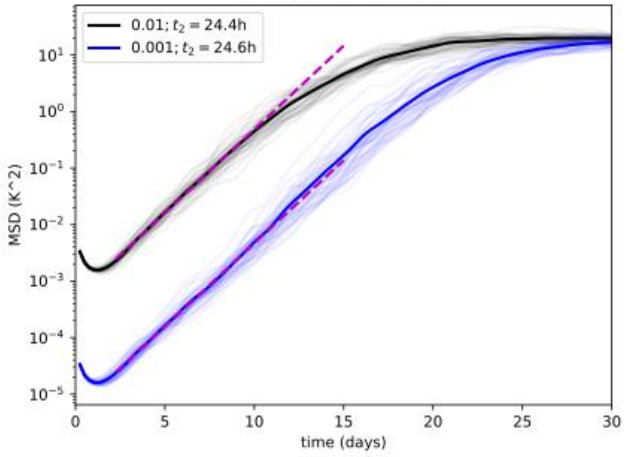
Exploring the Limit of Atmospheric Predictability with Machine Learning Models



Greg Hakim & Trent Vonich
University of Washington

Chris Snyder
National Center for Atmospheric Research

21 July 2025
US CLIVAR Summit

Outline

- Theory and prior results: prediction time limit is set by small scales
  - ~14 days, known and confirmed repeatedly since Lorenz (1969)

"the predictability speed of light"

- Testing the limit with new ML models & gradient sensitivity
- Pacific Northwest heatwave of June 2021 (Vonich & Hakim, 2024)
- Large sample confirmation (Vonich & Hakim, 2025: arxiv.org/abs/2504.20238)

Atmospheric Predictability Theory I (Lorenz 1969)

- defined by energy spectrum power laws (–3 ~ synoptic scale, –5/3 ~ mesoscale)
- isotropic; homogeneous; spectrally local triad interactions

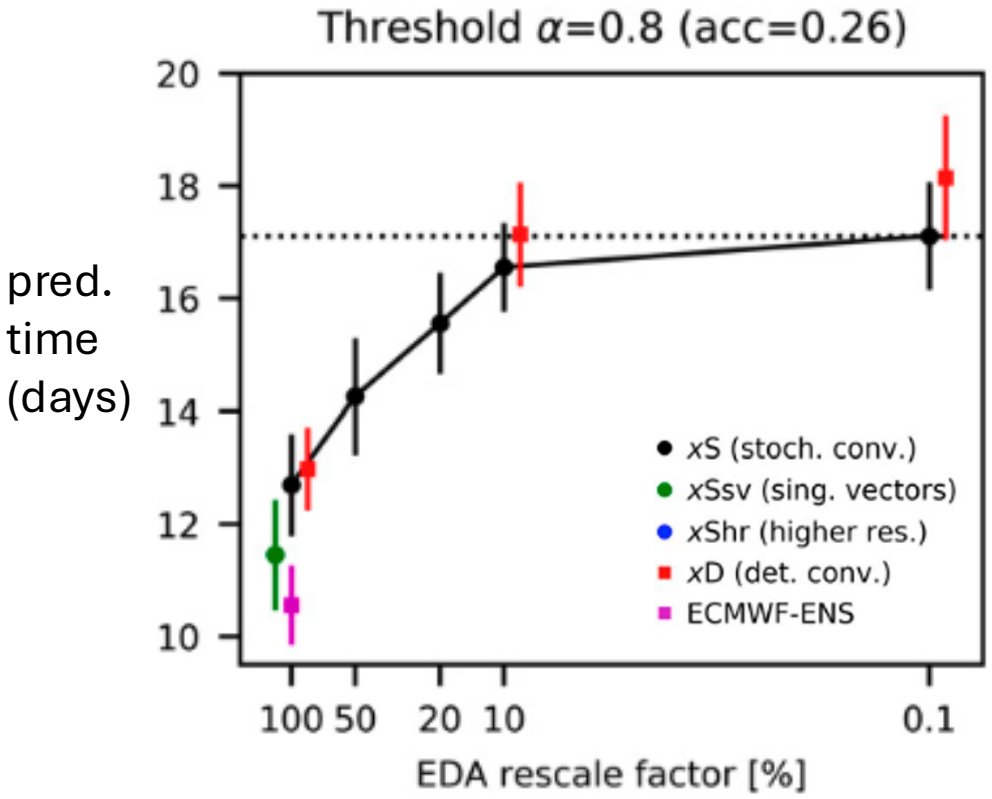
Timescale, T, for errors at small scale wavenumber, k_s, to saturate large scale, k_l:

$$T_{-3} \sim ln\left(\frac{k_s}{k_l}\right) \qquad \lim_{k_s \to \infty} T_{-3} \to \infty$$

$$T_{-5/3} \sim k_l^{-2/3} - k_s^{-2/3} \qquad \lim_{k_s \to \infty} T_{-5/3} \to k_l^{-2/3}$$

finite upper bound
"speed of light"

Atmospheric Predictability Limit Estimates



Selz et al. (2022) perfect model experiments

- reduce analysis error up to 90%
- **limit ~17 days**; 5 days longer than current IFS

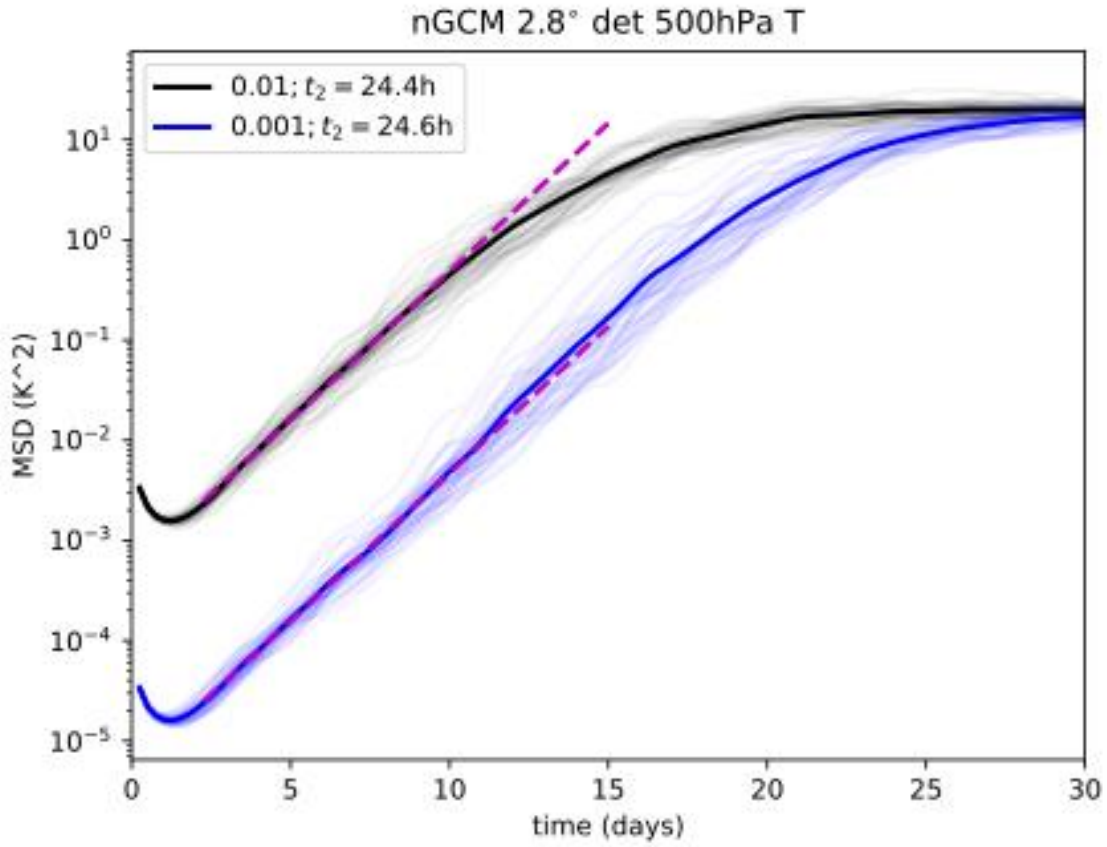
"These results confirm that **planetary-scale predictability is intrinsically limited by rapid error growth due to latent heat release in clouds <u>through an upscale-interaction process</u>…**''

"…the intrinsic limit could be reached in about 40–50 years if this [recent] rate [of improvement] were to continue"

Similar findings: Zhang et al. (2019); Judt (2020)

Atmospheric Predictability Theory II (Lorenz 1963)

- tangent linear, t → ∞, stability
- errors grow exponentially at the leading Lyapunov exponent
- “chaos”
- very clear in ML models
  - perfect-twin experiments
  - no initial fast growth



New Opportunities from ML Models

- ML models provide a new approach to predictability
  - no -5/3 spectrum; strongly damped at small scales (Bonavita 2024)
  - different error growth at short leads (Selz & Craig 2023)

- ML forecast skill beyond the limit of physics models suggests that predictability of the **true system is not limited by small scales.**

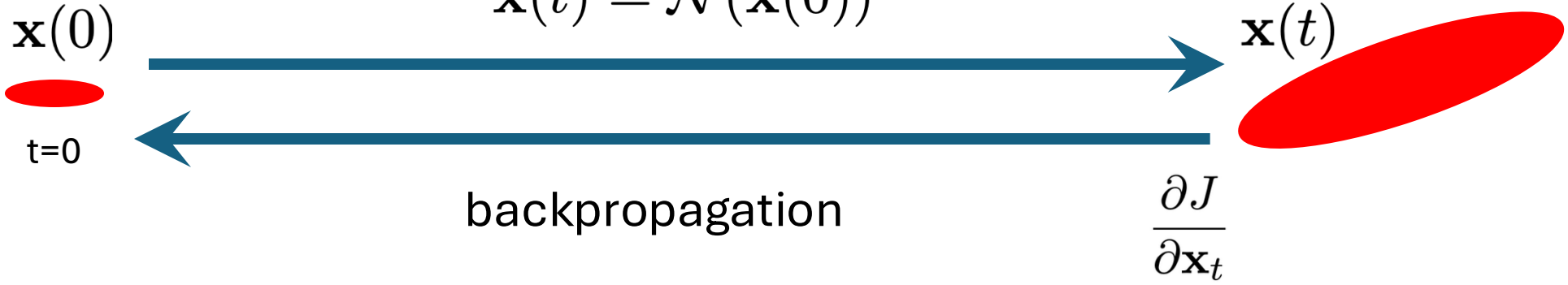
- Gradient facilities to compute derivatives of all aspects (model & state)

- Very computationally efficient (large ensembles)

Deep Learning Sensitivity Analysis

$$J(\mathbf{x}(t)) \sim \sum (\mathbf{x}_i - \tilde{\mathbf{x}}_i)^{\mathrm{T}} (\mathbf{x}_i - \tilde{\mathbf{x}}_i)$$

$$\mathbf{x}(t) = \mathcal{N}(\mathbf{x}(0))$$



$$\mathbf{x}_{i+1} = \mathbf{x}_i - \eta \frac{\partial J(\mathbf{x}_i(t))}{\partial \mathbf{x}(0)}$$

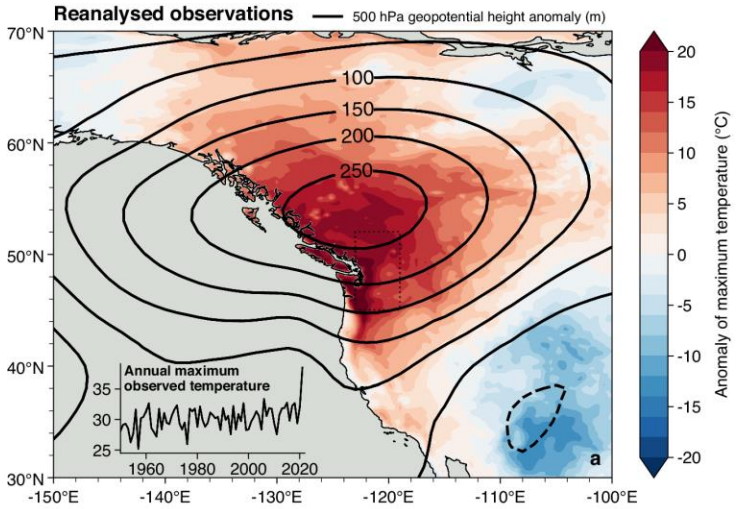
Map large errors (future) to when they are small (analysis error)

Application: Pacific Northwest Heatwave (2021)

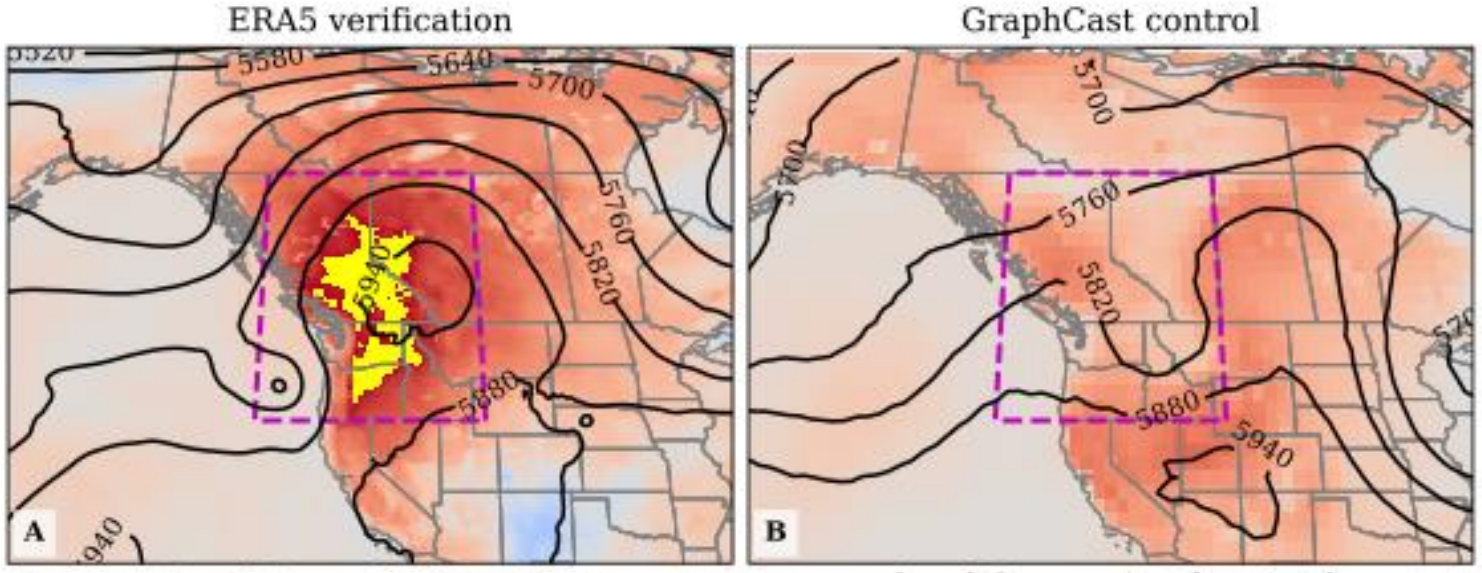
- Top-6 global extremes since 1960 (Thompson et al. 2022)
- Highest Canadian temperature (49.6ºC, Lytton, British Columbia)
- >1400 deaths
- Not in ML training data

Optimal initial condition evaluation

- GraphCast 1º model (Lam et al. 2023)
- Loss: ~weighted mean-squared error
- Derivatives: JAX framework (Bradbury et al. 2018)
- Initial conditions and verification: ERA5



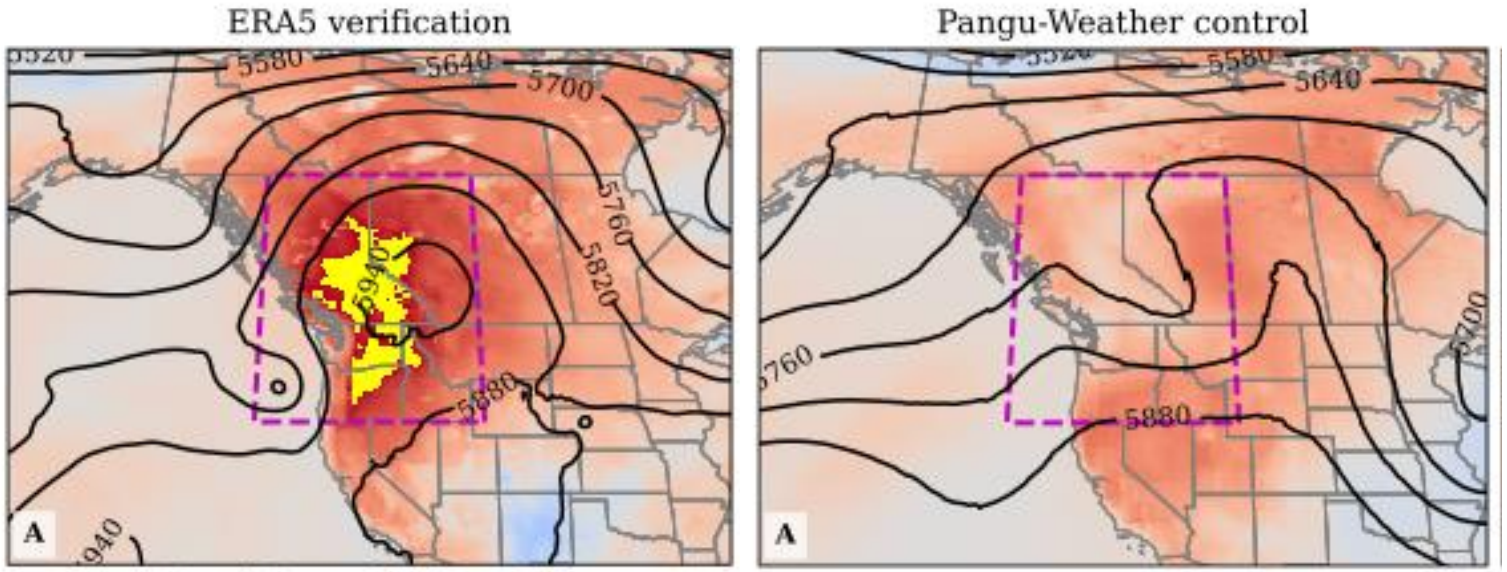
Leach et al. (2024)



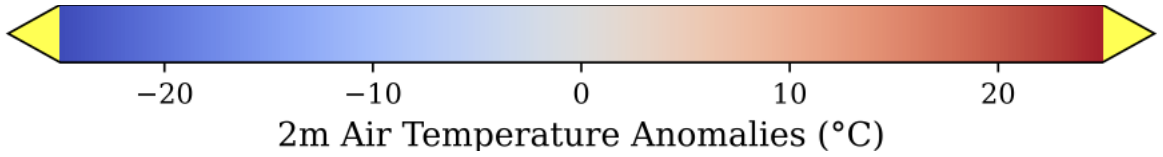
10-day forecast from ERA5

~90% error reduction

Vonich & Hakim (2024)



10-day forecast from ERA5

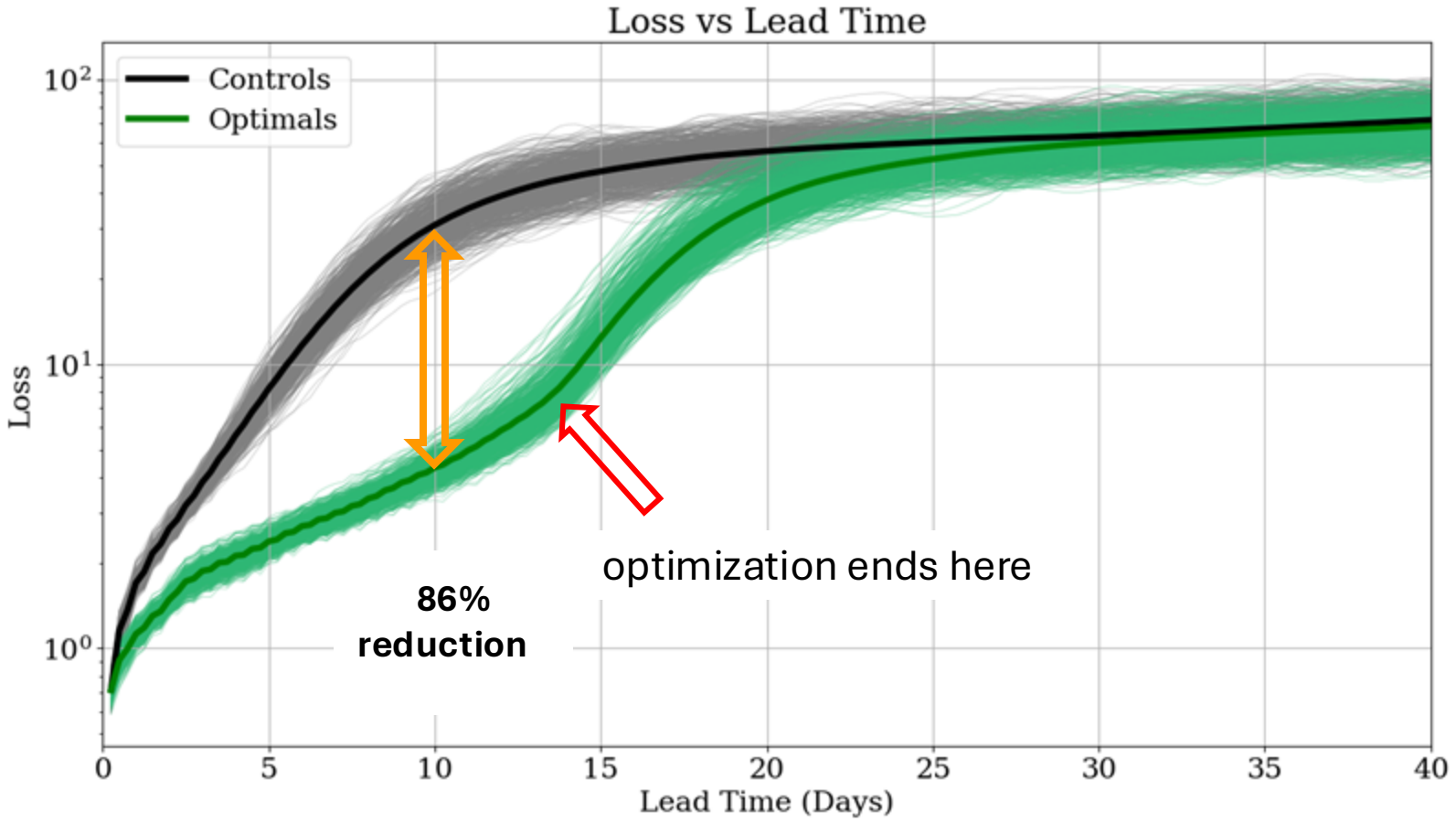


Vonich & Hakim (2024)

Generality beyond the Heatwave Case

GraphCast initial-condition optimization daily for 2020

Optimal Predictability

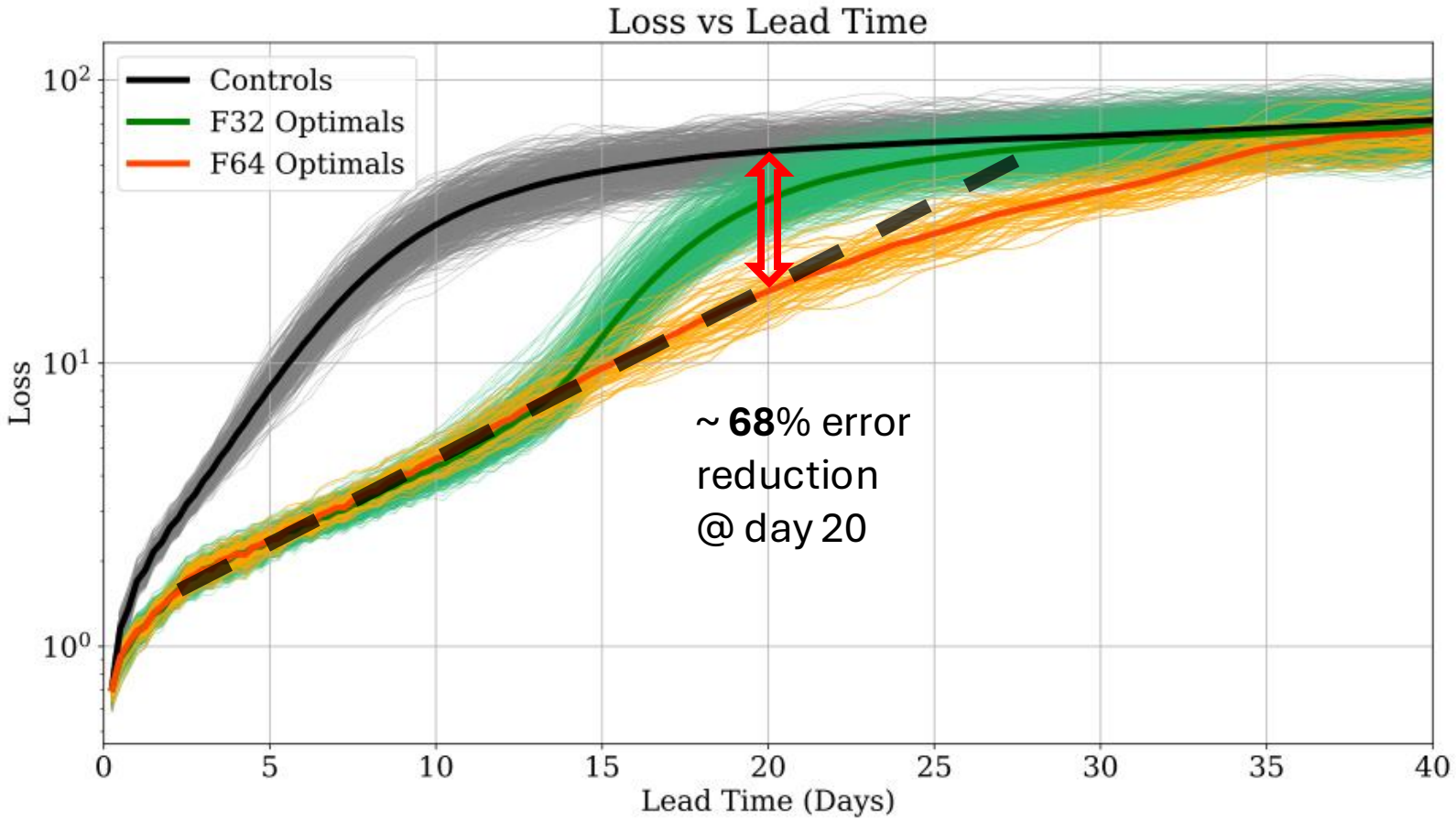


Every case has large improvement

Errors grow like the control after optimization stops

float-32 precision limit

Optimal Predictability Limit ~35+ days



Float precision matters!

- float32 fails ~14 days
- float64 fails ~32 days (out of memory)

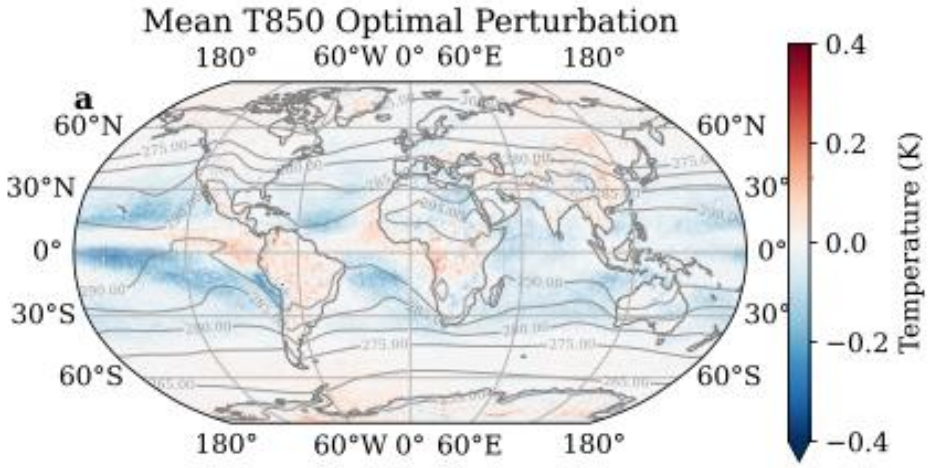
Faster error growth for first 3 days

Constant error growth rate ~3-20 days

- doubling t ~ 5.8 days
- perfect twins ~1.0 day

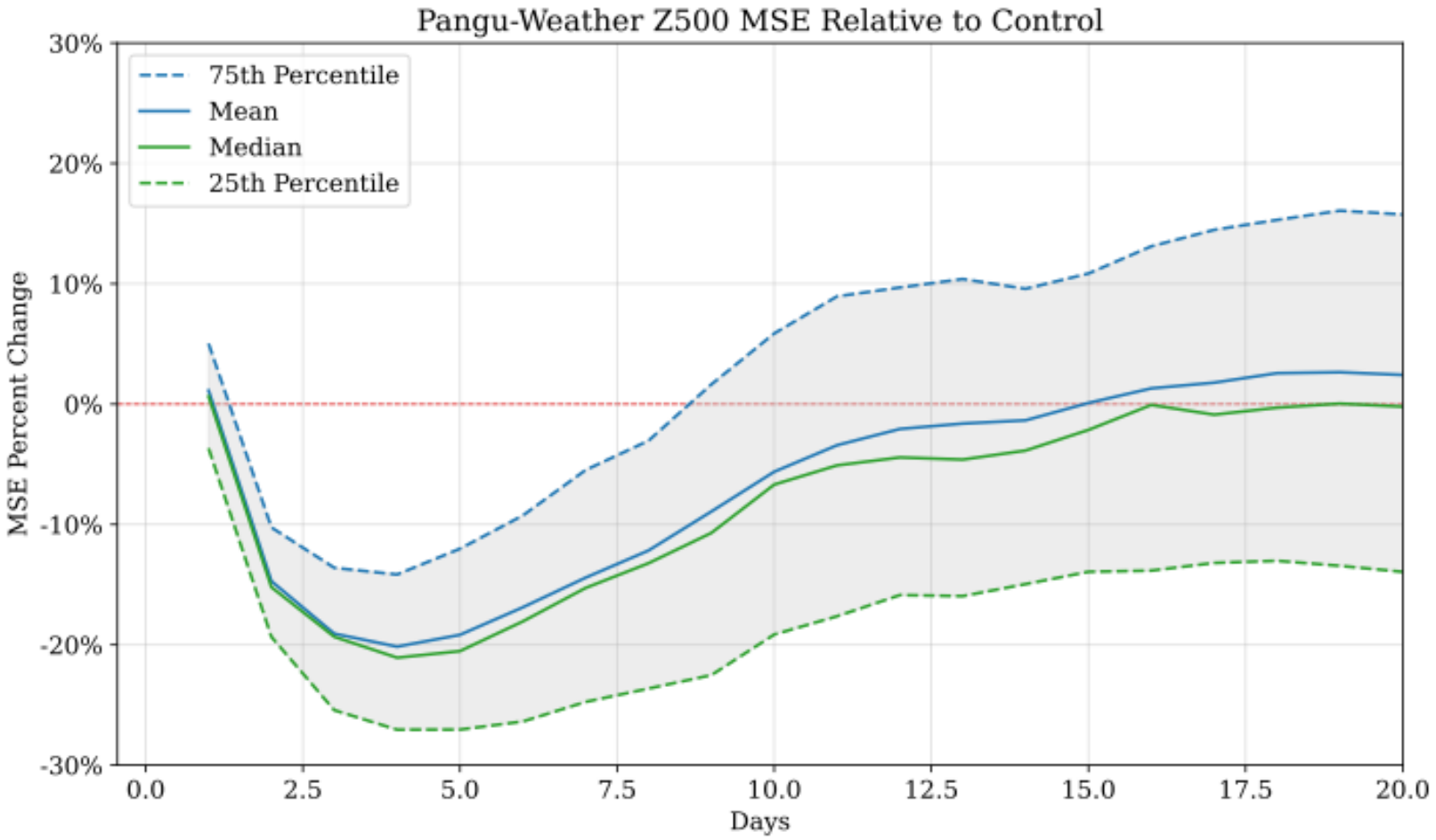
Day 20 skill ~ day 7 for the control

GraphCast Optimals: Time-Mean Structure



	mean	std
200 hPa Zonal Wind	$0.04\,\mathrm{m\,s^{-1}}$	$0.39\,\mathrm{m\,s^{-1}}$
200 hPa Meridional Wind	$0.03\,\mathrm{m\,s^{-1}}$	$0.28\,\mathrm{m\,s^{-1}}$
500 hPa Geopotential Height	$0.63\,\mathrm{m}$	$5.0\,\mathrm{m}$
500 hPa Pressure Vertical Velocity	$8 \times 10^{-4}\,\mathrm{Pa\,s^{-1}}$	$5 \times 10^{-3}\,\mathrm{Pa\,s^{-1}}$
700 hPa Specific Humidity	$0.02\,\mathrm{g\,kg^{-1}}$	$0.08\,\mathrm{g\,kg^{-1}}$
850 hPa Temperature	$0.04\,\mathrm{K}$	$0.33\,\mathrm{K}$

Pangu-Weather Forecasts from GC Optimals



- Float32 (optimized to 14 days)
- Significant improvements
- Much smaller than for GraphCast
- Suggests importance of model error

Summary & Outlook

- Gradient sensitivity over long windows yields deterministic predictability limits
- Small corrections to ERA5: ~80—90% reduction in 10-day forecast error
  - on the order of analysis error
  - average optimal strengthens the Hadley circulation in ERA5
  - error doubling time ~5-6 days (cf ~1-1.5 days for operational forecasts; ML twins)
- **Deterministic predictability limit > 30 days**
- An interpretation: The optimals are shadowing trajectories
  - But drift from ERA5 due to model error; suggests state—model optimization
- Outlook
  - **reanalysis**: seems like a compelling application (optimal analysis & corresponding tuned models)
  - **forecasting**: less clear, but several possibilities

Thank You!