is asymmetric: deficits e-fold in about 126 h and surpluses in about 30 h. Like other uniform global perturbations [Mahesh et al. \(2026\)](#), these designs form a reusable protocol for auditing which inputs MLWP models rely on. Applying it to models with higher tops and to other architectures is a natural next step.

Acknowledgements. The author thanks ECMWF for ERA5, Google DeepMind for GraphCast, and the WeatherBench2 team for the data infrastructure.

Declarations

- **Funding** No funding was received for conducting this study.
- **Competing interests** The author has no relevant financial or non-financial interests to disclose.
- **Ethics approval** Not applicable.
- **Consent to participate** Not applicable.
- **Consent for publication** Not applicable.
- **Data availability** ERA5 initial conditions were read from the public WeatherBench 2 archive [Rasp et al. \(2024\)](#) ([gs://weatherbench2/datasets/era5/1959-2023_01_10-6h-240x121_equiangular_with_poles_conservative.zarr](https://weatherbench2/datasets/era5/1959-2023_01_10-6h-240x121_equiangular_with_poles_conservative.zarr)). The GraphCast-small weights and normalization statistics were read from the public `dm_graphcast` Google Cloud bucket. Per-date numerical results (JSON) supporting the figures and tables are held in a private repository at <https://github.com/yizou728282/graphcast-strat-moisture> and are available from the author upon reasonable request until acceptance; a public archive with a DOI will be added upon publication.

- **Code availability** The `graphcast` code is available at <https://github.com/google-deepmind/graphcast> (tag v0.2). Experiment and analysis scripts are held in the same private repository and are available from the author upon reasonable request until acceptance; a public archive with a DOI will be added upon publication.
- **Author contribution** Yi Zou designed the study, performed the experiments and analysis, and wrote the manuscript.

References

- Amiramjadi M, Roth C, Nowack P (2026) Architectural sensitivity of AI weather prediction models to 3D structural and seasonal climate forcing. EGU General Assembly 2026, Vienna, Austria, 3–8 May 2026, EGU26-18038, <https://doi.org/10.5194/egusphere-egu26-18038>, <https://meetingorganizer.copernicus.org/EGU26/EGU26-18038.html>
- Baldwin MP, Dunkerton TJ (2001) Stratospheric harbingers of anomalous weather regimes. *Science* 294(5542):581–584. <https://doi.org/10.1126/science.1063315>
- Baño-Medina J, Sengupta A, Doyle JD, et al (2025) Are AI weather models learning atmospheric physics? A sensitivity analysis of cyclone Xynthia. *npj Climate and Atmospheric Science* 8:92. <https://doi.org/10.1038/s41612-025-00949-6>
- Elms J (2026) Results from idealized testing of machine learning atmosphere models. Master’s thesis, Indiana University, Department of Earth and Atmospheric Sciences, Bloomington, IN, m.S. thesis. <https://hdl.handle.net/2022/38490>
- Hakim GJ, Agrawal A (2026) Gray swan factory: Making extreme events from ordinary cyclones. arXiv preprint arXiv:2604.00348, <https://doi.org/10.48550/arXiv.2604.00348>
- Hakim GJ, Masanam S (2024) Dynamical tests of a deep learning weather prediction model. *Artificial Intelligence for the Earth Systems* 3(3). <https://doi.org/10.1175/AIES-D-23-0090.1>
- Hersbach H, Bell B, Berrisford P, et al (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146(730):1999–2049. <https://doi.org/10.1002/qj.3803>
- Hitchcock P, Butler A, Charlton-Perez A, et al (2022) Stratospheric nudging and predictable surface impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts. *Geoscientific Model Development* 15(13):5073–5092. <https://doi.org/10.5194/gmd-15-5073-2022>
- Höver M, Klöwer M, Schroeder de Witt C, et al (2026) Spatial generalization tests for machine learning-based weather models to assess physical consistency. arXiv preprint arXiv:2607.20716, <https://doi.org/10.48550/arXiv.2607.20716>

- Huang Q, Liu M, Kwon Y, et al (2026) Steering tropical cyclones using small perturbations in an AI weather model. arXiv preprint arXiv:2605.29248, <https://doi.org/10.48550/arXiv.2605.29248>
- Jones M (2026) Do machine learning-based forecasts capture the link between stratospheric variability and extreme weather? Royal Meteorological Society Early Career and Student Conference, University of Exeter, 6–8 July 2026 (oral presentation), <https://www.rmets.org/conference/early-career-and-student-conference-2026/papers/do-machine-learning-based-forecasts>
- Kidston J, Scaife AA, Hardiman SC, et al (2015) Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience* 8(6):433–440. <https://doi.org/10.1038/ngeo2424>
- Lam R, Sanchez-Gonzalez A, Willson M, et al (2023) Learning skillful medium-range global weather forecasting. *Science* 382(6677):1416–1421. <https://doi.org/10.1126/science.adf2336>
- Mahesh A, Collins WD, O’Brien TA, et al (2026) Examining fast radiatively driven responses using machine-learning weather emulators. arXiv preprint arXiv:2602.16090, <https://doi.org/10.48550/arXiv.2602.16090>
- Perkan U, Zaplotnik Z, Skok G (2025) Forecast error diagnostics in neural weather models. arXiv preprint arXiv:2506.11987, <https://doi.org/10.48550/arXiv.2506.11987>
- Rackow T, Koldunov N, Lessig C, et al (2024) Robustness of AI-based weather forecasts in a changing climate. arXiv preprint arXiv:2409.18529, <https://doi.org/10.48550/arXiv.2409.18529>
- Rasp S, Hoyer S, Merose A, et al (2024) WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems* 16(6):e2023MS004019. <https://doi.org/10.1029/2023MS004019>
- Sodemann H (2020) Beyond turnover time: Constraining the lifetime distribution of water vapor from simple and complex approaches. *Journal of the Atmospheric Sciences* 77(2):413–433. <https://doi.org/10.1175/JAS-D-18-0336.1>
- Subich C (2025) Efficient fine-tuning of 37-level GraphCast with the Canadian global deterministic analysis. *Artificial Intelligence for the Earth Systems* 4(3). <https://doi.org/10.1175/AIES-D-24-0101.1>, preprint: arXiv:2408.14587
- Wu Z (2026) Assessing subseasonal predictions of stratosphere-troposphere coupling of GraphCast. *Journal of Geophysical Research: Atmospheres* 131(6):e2025JD044852. <https://doi.org/10.1029/2025JD044852>

Table 1 Pre-registered contrasts: ratios of date-mean NH RMS ΔZ_{500} with paired-bootstrap 95% CIs. p is evaluated at 120 h and Holm-corrected over seven contrasts

Contrast	24 h	72 h	120 h	240 h	p_{Holm}
zm_150 / rand_150	1.04 [1.00, 1.07]	1.57 [1.50, 1.66]	1.49 [1.40, 1.59]	1.16 [1.11, 1.22]	<0.001
tropo / zm_150	4.58 [4.42, 4.77]	2.61 [2.49, 2.74]	1.84 [1.76, 1.92]	1.09 [1.05, 1.13]	<0.001
zm_150 / clim_150	1.07 [1.01, 1.12]	0.99 [0.95, 1.03]	0.97 [0.93, 1.02]	1.02 [0.99, 1.06]	0.22
oy_150 / clim_150	1.30 [1.23, 1.37]	1.20 [1.14, 1.26]	1.14 [1.09, 1.20]	1.03 [0.99, 1.08]	<0.001
sh_150 / zm_150	1.54 [1.49, 1.58]	1.41 [1.38, 1.45]	1.29 [1.25, 1.33]	1.07 [1.03, 1.11]	<0.001

Table 2 Z500 eddy-variance ratio (perturbed/control) over 30–70° latitude, as the mean with a 95% bootstrap CI

Experiment	NH		SH	
	120 h	240 h	120 h	240 h
$a = 0$	0.661 [0.636, 0.685]	0.635 [0.601, 0.669]	0.607 [0.578, 0.637]	0.530 [0.491, 0.573]
$a = 0.5$	0.833 [0.811, 0.854]	0.848 [0.813, 0.884]	0.816 [0.790, 0.842]	0.747 [0.700, 0.796]
$a = 1.25$	1.093 [1.075, 1.112]	1.078 [1.034, 1.122]	1.083 [1.055, 1.115]	1.041 [0.987, 1.095]
$a = 1.5$	1.067 [1.039, 1.095]	1.096 [1.033, 1.169]	1.194 [1.139, 1.257]	1.073 [0.997, 1.146]
T-noise ₀	0.215 [0.194, 0.238]	0.245 [0.211, 0.280]	0.229 [0.210, 0.249]	0.405 [0.371, 0.440]
T-noise _{0.5}	0.342 [0.319, 0.365]	0.498 [0.460, 0.536]	0.579 [0.544, 0.613]	0.974 [0.908, 1.041]

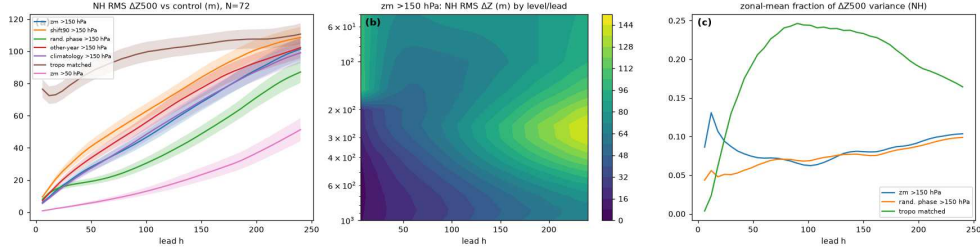


Fig. 1 Stratospheric knockouts ($N = 72$). (a) NH RMS ΔZ_{500} relative to the control, with 95% bootstrap bands. (b) NH RMS ΔZ (m) for zm_150 by pressure level and lead time. (c) Zonal-mean fraction of the NH ΔZ_{500} variance

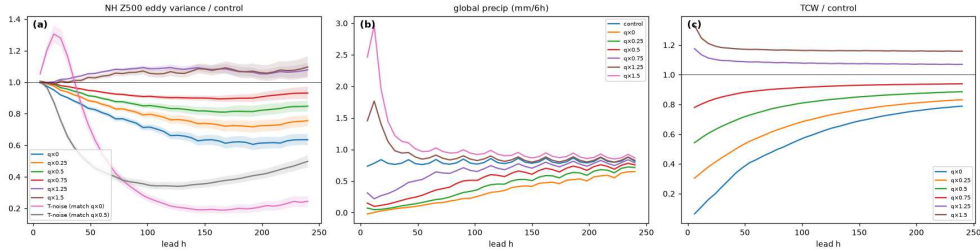


Fig. 2 Humidity scaling ($N = 72$). (a) NH Z500 eddy variance relative to the control, including the temperature-noise controls. (b) Global-mean precipitation (mm per 6 h). (c) Global-mean TCW relative to the control

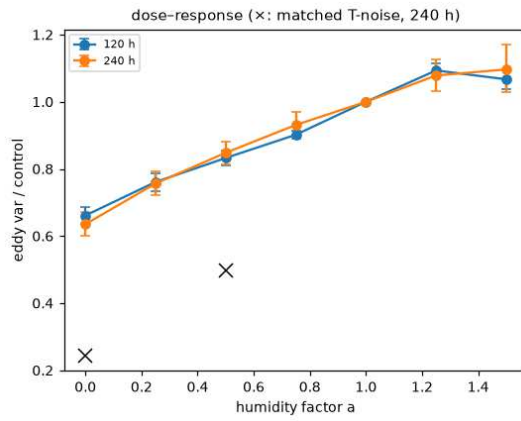


Fig. 3 NH Z500 eddy-variance ratio as a function of the humidity factor a at 120 and 240 h (95% CIs). Crosses mark the temperature-noise controls at 240 h