Predictability constraints on medium-range weather prediction

George Craig

Tobias Selz

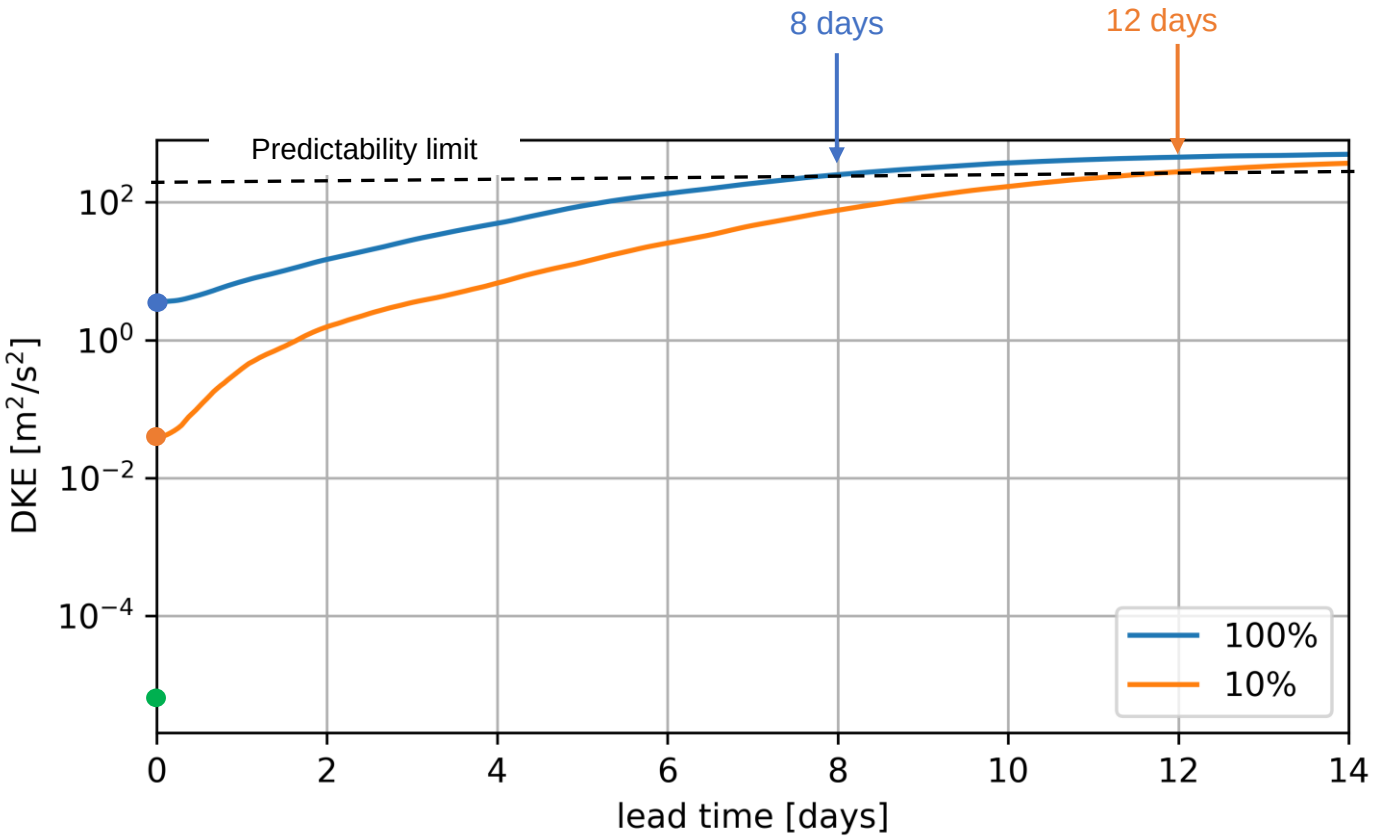
Meteorologisches Institut München, Ludwig-Maximilians-Universität München , Germany

Questions

Focus on medium-range forecasts of synoptic pressure, temperature, wind in the mid-latitudes

1. How close is the limit of predictability? – *1 more week*
2. Where is the most potential for improvement? – *DA*
3. Can data-driven models help? – *depends on the data*

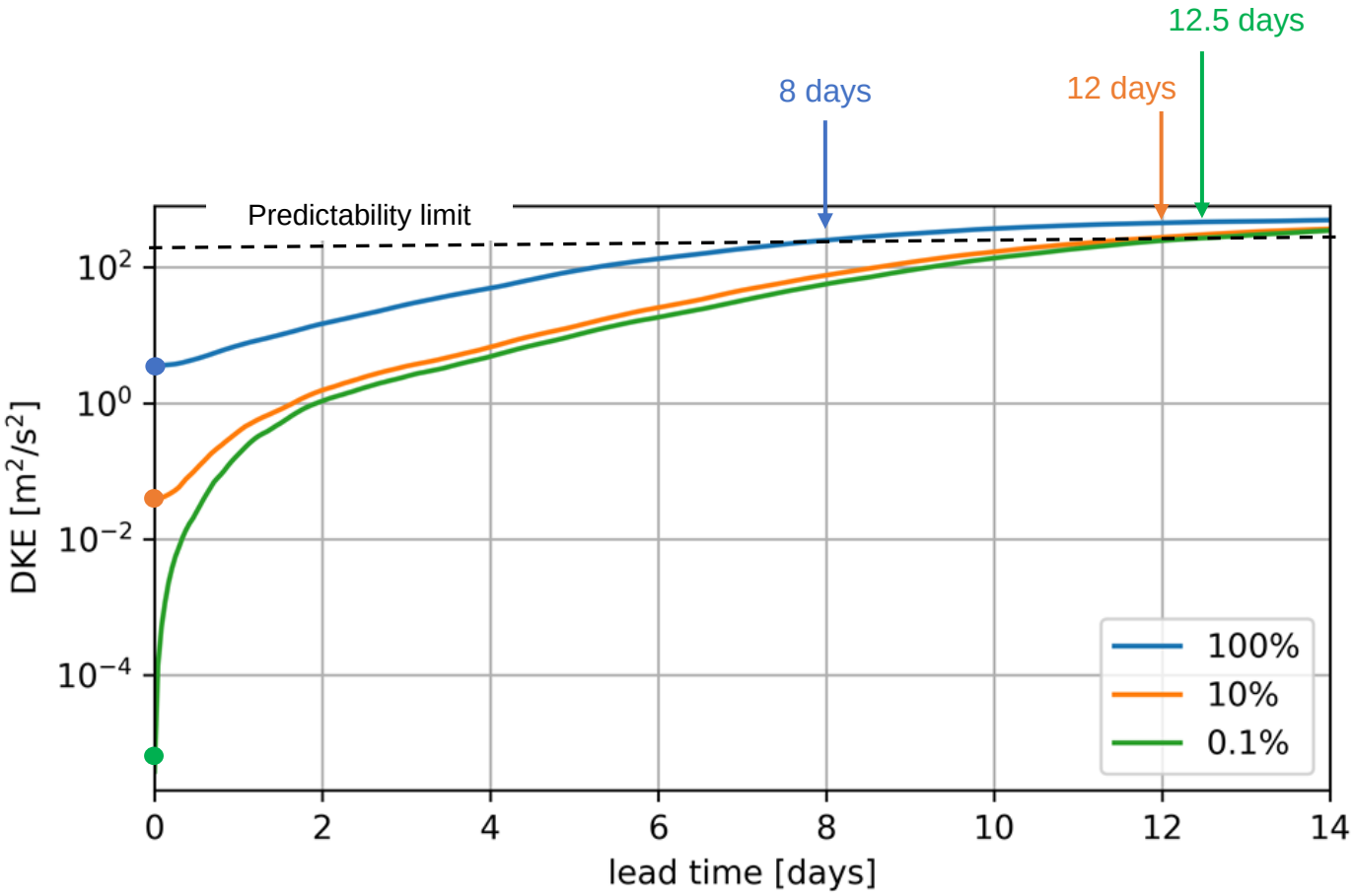
Error growth and predictability



Chaos (Lorenz 1963)

- Sensitive dependence on initial conditions
- Exponential error growth
  Expect saturation after finite time – limited predictability
- Can extend length of accurate forecast by reducing initial condition errors

Error growth and predictability

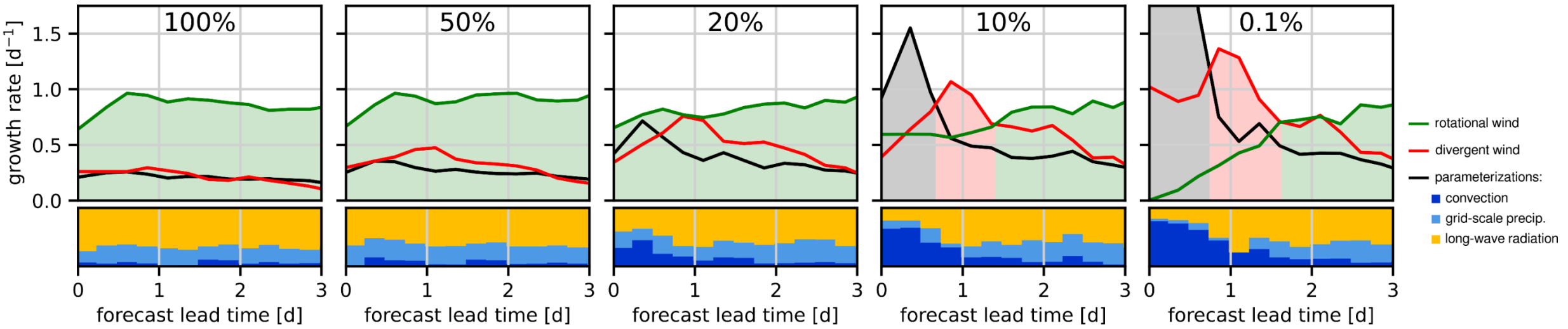


The “real” butterfly effect (Lorenz 1969)

- Faster error growth on small scales and rapid saturation
- Upscale growth by nonlinear interactions
- Diminishing returns from improved initial conditions

PV Error growth mechanisms

Contributions to perturbation/error growth in tropopause-level Potential Vorticity (Baumgart et al. 2019)



Current initial errors sampled from EDA

Error growth from

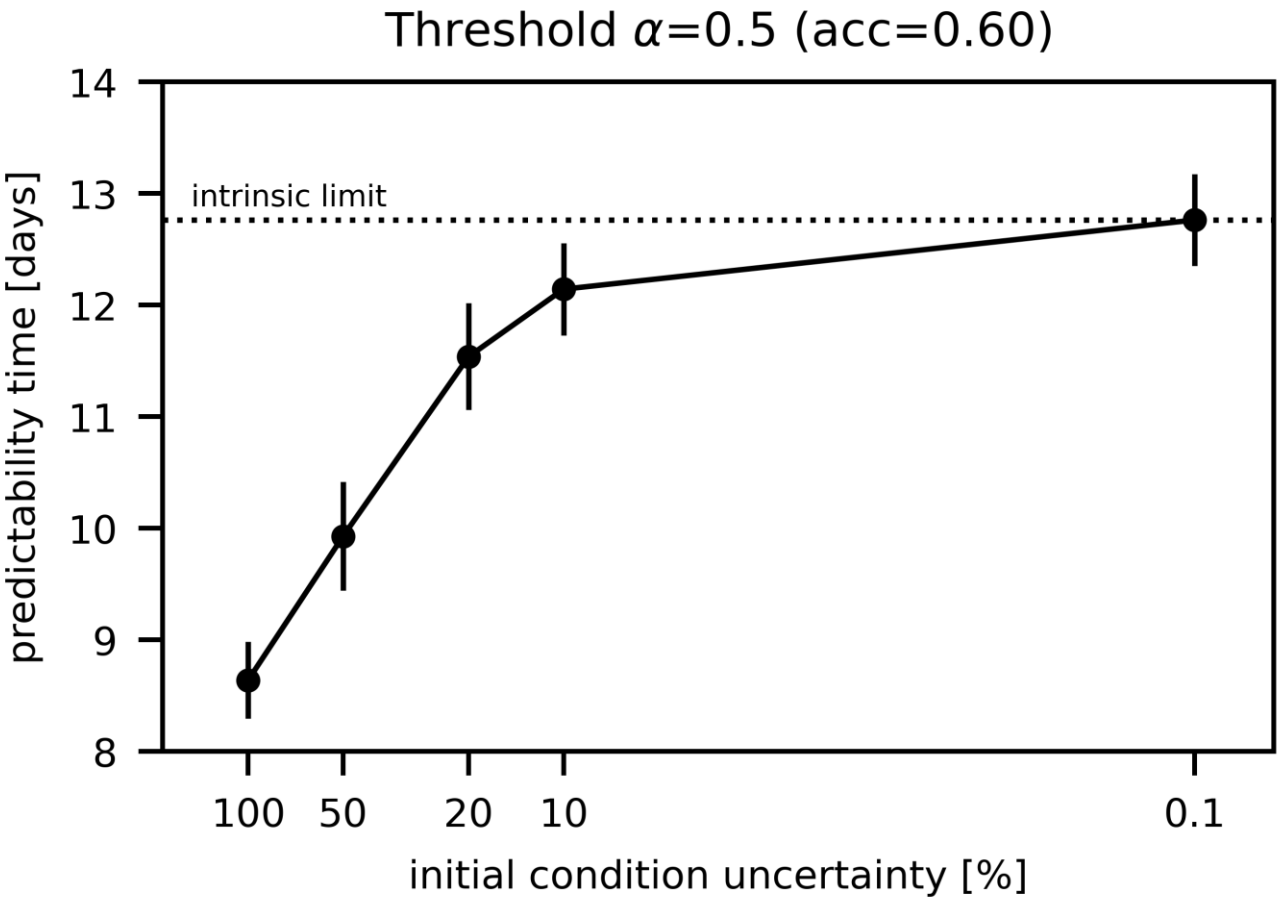
- Advection by perturbation rotational wind (quasi-barotropic, Lorenz theory)

Small initial errors scaled in amplitude

Three stage error growth

1. Convection scheme
2. Advection by perturbation divergent wind
3. Advection by perturbation rotational wind

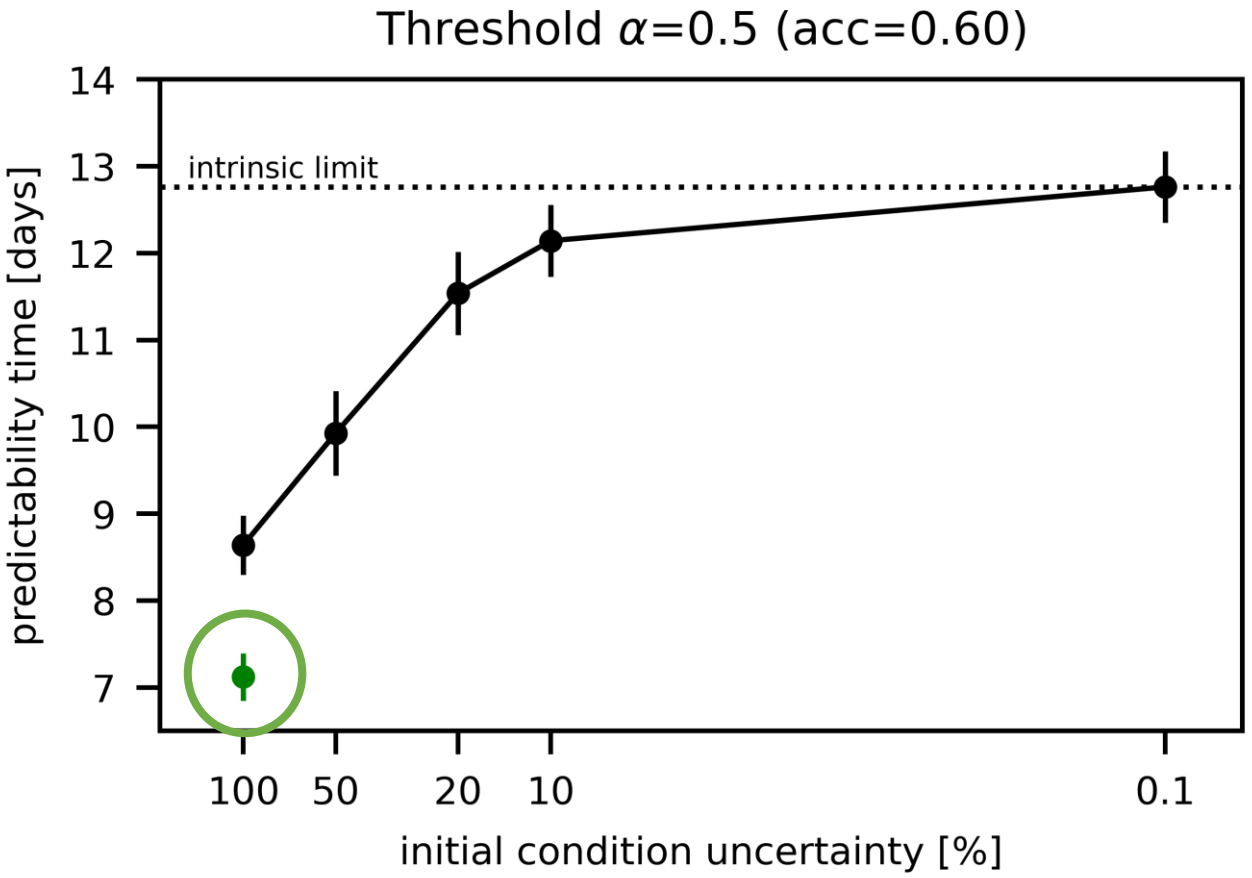
How close is the limit?



Predictability time

- Time when difference energy reaches 50% of saturation (ACC ~ 0.6)
- Intrinsic limit: 13 days
- Current initial errors: 8.5 days
- Reach limit when initial condition error reduced to about 10% of current level

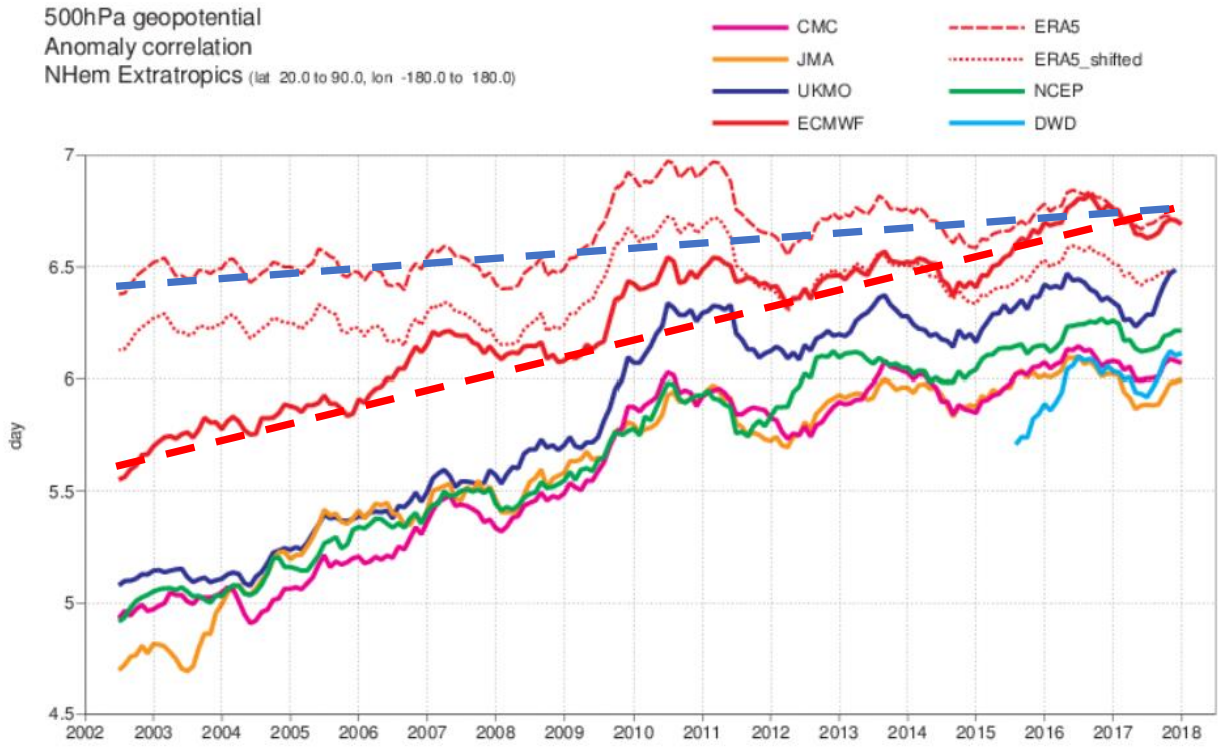
Improvement from perfecting model



Predictability time

- Time when difference energy reaches 50% of saturation
- Intrinsic limit: 13 days
- Perfect model: 8.5 days (4-5 days improvement possible with better initial conditions)
- ECMWF: 7 days (1-2 days additional improvement if model perfected)

Use of observations vs new observations



ERA5 reforecasts vs operational skill

- Day where 500 hPa ACC for Z500 drops below 80%
- Operational skill has improved by 1.1 days (2002-2018)
- Reforecast skill has improved by 0.3 days
- ~25% of improvement due to observation network

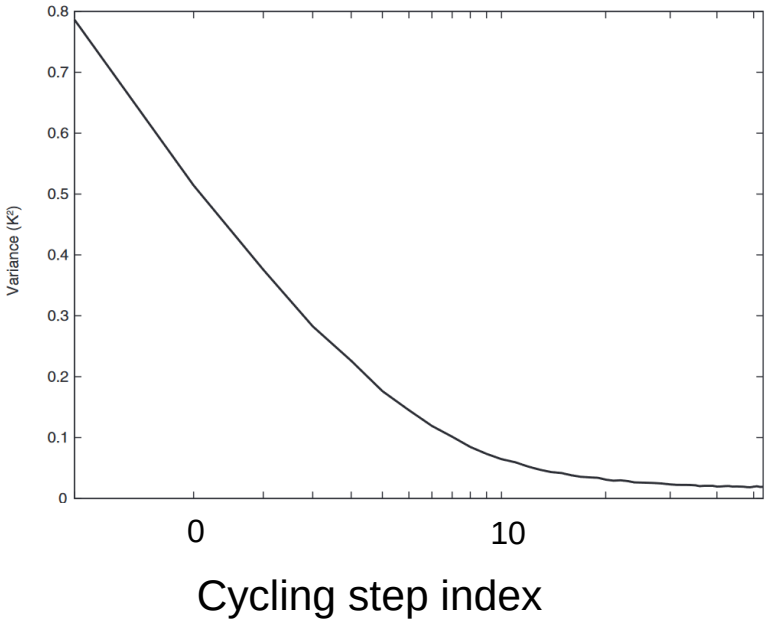
(Hersbach et al. 2018)

Use of observations can be improved

Decay of observation impact in data assimilation cycling

$\mathbf{e}_0^a$	Initial analysis error $\mathbf{e}_0^a$
$\mathbf{e}_1^b = (\mathbf{M}_0\mathbf{e}_0^a + \mathbf{e}_0^m)$	Increased by model dynamics $\mathbf{M}_0$ and model error $\mathbf{e}_0^m$
$\mathbf{e}_1^a = (\mathbf{I} - \mathbf{K}_1\mathbf{H}_1)\,\mathbf{e}_1^b + \mathbf{K}_1\mathbf{e}_1^o$	Reduced by factor $(\mathbf{I} - \mathbf{K}_1\mathbf{H}_1)$ during assimilation, and increased by observation error $\mathbf{e}_1^o$

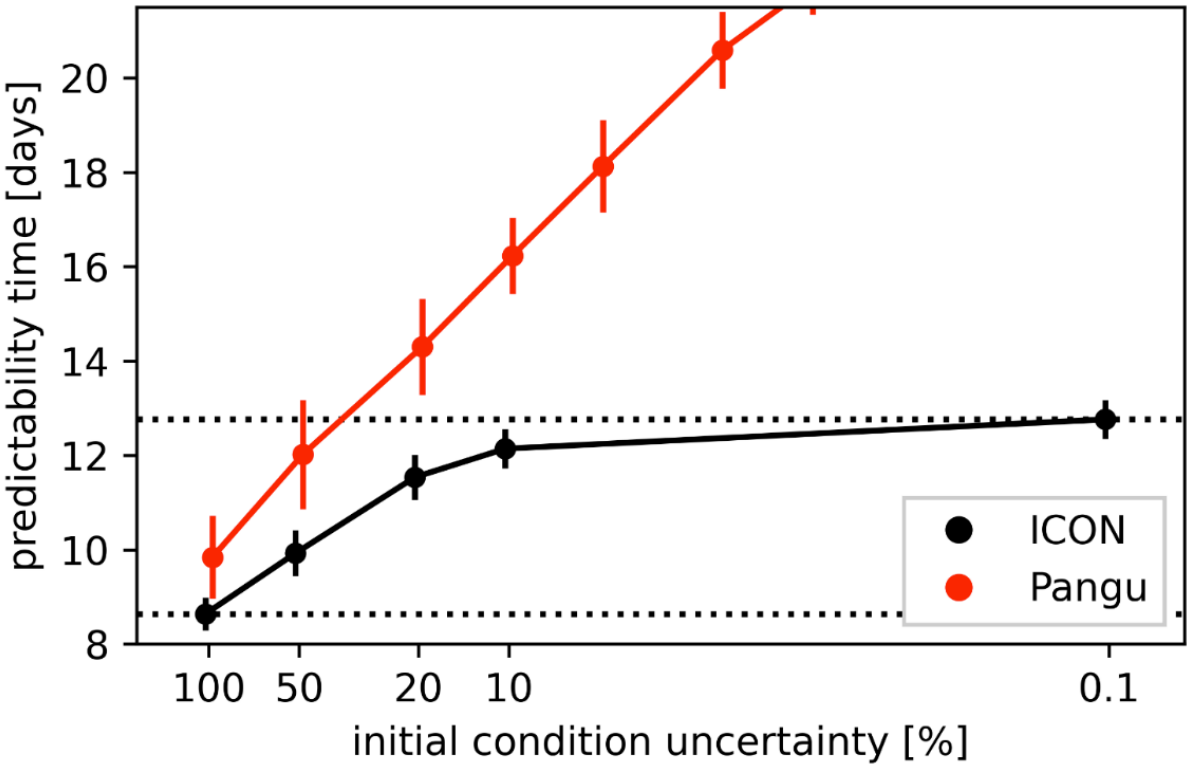
Contribution of background state to 500 hPa temperature variance (K^2)



(Berre 2018, Hubans et al. 2022, MétéoFrance)

- Exponential decay of background information
- Contribution negligible after 8-10 12h cycles
- On average, a forecast uses 4-5 days of data
- Significantly less than predictability time

Can AI models simulate the butterfly effect?



Predictability time

- Time when difference energy reaches 50% of saturation (ACC ~ 0.6)

Pangu:

- Slower error growth for current IC uncertainty
- No evidence of predictability limit

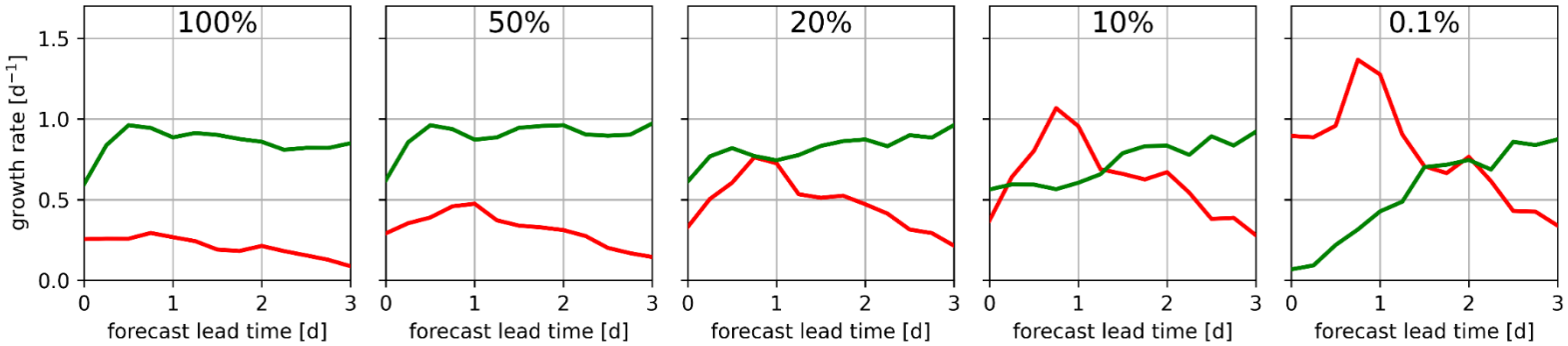
Results for Pangu, GraphCast, FourCastNet, NeuralGCM, Aurora, GenCast

See poster of Tobias Selz, Thu 12:30-14:00

(Selz and Craig 2023)

PV diagnostic from AI model

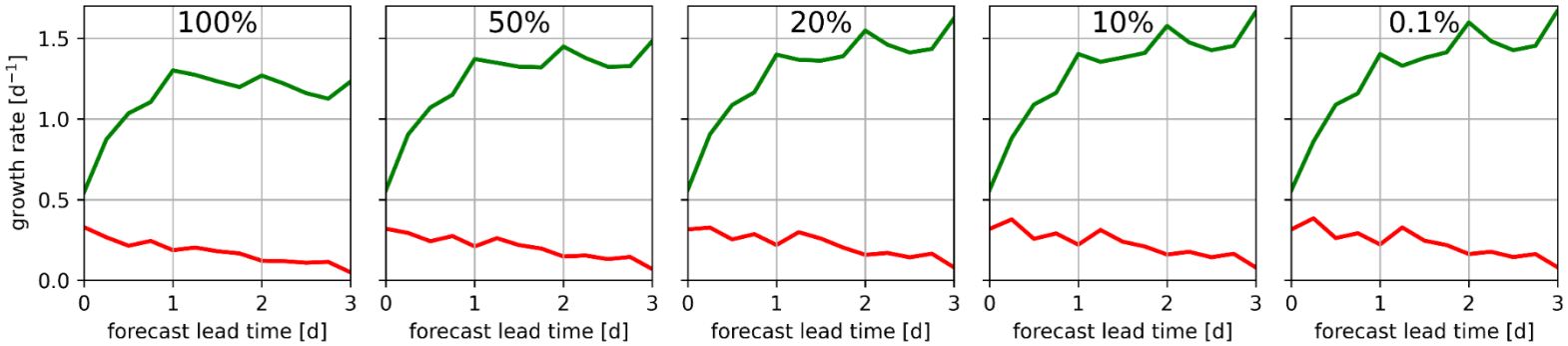
ICON simulations



PV error growth rate due to advection by errors in:

- rotational wind
- divergent wind

Pangu simulations



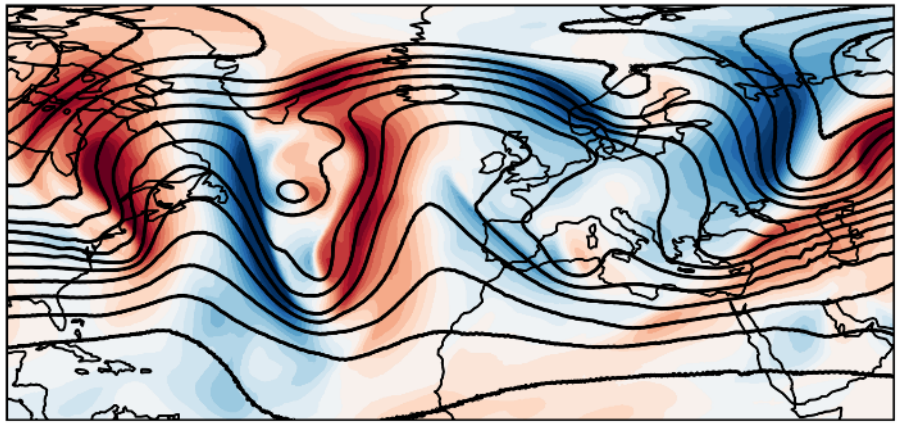
- Simplified version of PV error growth diagnostic
- Pangu fails to capture rapid initial error growth mechanism – why?

(Selz and Craig 2023)

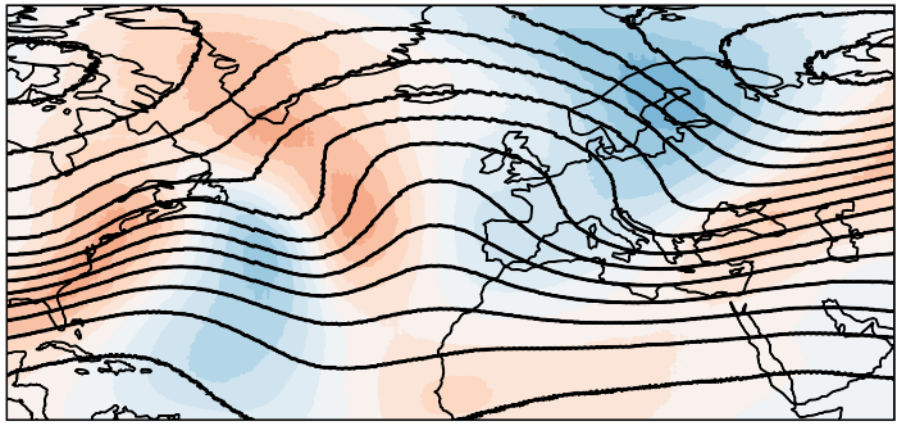
Effective resolution of AI models

Aurora 10d forecasts of meridional wind, geopotential

Aurora-S



Aurora-L



Base model
trained on 12h step

LoRA fine-tuned
trained on 10d rollout

(Selz et al. 2025, https://essopenarchive.org/doi/full/10.22541/essoar.174139239.94807670)

Loss function and ensemble mean

- Loss function is sum of errors over roll-out training period – typically up to 10d in 6-12h steps

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{p(x_{t_n},\ldots,x_{t_1},x_{t_0})} \left[\sum_{t'=t_1}^{t_n} ||x_{t'} - \hat{x}^{\theta}_{t'}(x_{t_0})||^2 \right]$$

- Equivalently, optimal prediction follows mean of predictive distribution over training period

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{t'=t_1}^{t_n} \mathbb{E}_{p(x_{t_0})} \left[||\mu_{t'|t_0} - \hat{x}^{\theta}_{t'}(x_{t_0})||^2 \right]$$

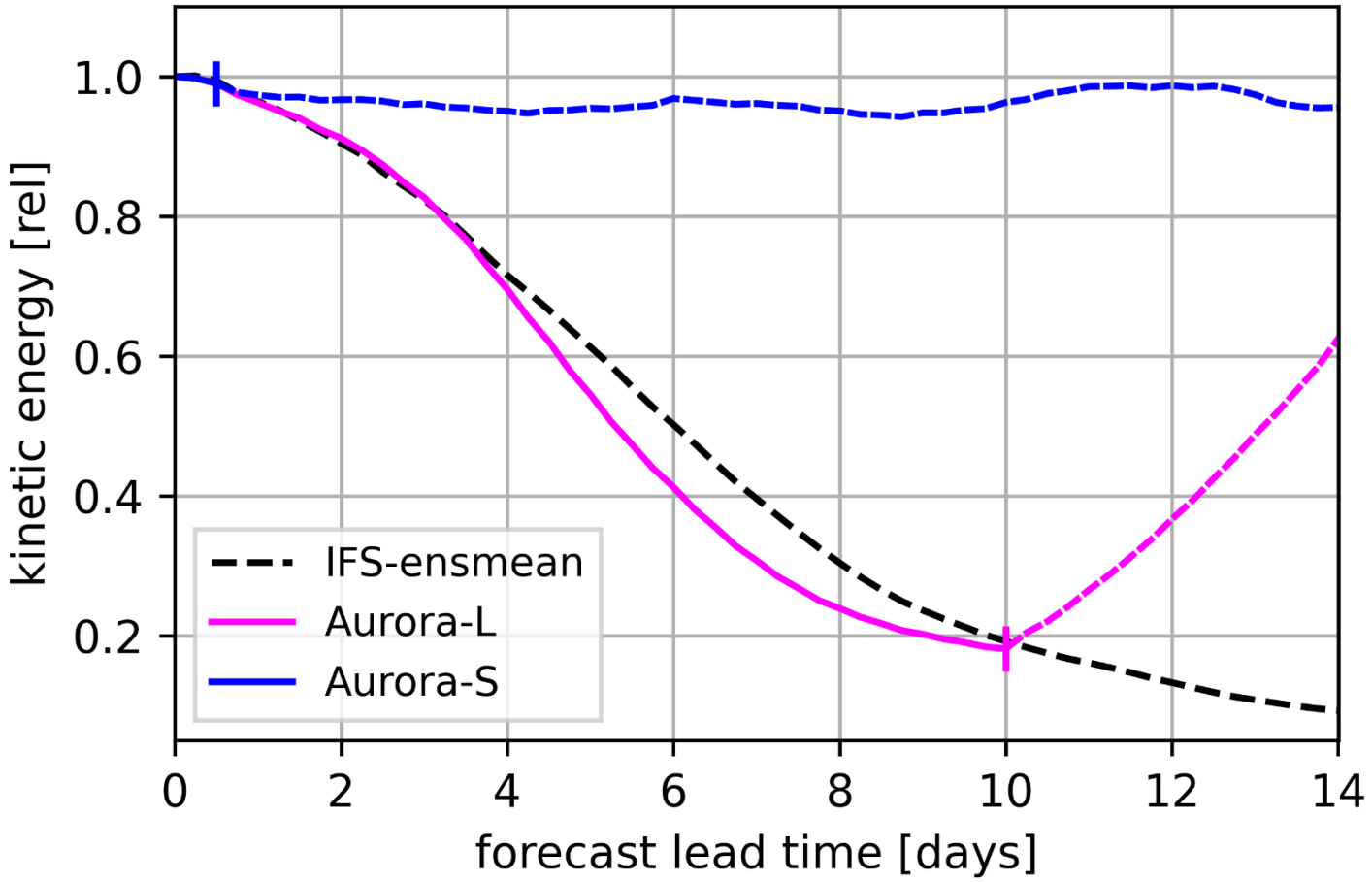
Low effective resolution is a side-effect of rapidly saturating modes being averaged out

Remarks:

1. Objective achieved in limit of large data, training time, model capacity (e.g. parameter count)
2. Objective approached from smooth side – gradient descent learns large scales first
3. After training period, behavior is unconstrained

Effective resolution of AI models

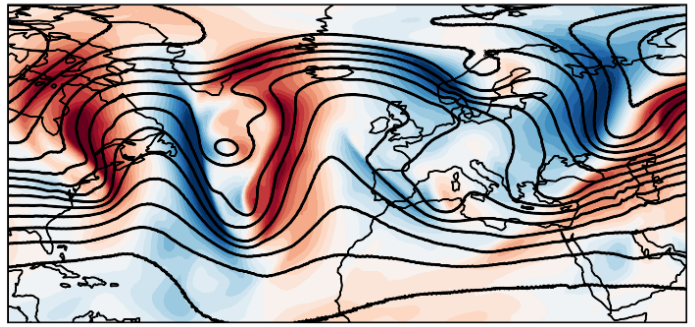
300 hPa kinetic energy between 400 km and 4000 km wavelength, relative to initial condition



(Selz et al. 2025)

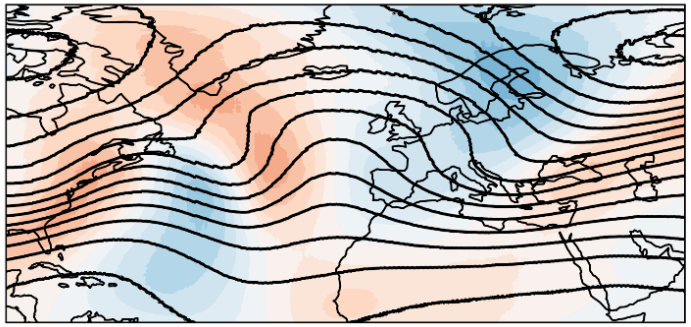
Aurora 10d forecasts of meridional wind, geopotential

Aurora-S



- trained on 12h step

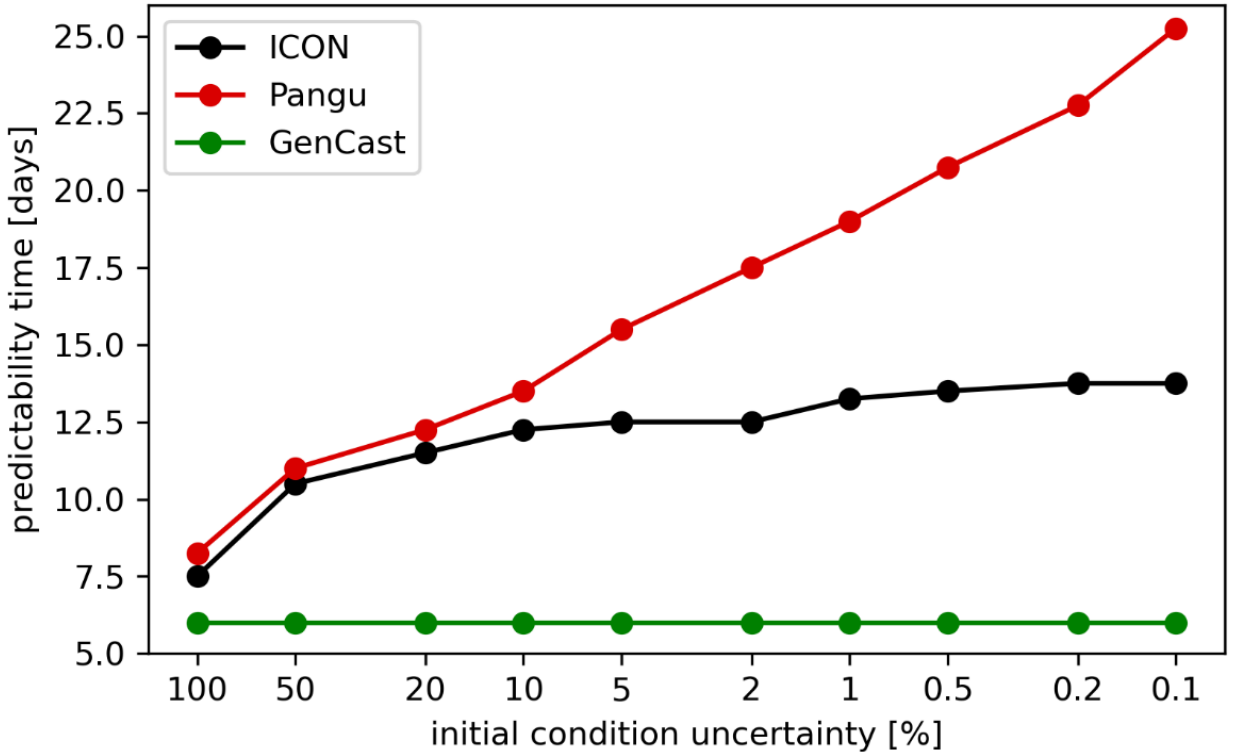
Aurora-L



- trained on 10d rollout

Can generative models simulate the butterfly effect?

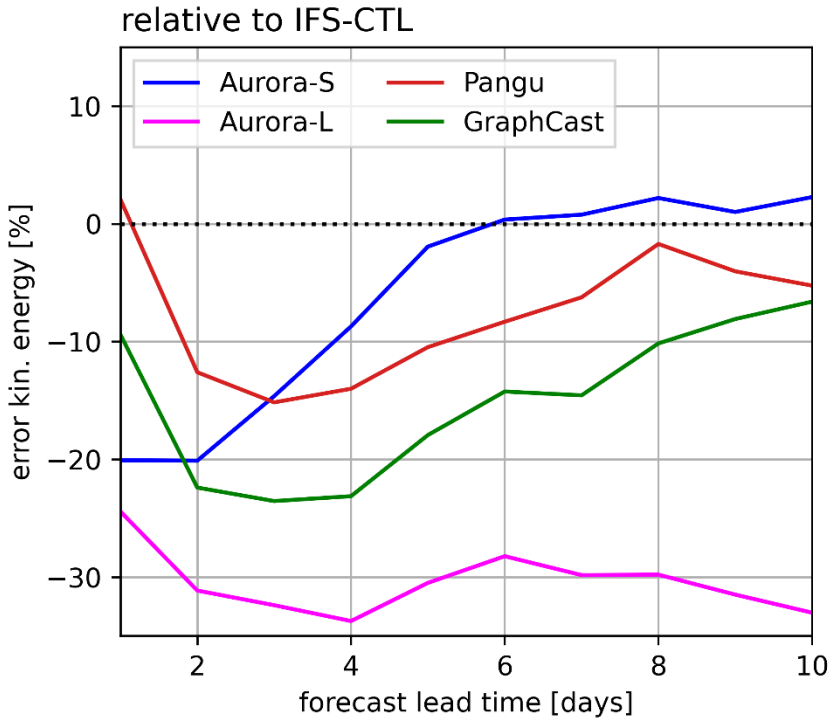
- Reverse normalizing flow maps Gaussian noise to a sample drawn from the forecast distribution, $p(x_f \mid x_i)$
- Trained on x_f , x_i pairs drawn from reanalysis
- IC errors sample current analysis uncertainty and forecast errors are conditioned on that uncertainty
- **Result:** forecast error is almost independent of IC error

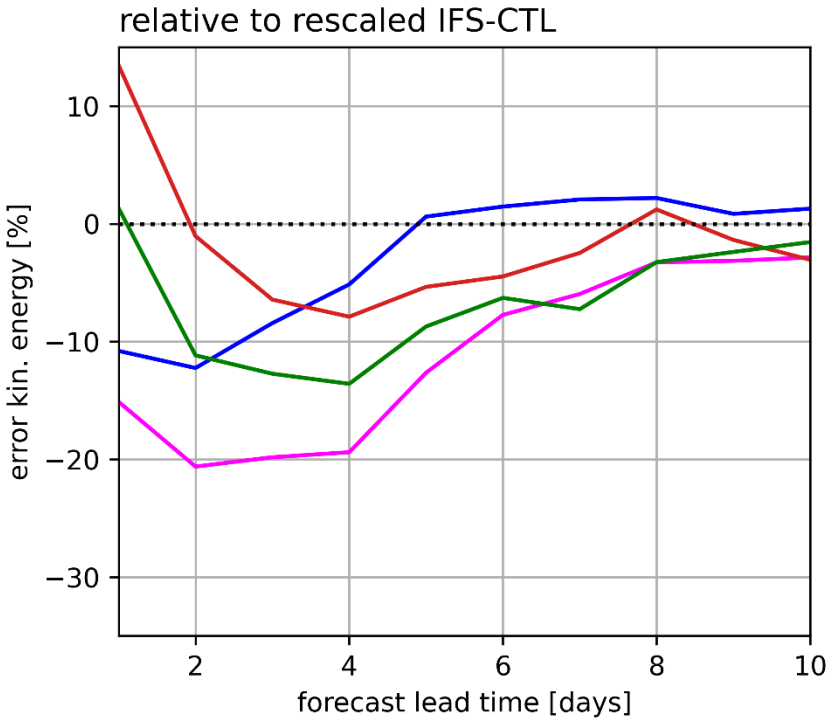


Impact of low effective resolution

Forecast error of AI models relative to IFS-CTL

Forecast error of AI models relative to an IFS-CTL that has been spectrally smoothed to same effective resolution





Implications

- Predictability can be extended by up to another week – limited by transition to rapid initial error growth
- Most of improvement (~4 days) can come from improving DA
- Improvement from model alone smaller – AI models have already significantly closed gap to perfect model
- Need to work from obs. (advantage to weather services) – cycling or 4-5 days of observations needed
- Reanalysis data samples current IC errors – data-driven models target properties of forecast distribution for current levels of IC error – subject to training/capacity limits

Extras

PV error diagnostic (Baumgart et al. 2019)

- Generalization of Lorenz (1969) framework to full equations
- Potential vorticity:

$$P = (\zeta_\theta + f)/\sigma$$

- PV equation:

$$\frac{\partial P}{\partial t} = -\boldsymbol{v} \cdot \nabla_\theta P + N + res$$

  - Piecewise PV inversion

$\boldsymbol{v} = \boldsymbol{v}_{neartropopause} + \boldsymbol{v}_{troposphericdeep} + \boldsymbol{v}_{divergent}$

  - Non-conservative

$N = convection + g.s.precip. + l.w.radiation + \ldots$

  - Residual (Mainly diffusive)

PV error equation

(Rate of change of area-averaged error potential enstrophy near tropopause)

= *(Barotropic advection)*

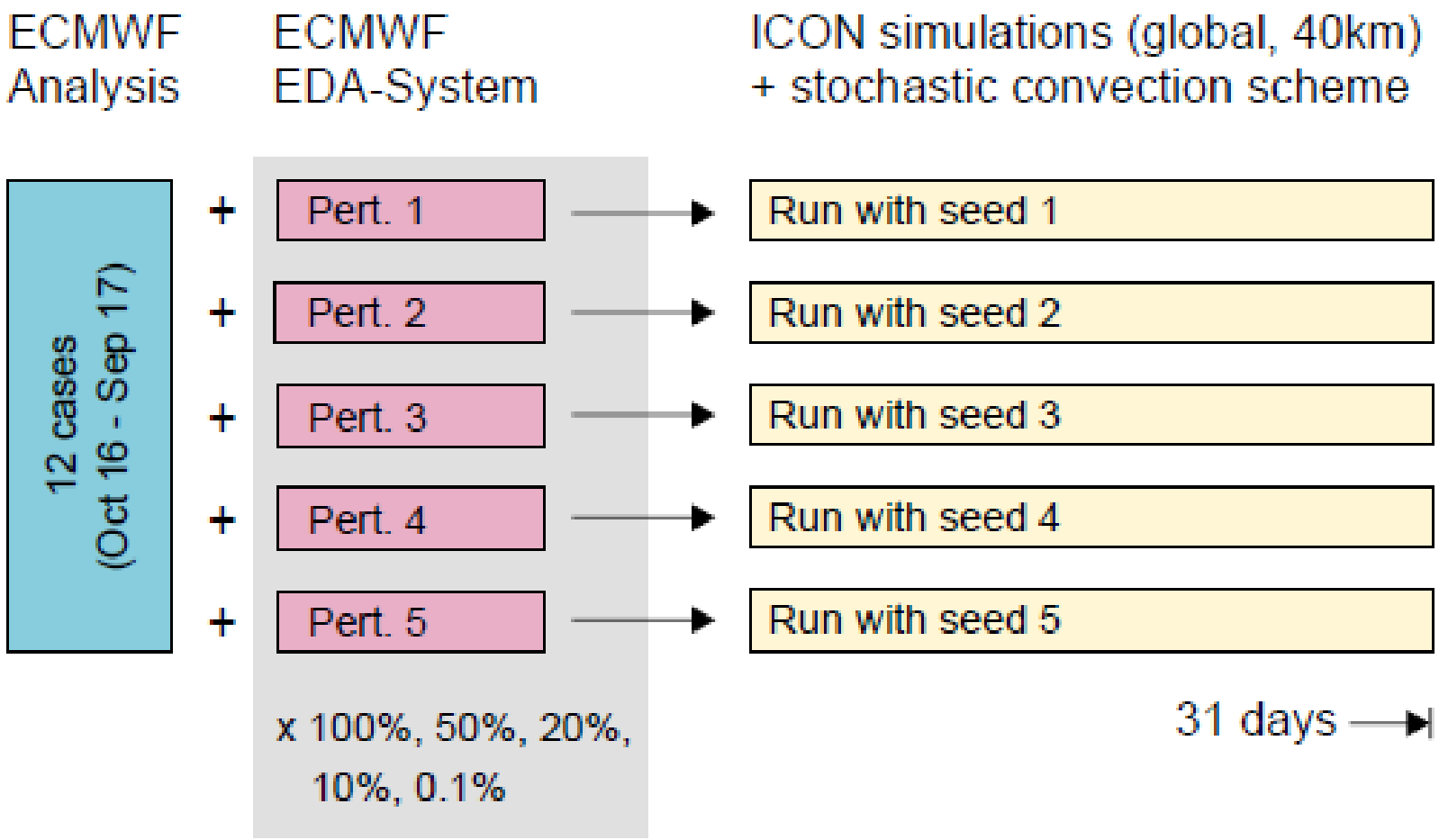
+ *(Baroclinic advection)*

+ *(Divergent advection)*

+ *(Non-conservative processes)*

+ *(Boundary and residual terms)*

Error growth experiments

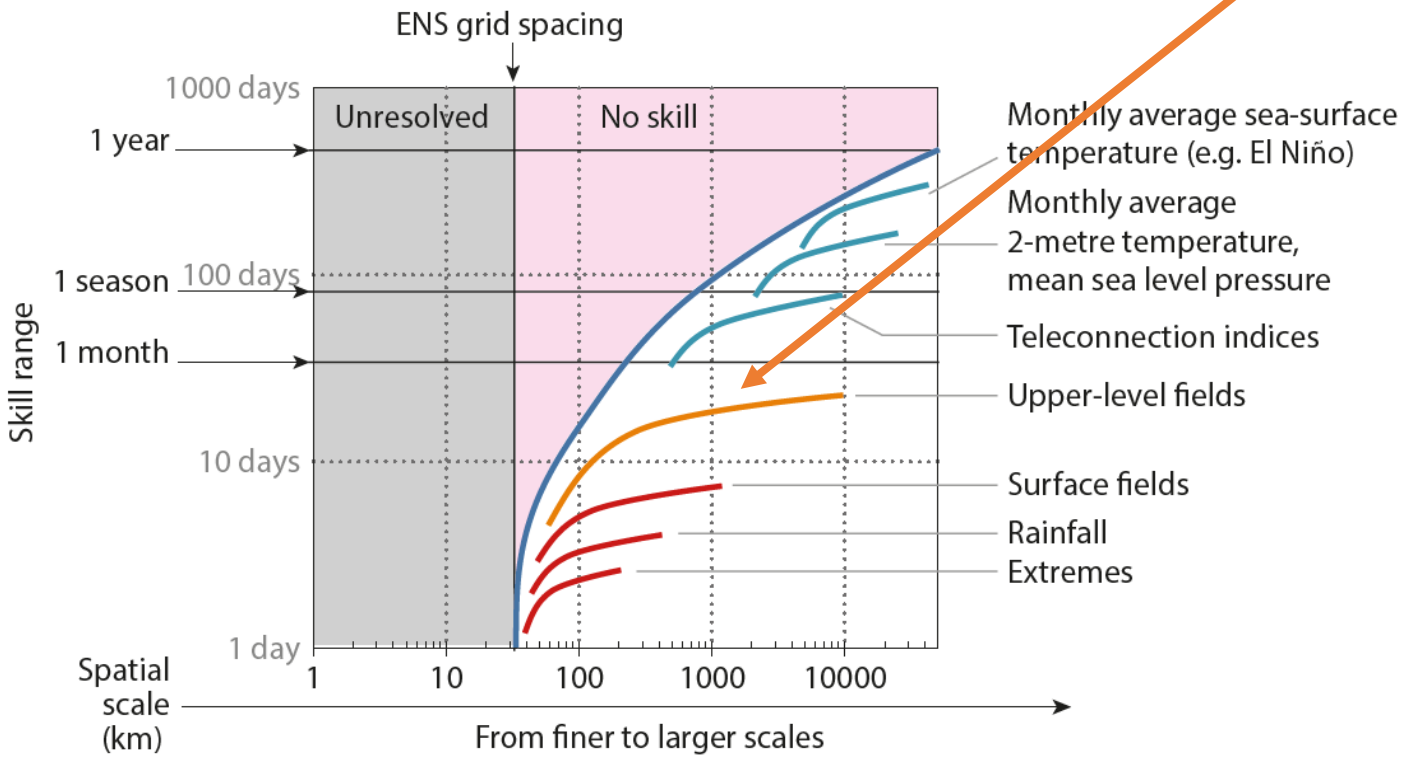


Practical predictability
Realistic initial errors

- Initial condition uncertainty from ECMWF EDA
- 5 experiments with re-scaled initial error amplitude
- 12 start times
- 5-member ensembles
- 40 km resolution
- stochastic convection

Predictability depends on what you predict

Limit of detectable skill for ECMWF operational forecasts (Buizza and Leutbecher 2015)



We have considered classical variables like pressure, temperature and wind

- Forecasts for long ranges and local extremes depend on many imperfect model components (Earth system modeling)
- Skill levels are low – need to take into account value for users to decide what is useful