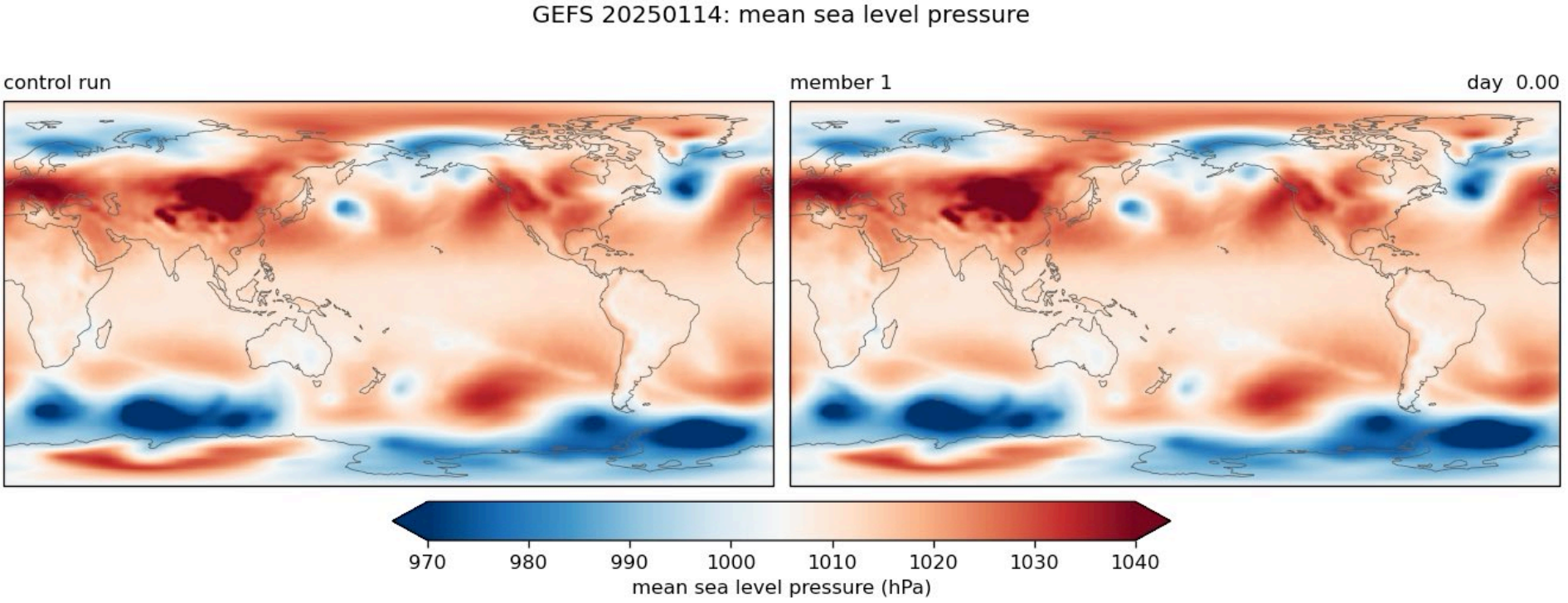
Tropical predictability and error growth in AI weather models

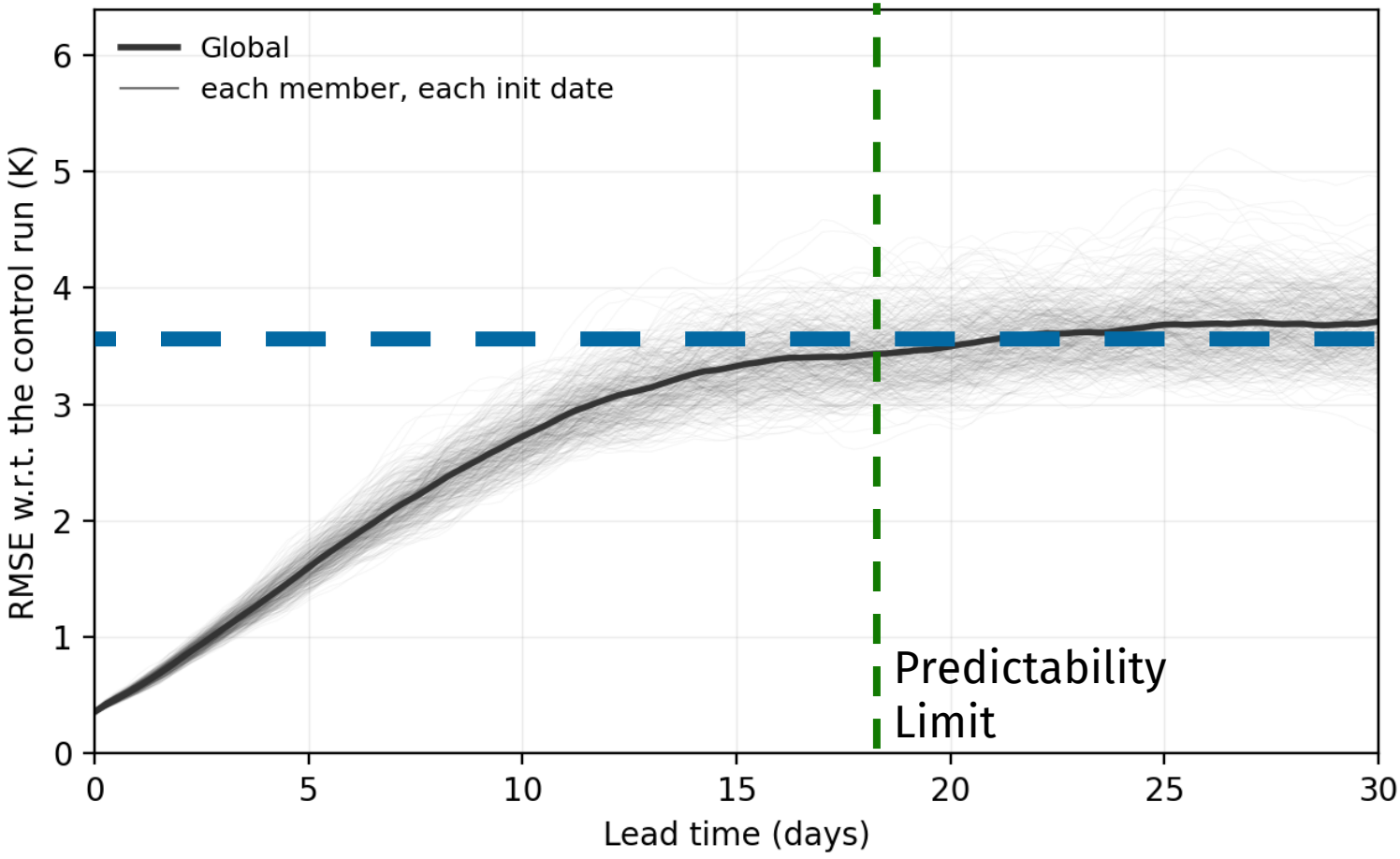
Aditi Sheshadri, Rob King, and Aman Gupta

Ensemble divergence



$$Error = \frac{1}{N_{members} N_{points}} \sum_{i \in E} \sum_{\vec{x} \in D} w(\vec{x}) \left(y^{(i)}(\vec{x}) - y^{(ctl)}(\vec{x}) \right)^2$$

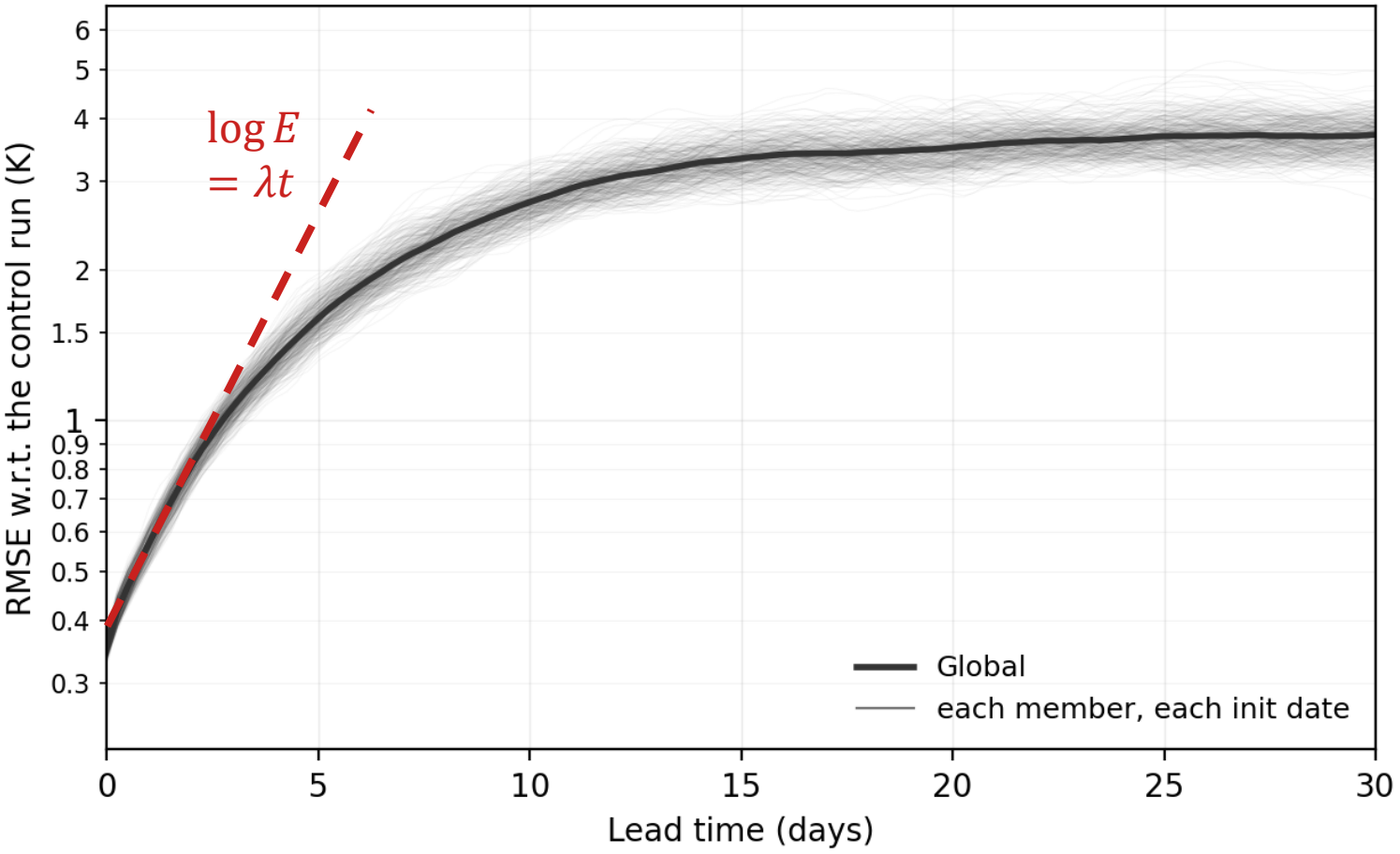
Window of predictability



Error saturation

GEFS Ensemble 2025

Window of predictability



λ Lyapunov exponent

$\tau_{2\times} = \frac{\ln 2}{\lambda}$

Error doubling time

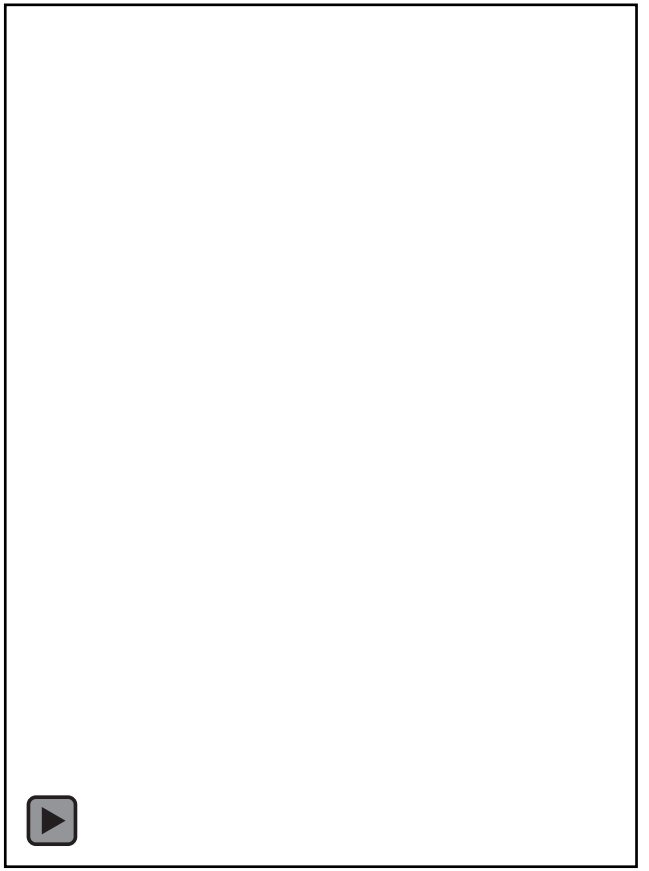
GEFS Ensemble 2025

Parameterization and predictability in numerical models

Two “free parameters”

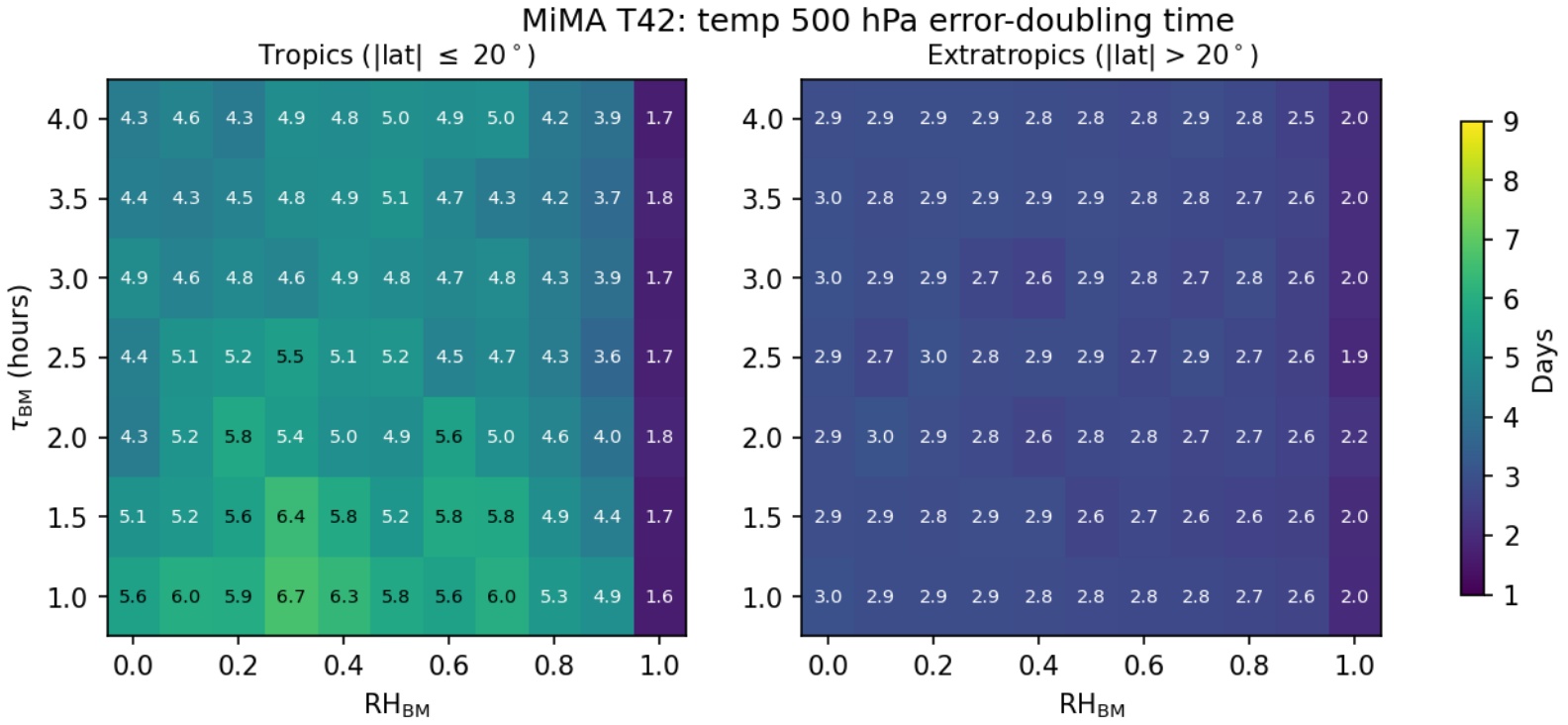
$$\Theta = (RH, \tau)$$

	MiMA
Resolution	T42 (~ 2.0 degrees at equator)
Spinup	10 years
Moisture	✅
Run	90 days (for Jan/Apr/Jul/Oct)
Perturbation	0.01K thermal noise at T1-21
Ocean	Slab Ocean + Q Fluxes
Land	Simple albedo + roughness scheme
Convective	**Betts-Miller**



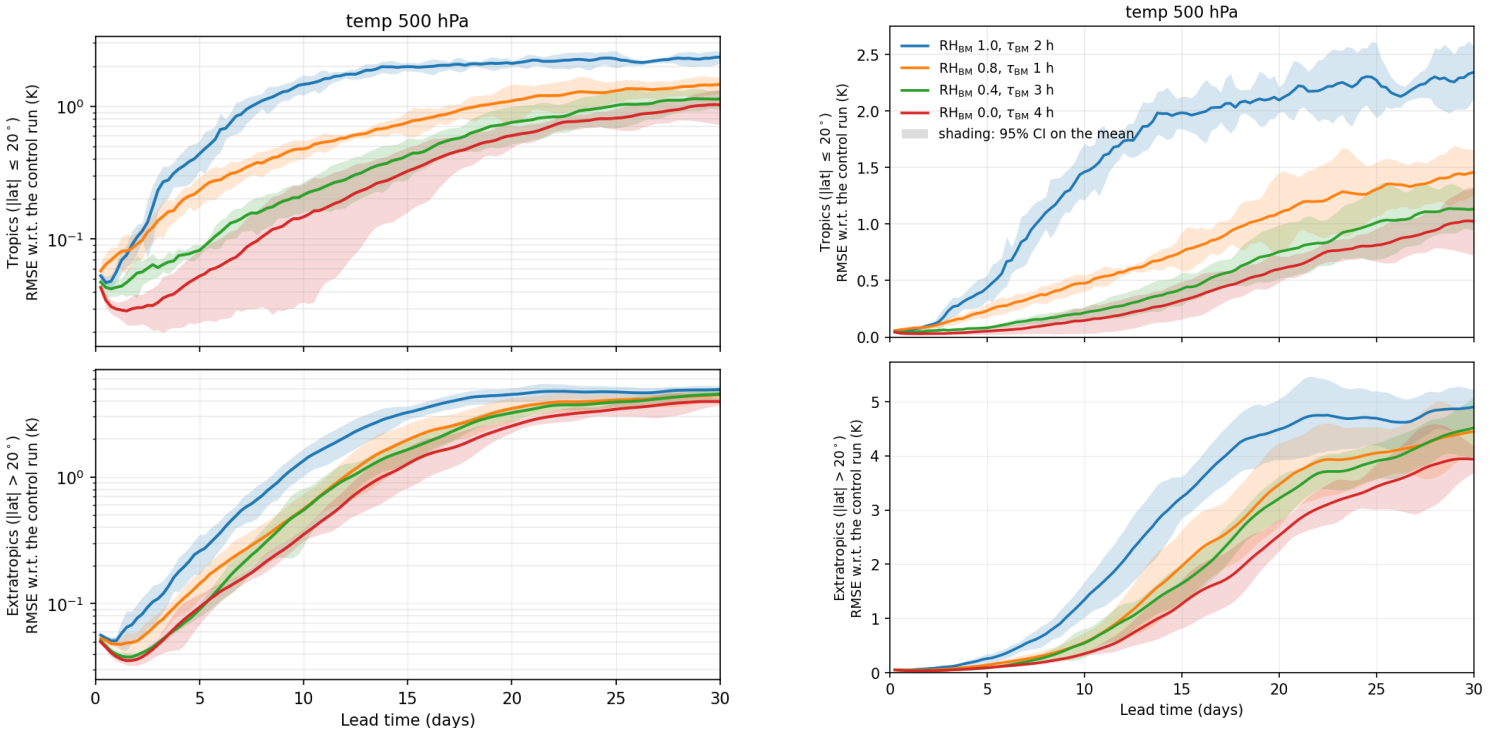
Frierson 2007 (JAS)

Parameterization and predictability in numerical models



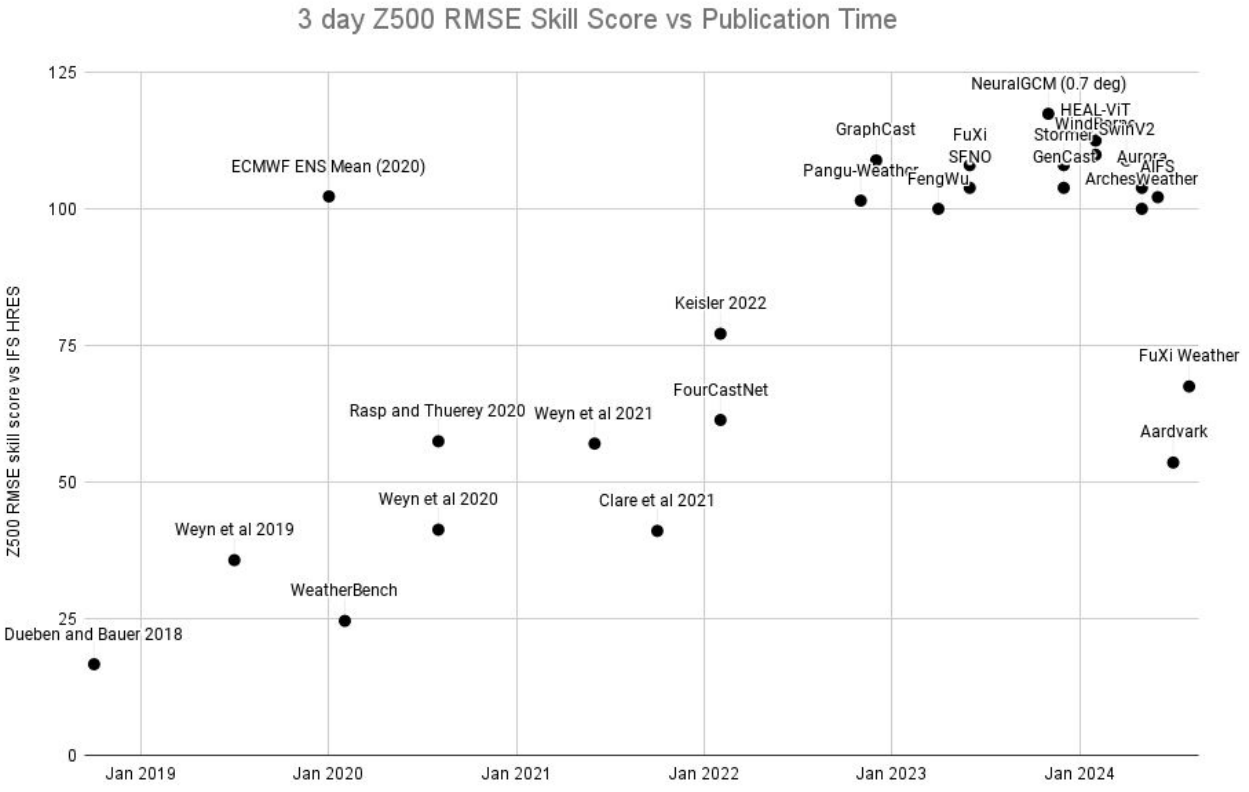
King et al. in prep

Parameterization and predictability in numerical models



King et al. in prep

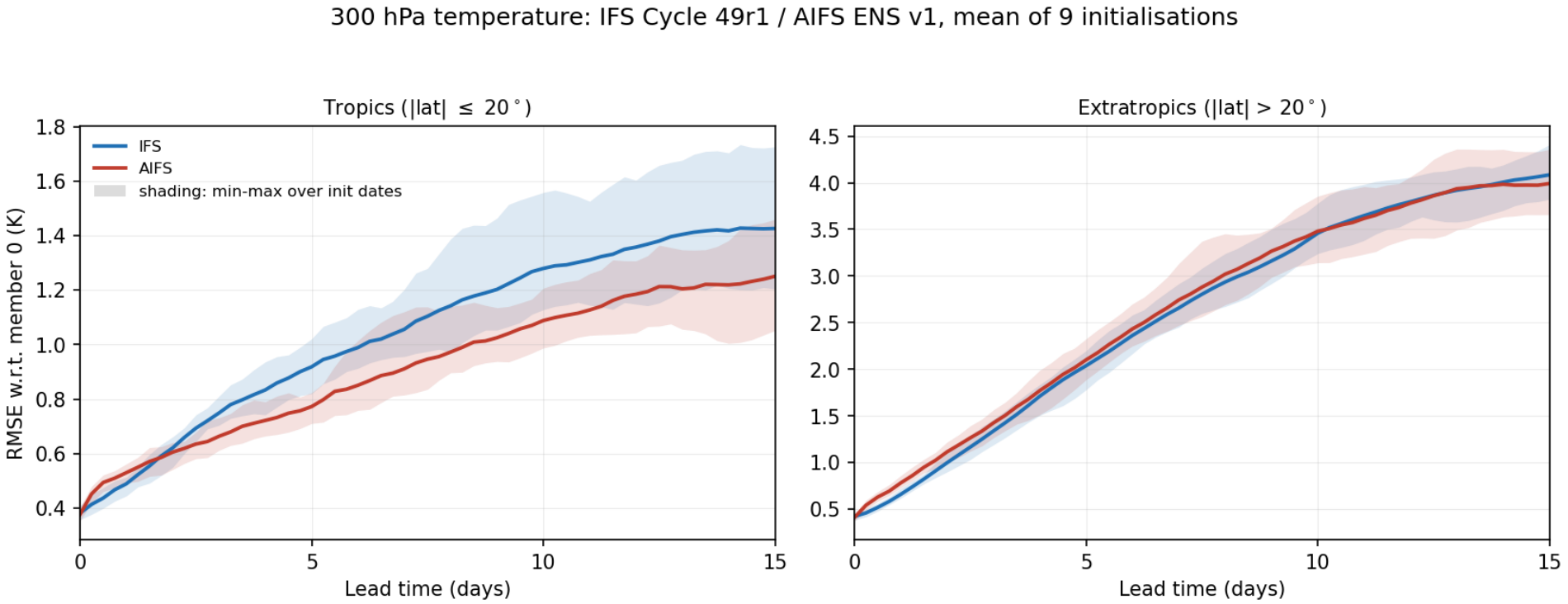
AIWP has improved medium range skill



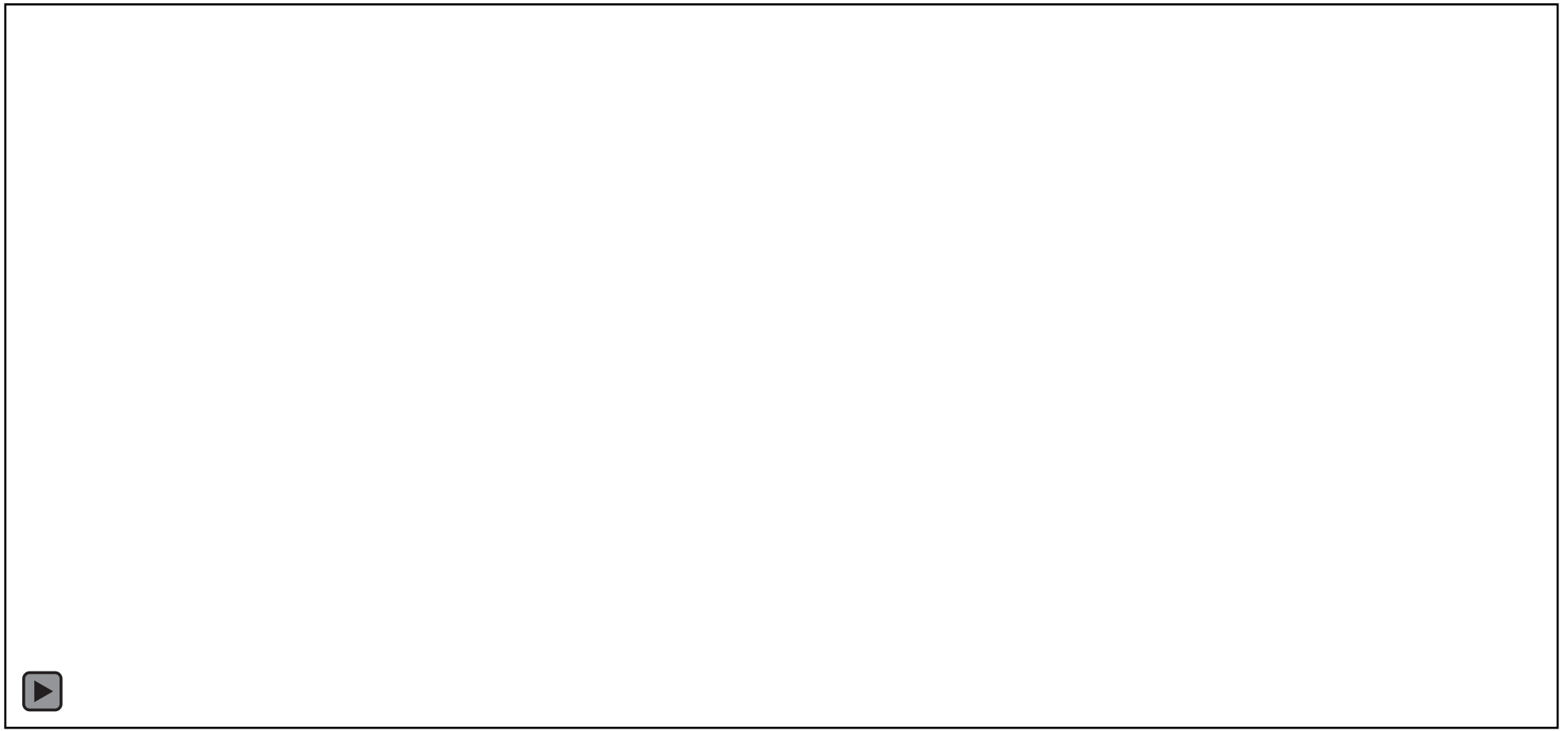
Is there diversity in the representation of convection in AIWP? Does it impact predictability?

(Image Source: Stephan Rasp)

Does that imply slower error growth?



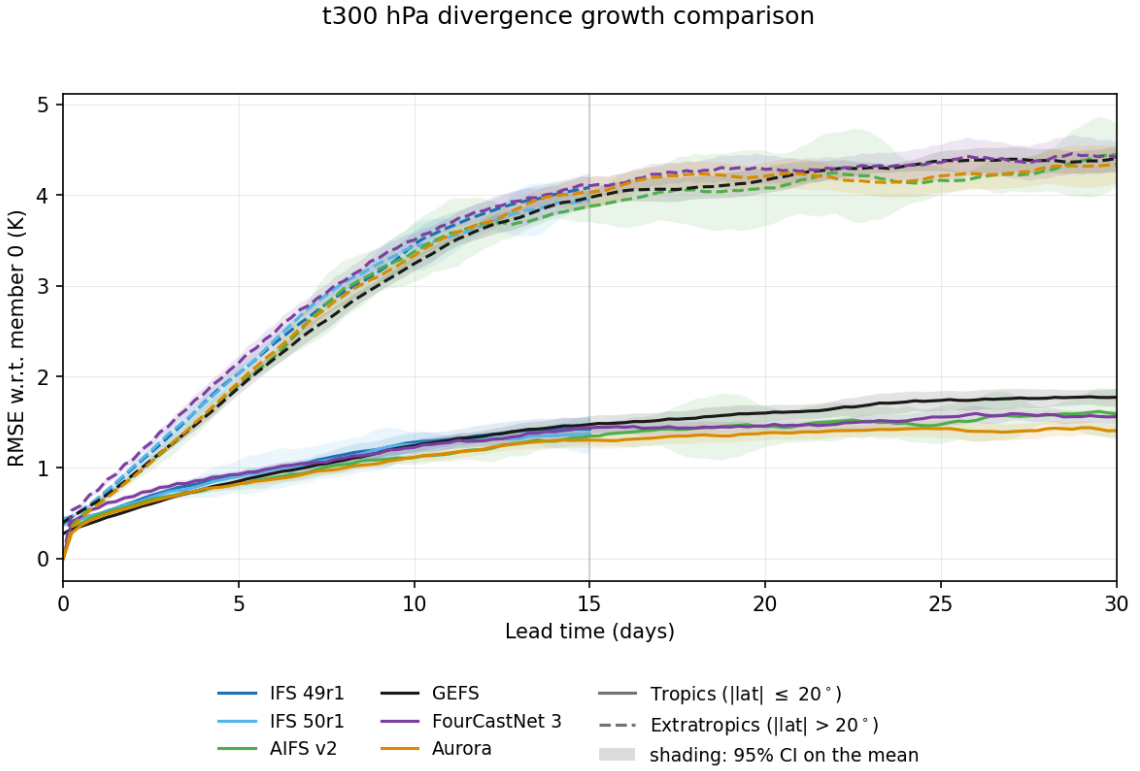
But AIFSv1 blows up!



Accuracy not because of slower error growth. So, less model error?

AIWP models have different architectures + training data

All posses similar error growth times!



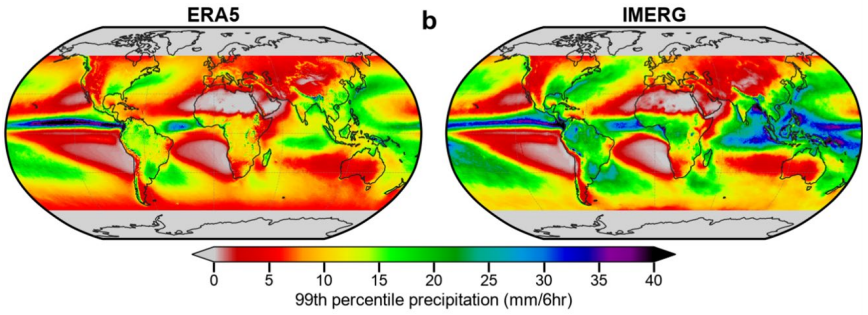
Physics based NWP

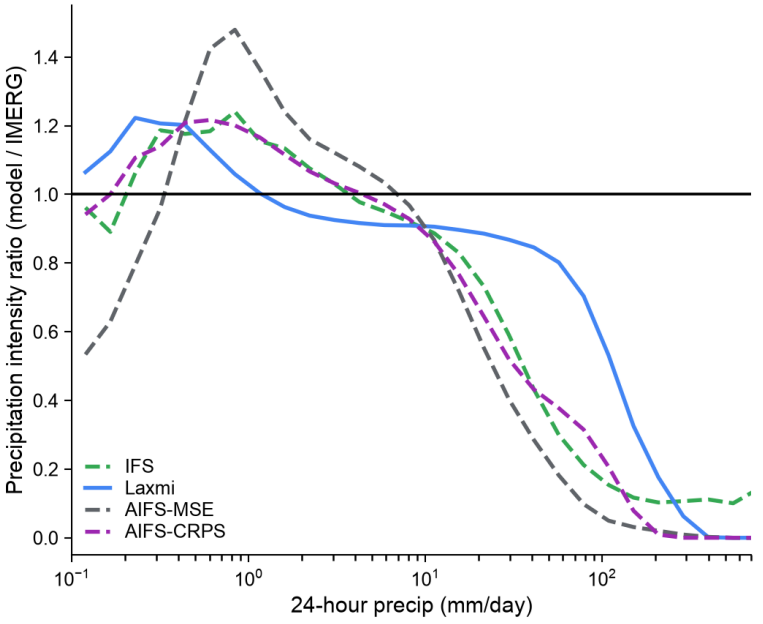
IFS 49r1
IFS 50r1
GEFS

AIWP

AIFS v2
FourCastNet 3
Aurora

The next frontier for AIWP – customizing with observational data

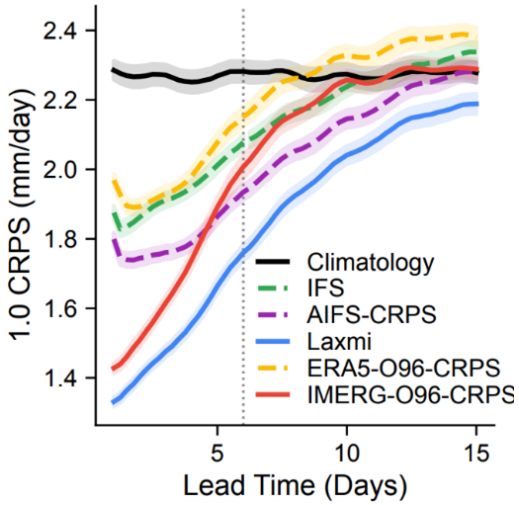




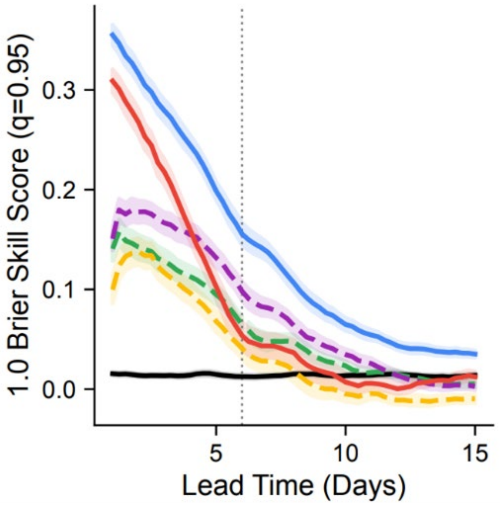
Most AI weather models **severely underestimate** extreme precipitation. Training data is part of the problem and can be addressed by training directly to observational datasets.

Our model, Laxmi, is a retraining of AIFS-CRPS (a graph-transformer) to IMERG that captures 68% of the observed precipitation frequency above 50 mm per day. IFS and AIFS-CRPS capture 30% at 0.25° resolution.

Laxmi also places precipitation extremes in the right place and time



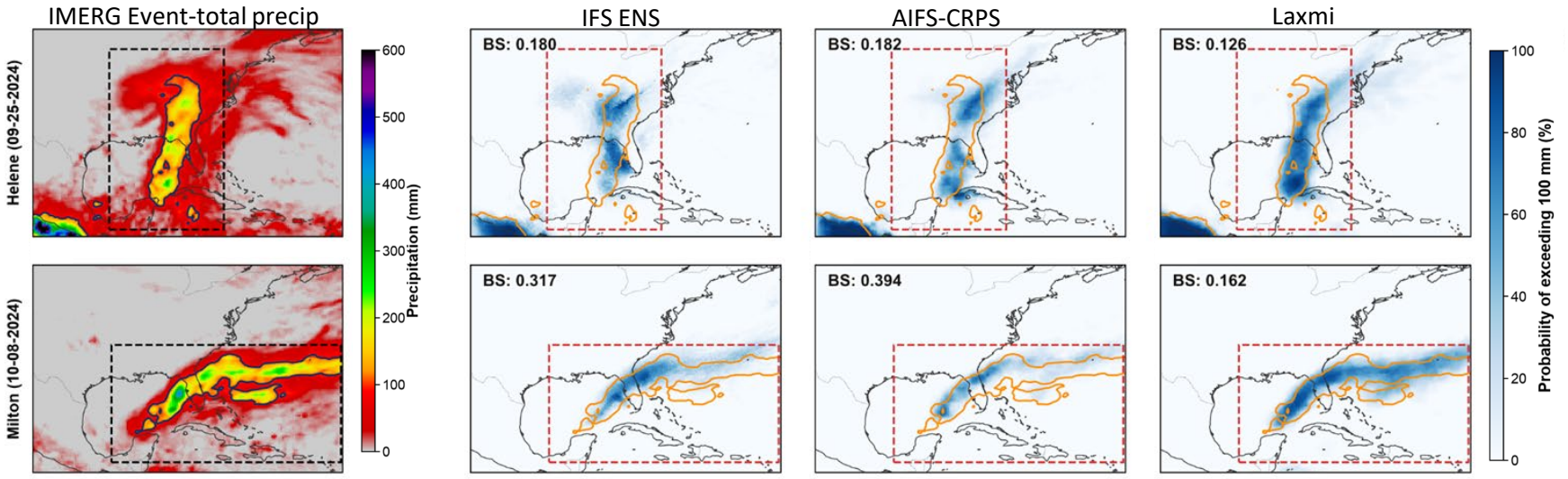
19% improvement in global CRPS at 0.25°



47% improvement in global Brier skill score at the 95th percentile at 0.25°

We train our architecture at 2 resolutions: 0.25° and 1°, finding higher resolution is essential for precipitation forecasts, particularly at longer lead times (blue vs red curves).

Substantial improvement in 3-day ahead forecasts for TCs



Observed precipitation exceeding 100 mm is outlined in blue at left and highlighted in gold at right. Brier scores are computed over the red dashed box for each model.

Price to train Laxmi

$153k

Training cost

128

A100 GPUs for 10 days

Julian Schmitt, Bertrand Delorme, Rob King, Yashica Patodia, Tapio Schneider, Aditi Sheshadri, Ravi Jain