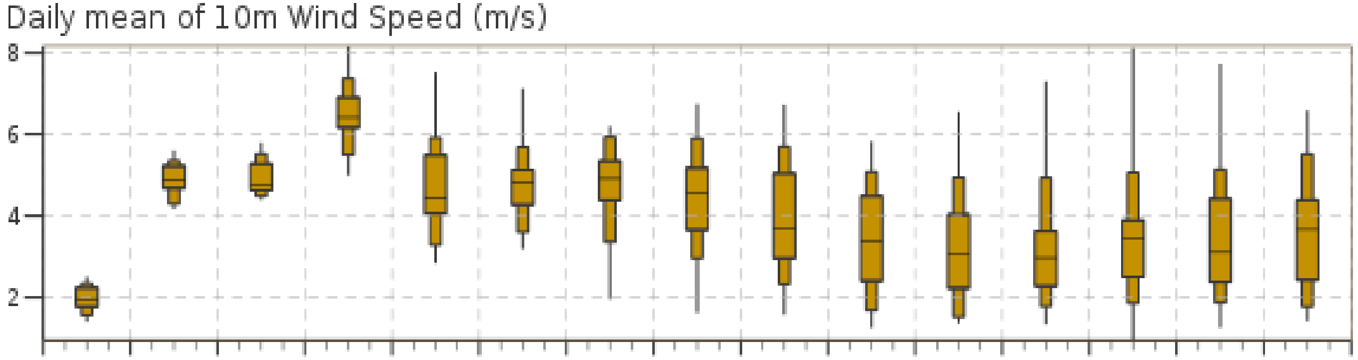
Introduction to intrinsically stochastic parameterization

George Craig
LMU Munich

Chaos, probability and model uncertainty

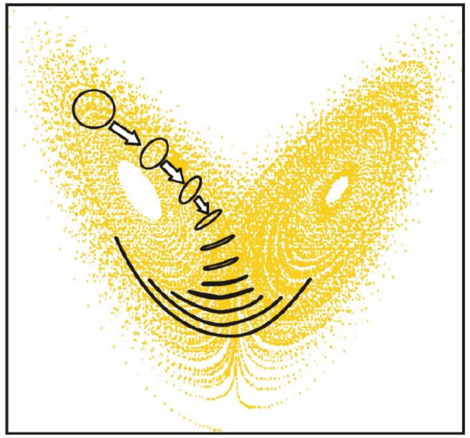


2. Forecasts need to show uncertainty

Initial conditions

Model

Post-processing

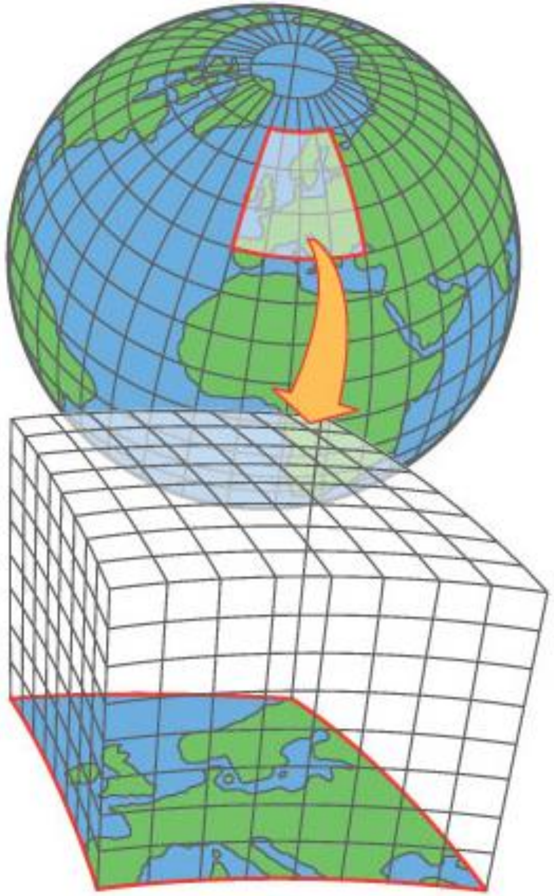


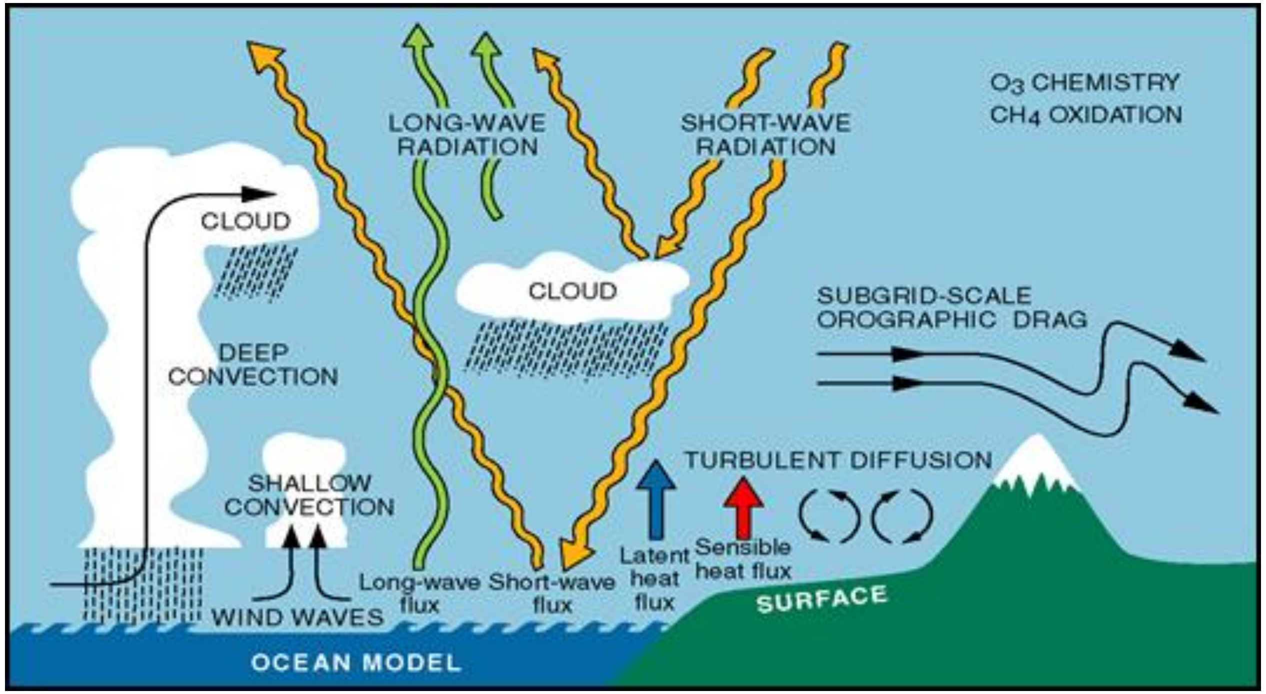
1. Predictability is intrinsically limited

3. Each part of the NWP chain contributes uncertainty

Model uncertainties

- Finite resolution – unresolved motions
- Physical processes – not part of fluid equations





Types of model error

Systematic errors

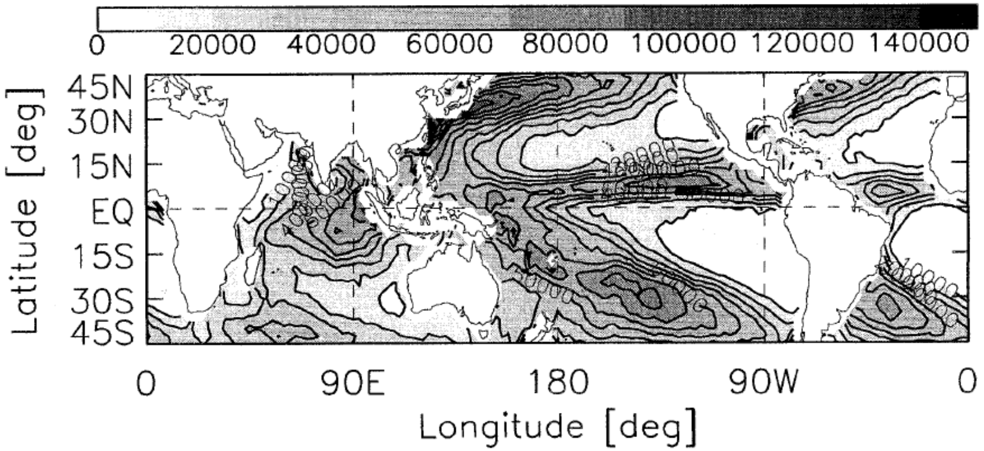
- e.g. dry bias in boundary layer leads to lack of convective initiation
- Need multiparameter or multimodel ensemble

Random variability

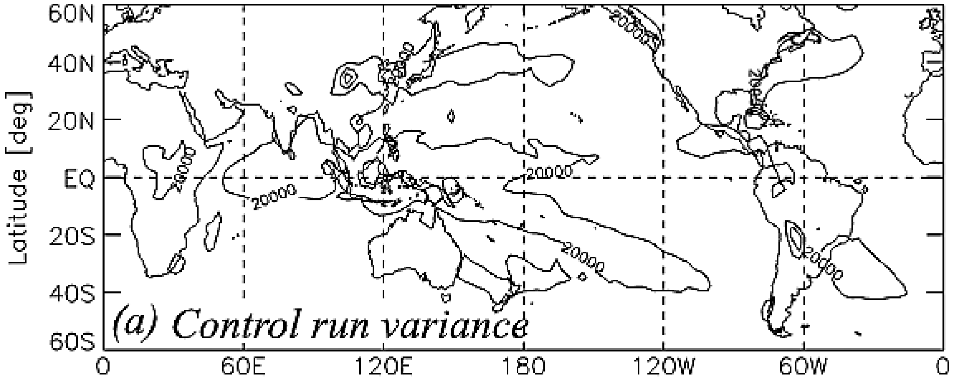
- e.g. unresolved motions trigger convection in wrong place
- Need stochastic variability among ensemble members

So add the missing variability, right?

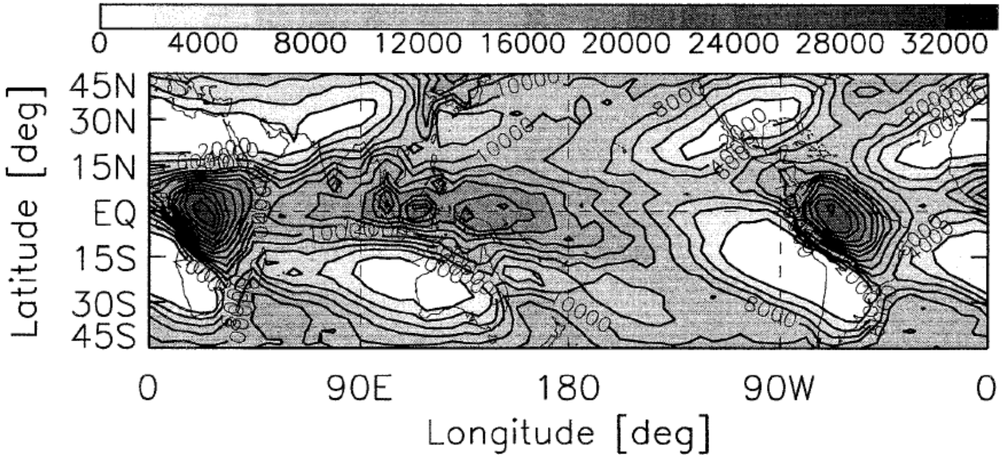
Observed precipitation variance



Default model has almost no variance



Model variance with stochastic forcing

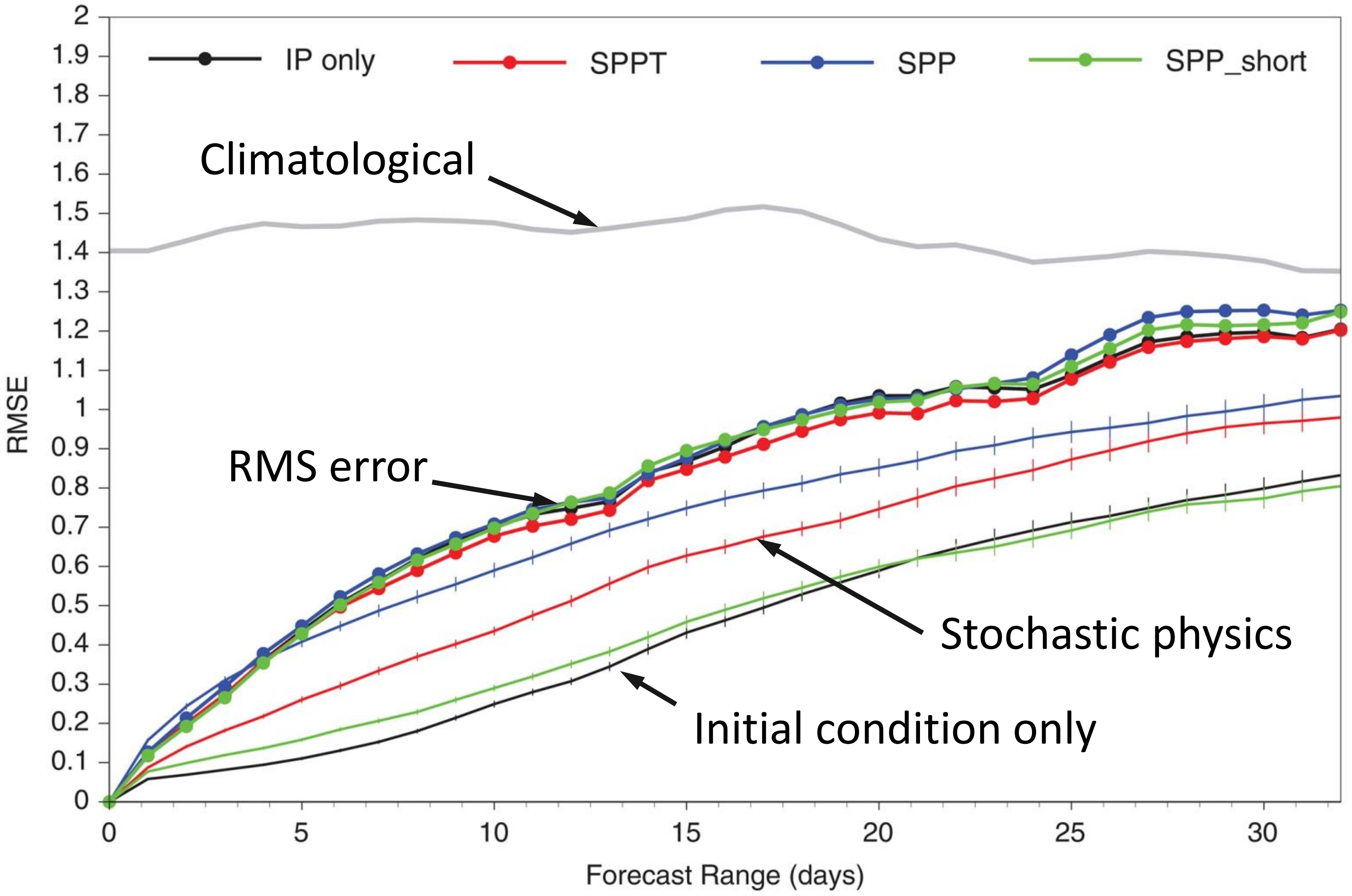


Lin and Neelin, *JAS* 2002

Response to forcing is nonlinear!

- Lin and Neelin (JAS 2002) added stochastic forcing to GCM based on observed precipitation variance
- Model variance increased in response to forcing but with different properties

Pragmatic solution – look at response



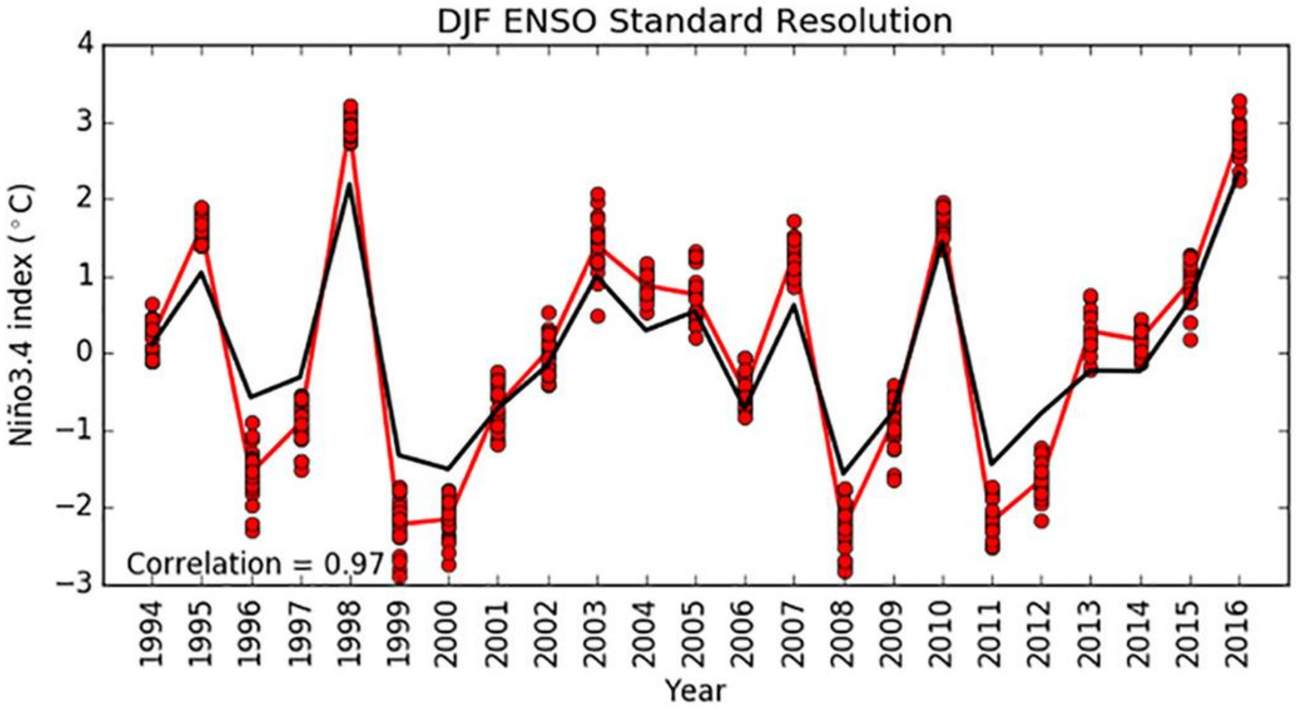
MJO index – Leutbecher et al., *QJRMS* 2017

- Ensemble is over-confident – spread does not capture uncertainty
- Calibrate ensemble by comparing spread and skill (eg. ECMWF well-calibrated)
- Reality should resemble random ensemble member

But:

- Calibration expensive
- Incomplete description of variability

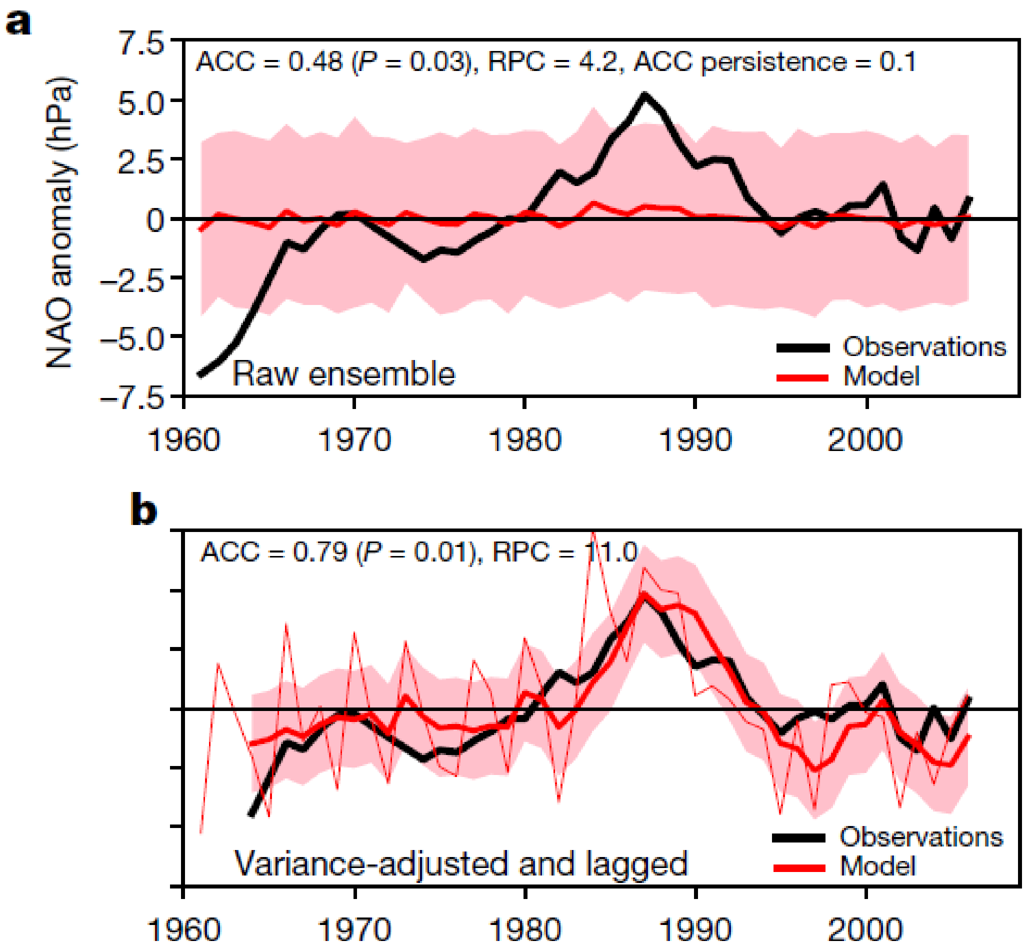
Variability on different scales I



- Hadley Centre ENSO forecast over-confident
- Problem is ENSO signal too strong, not lack of variability
- More stochastic perturbations could degrade forecast

Scaife et al., *ASL* 2019

Variability on different scales II

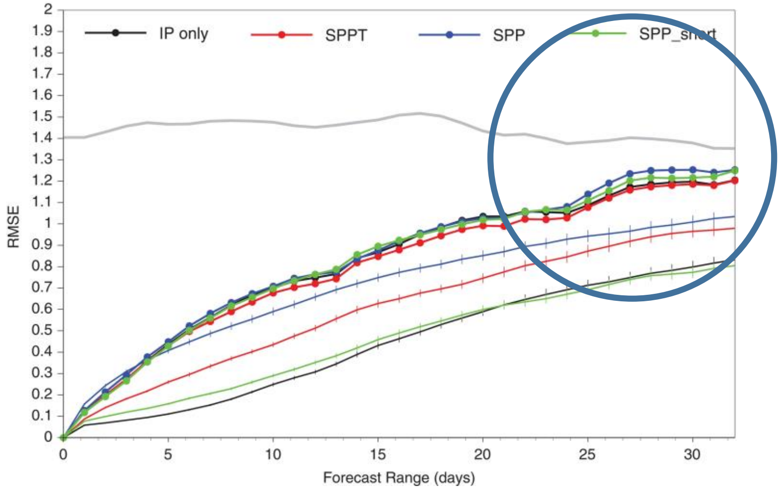


- Hadley Centre decadal NAO forecast has correct variance, but on short timescales
- Mean of large ensemble can average out variance and correlates with observations
- Reality very different from ensemble members (S/N ratio paradox)

Smith et al., *Nature* 2020

Variance on different scales

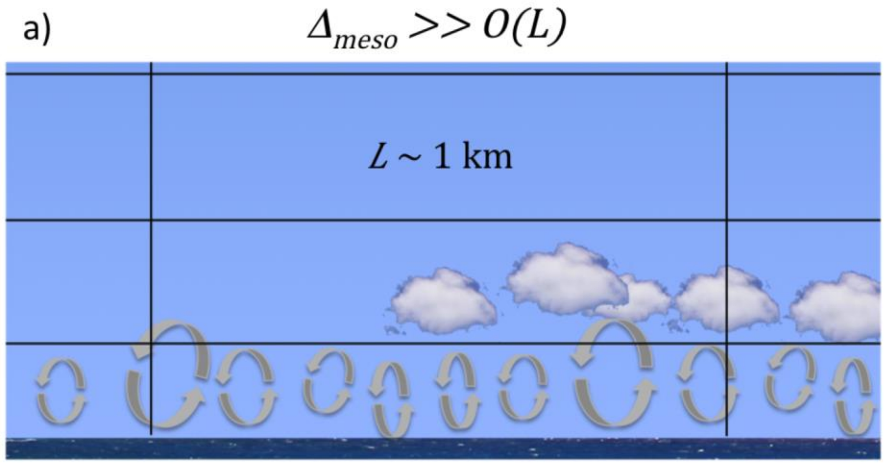
- Random model variability (stochastic physics) mainly “small-scale“
- Also have larger scale “climatological“ variability
- Not enough to consider spread-skill calibration
- *Analogy to convective vs synoptic scale?*



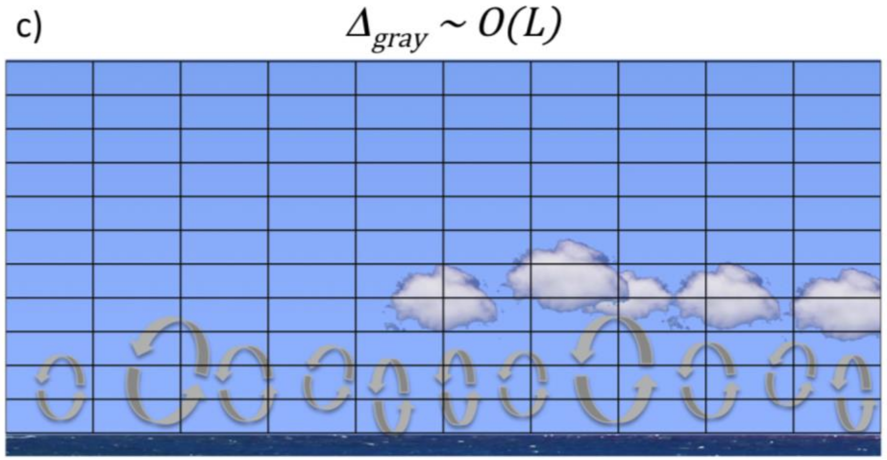
- Need to consider physical processes behind variability – intrinsically stochastic parameterization

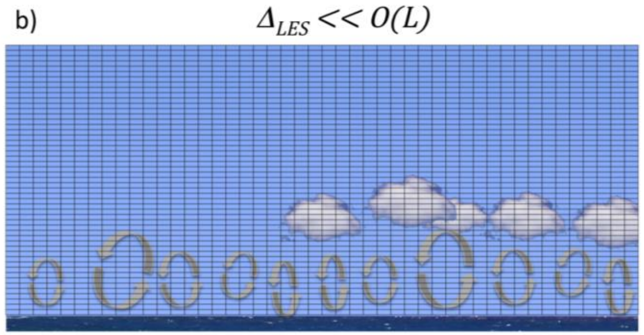
Parameterizations are intrinsically stochastic

- Parameterization gives net effect of many small elements (size *L*)
- Scale separation, grid length $\Delta >> L$, ensures well-defined result
- Gray zone, $\Delta \sim L$, range of possible results on grid scale – **stochastic**



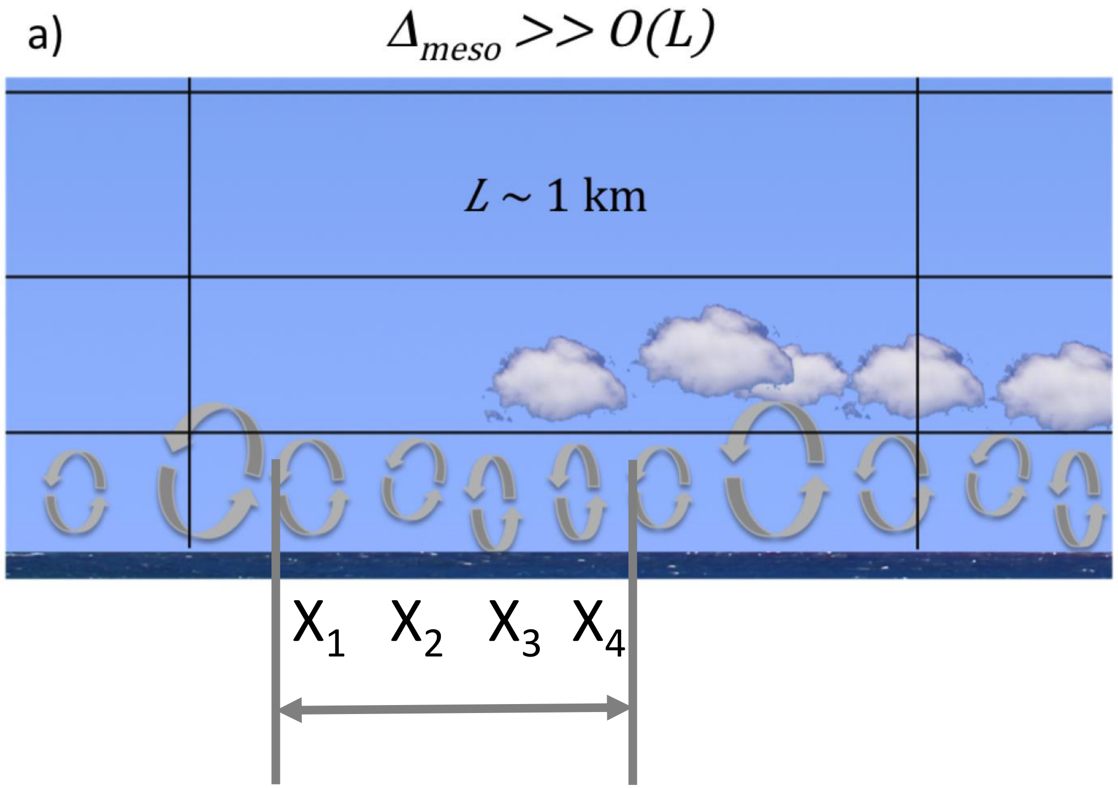
...or resolve elements $\Delta << L$, LES





Chow et al. *Atmosphere* 2019

Theory for fluctuations (Einstein 1905)



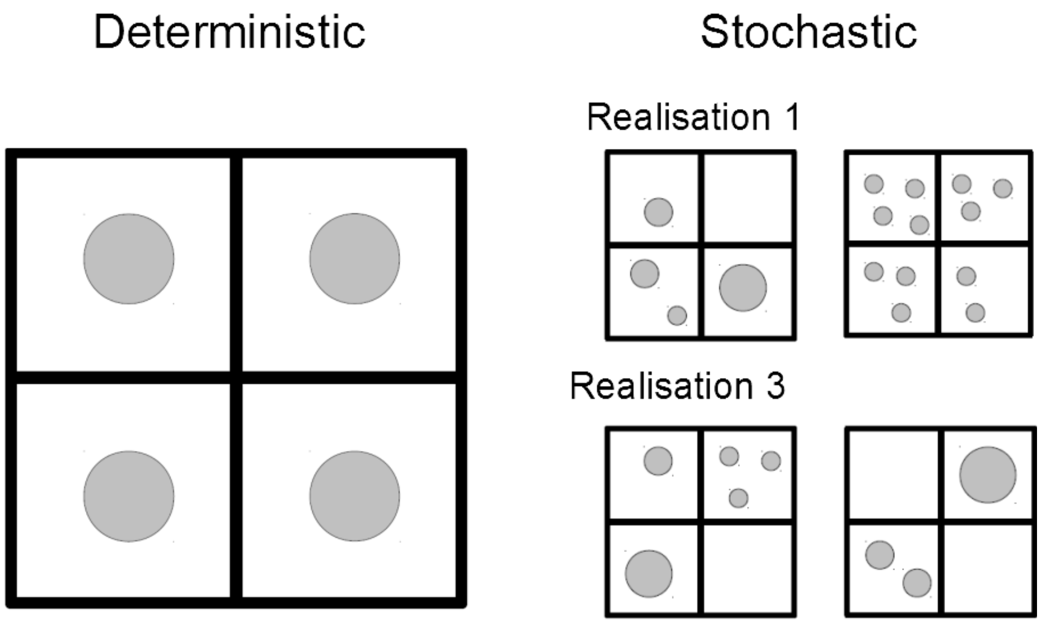
- Want net effect of (small) number *N* of subgrid features
- Sum of *N* random variables $X_1 + X_2 + \ldots$
- Central limit theorem says if *N* not too small, sum follows Gaussian distribution
- Standard deviation $\sigma = C/\mathrm{sqrt}(N)$
- *C* depends on environment (but not on *N*)

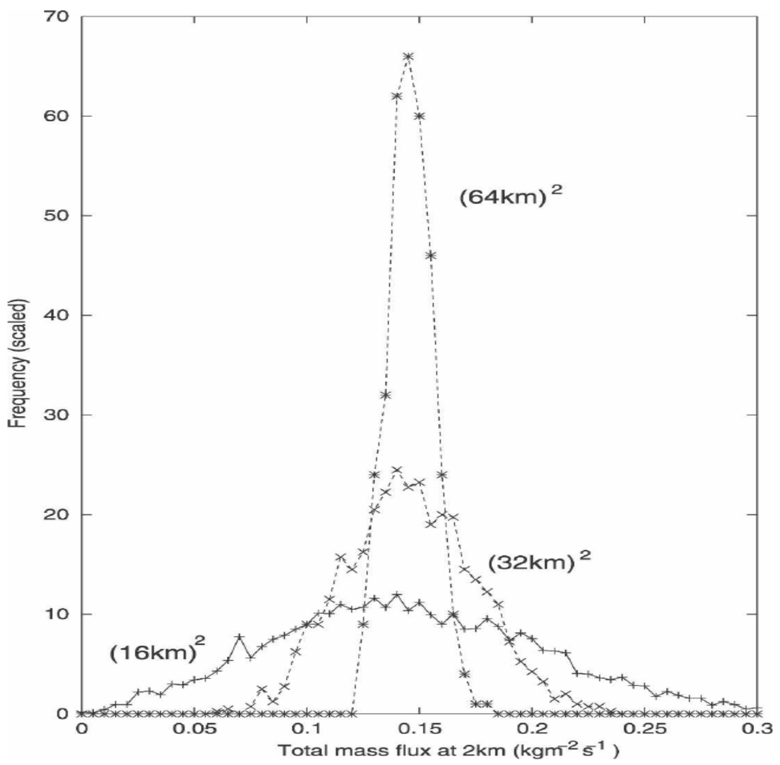
Stochastic convection (Plant and C. 2008)

- Average number of clouds: $N = M / m$

  = total mass flux (closure) / mass flux per cloud (constant)
- Analytic solution is Poisson-like process for M

  with $\sigma_M = 2M/sqrt(N)$





Plant and Craig *JAS* 2008

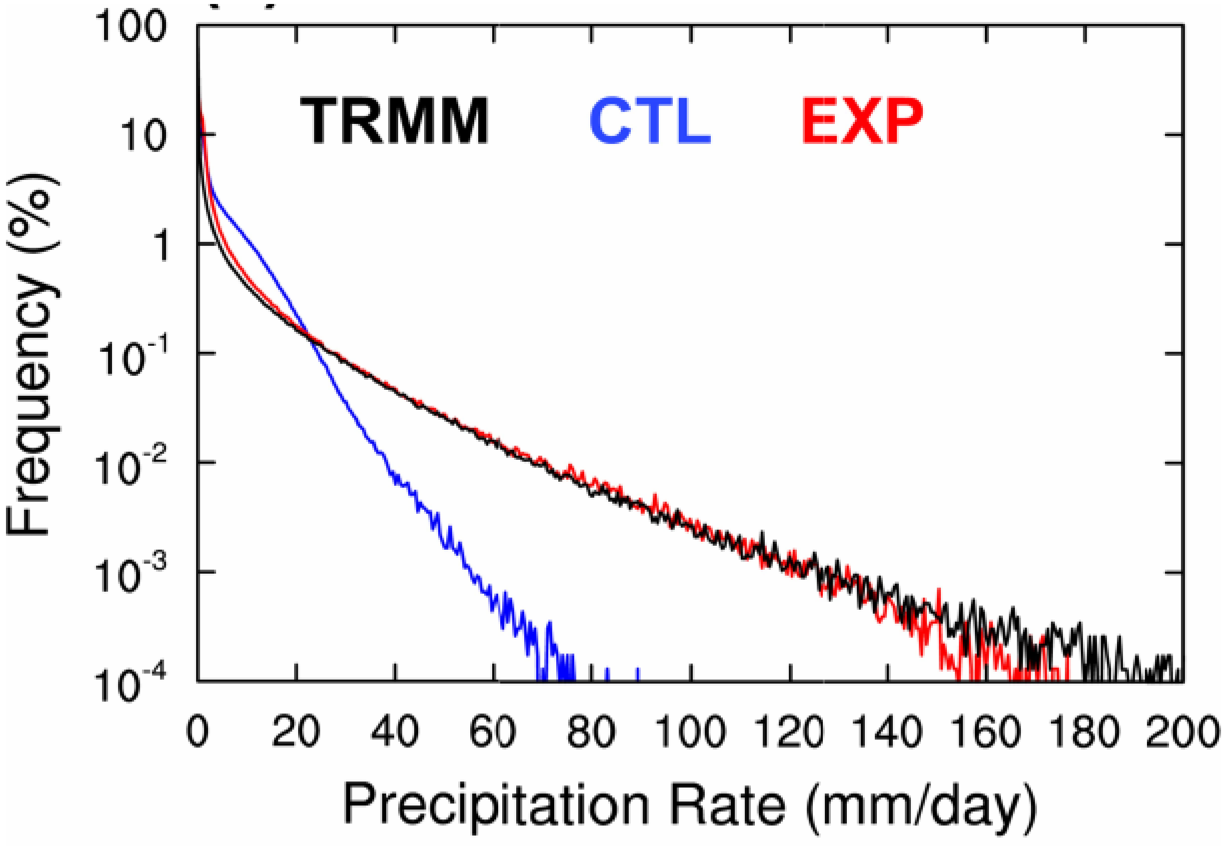
Impact on precipitation rates

- NCAR CAM5
- 10 year climate simulation
- Frequency of precipitation rates in tropics (20S – 20N)

Observations

Default Zhang-McFarlane convection scheme

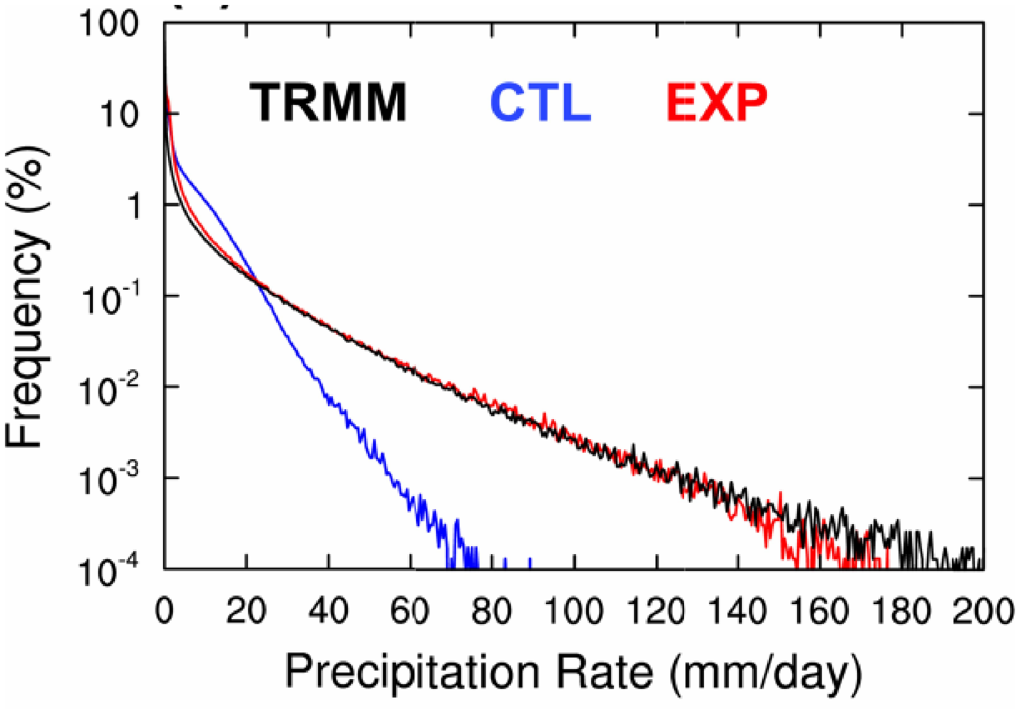
Stochastic ZMPC



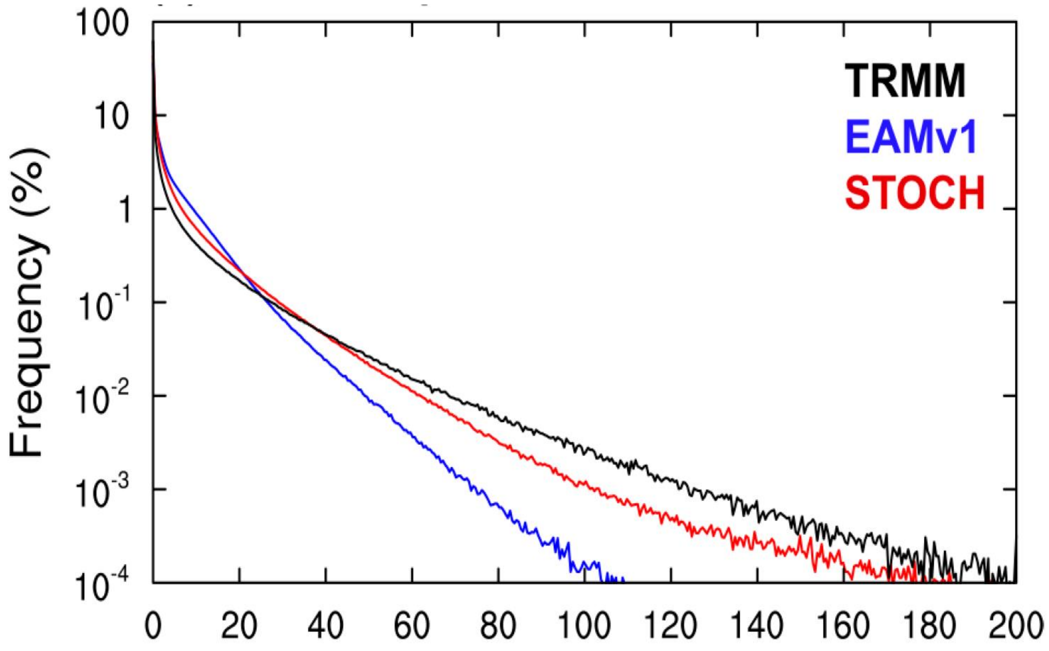
Wang et al. *GRL* 2016

Model dependence

NCAR CAM5



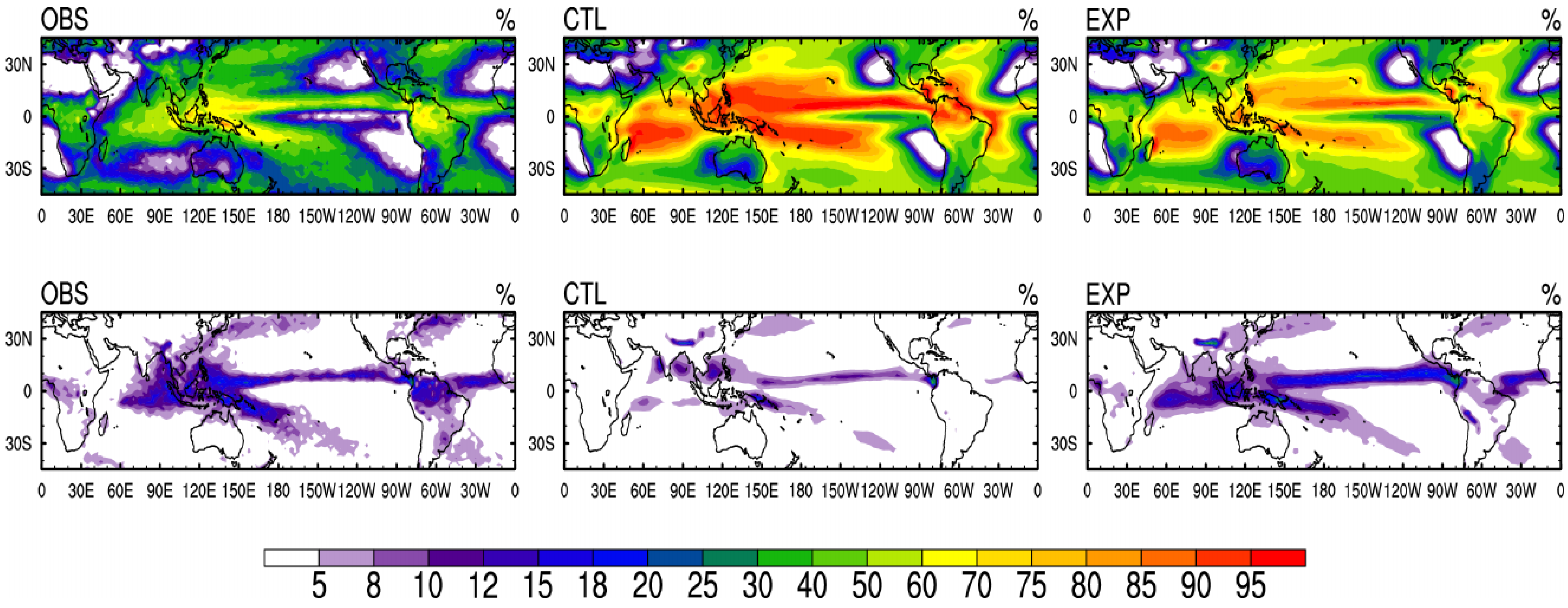
DOE E3SMv1



Sensitivity to vertical resolution

Wang et al. *GMD* 2021

Distribution of Light/Heavy precipitation

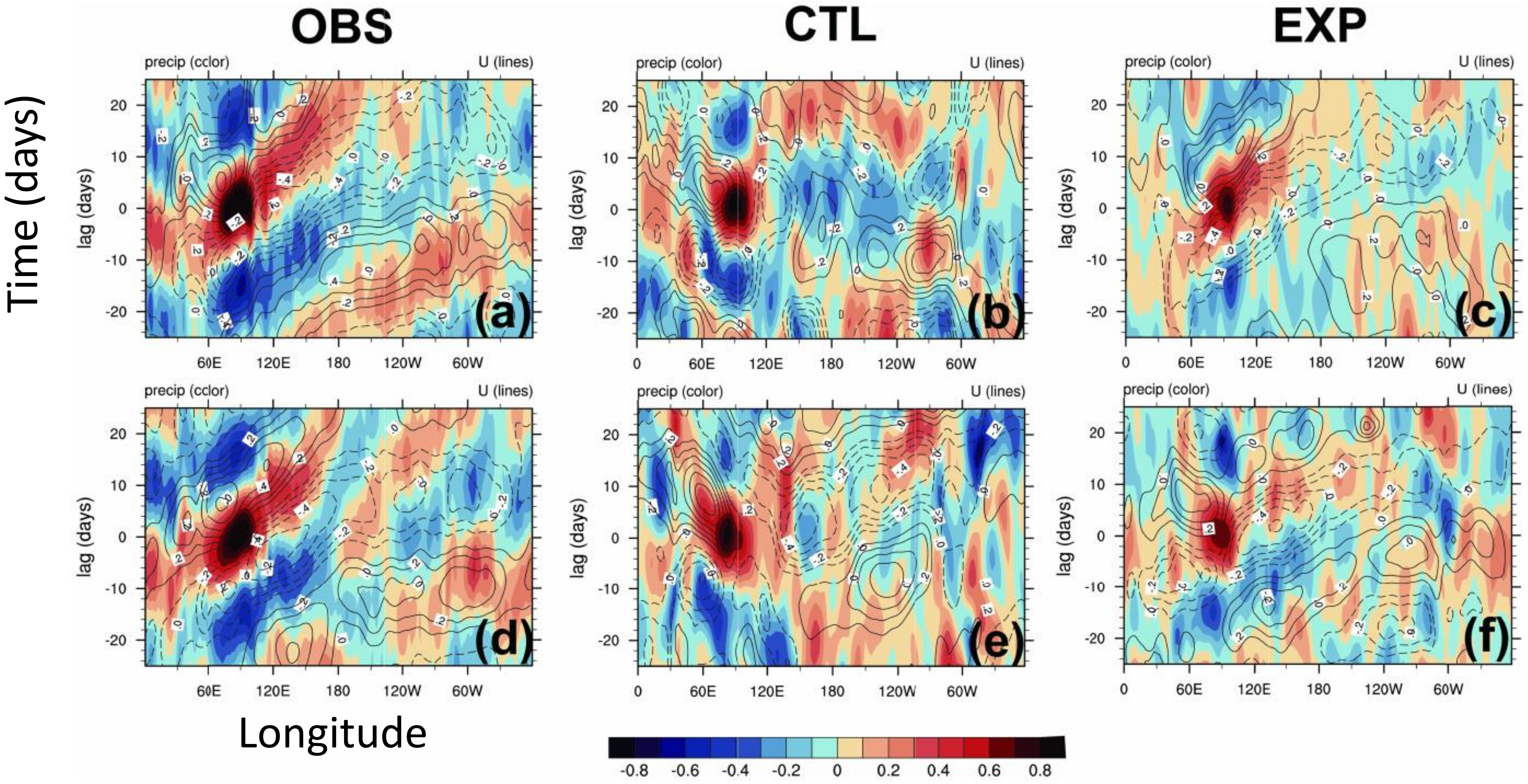


Frequency of precipitation > 1 mm/day (top row)

> 10 mm/day (bottom row)

Wang et al. *GRL* 2016

MJO variability



20-100 day bandpass filtered precipitation and U850 (20S-20N)

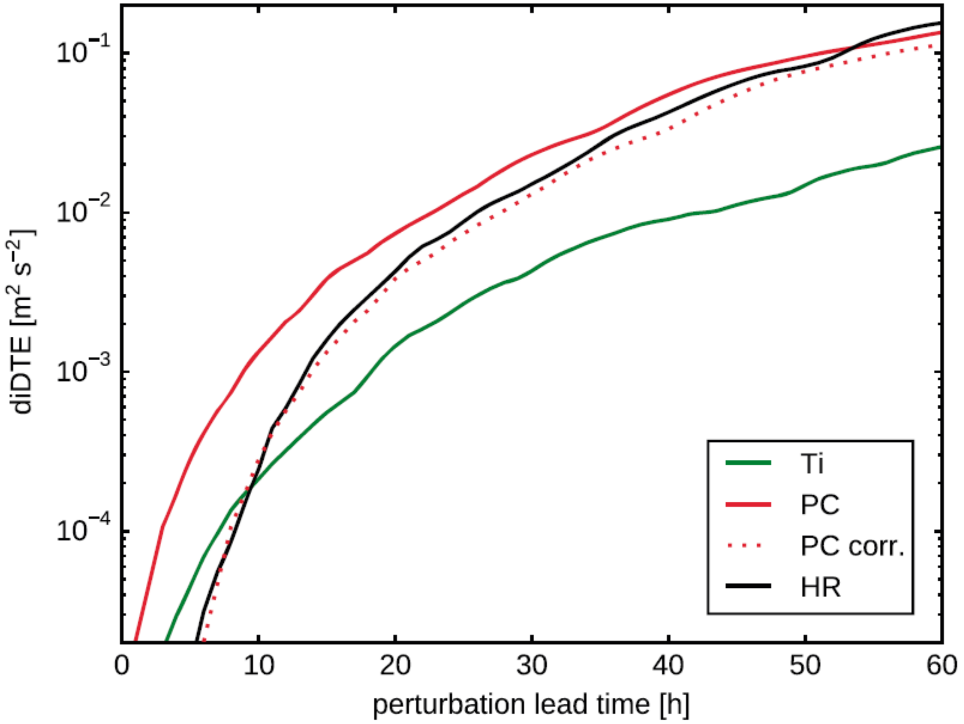
May-Oct (top, Nov-Apr (bottom)

Wang et al. *GRL* 2016

Perturbation growth experiments

Difference Total Energy (DTE):

$$DTE := \frac{1}{2}\left(\Delta u^2 + \Delta v^2 + \frac{c_p}{T_0}\Delta T^2\right)$$



- COSMO domain covering Europe and western Atlantic, 4 start times
- Perturbed by small-scale noise in initial conditions and stochastic convection (PC)

HR: 2.8 km resolution, explicit convection

Ti: 28 km resolution, default Tiedtke scheme

PC: 28 km resolution, PC stochastic scheme

PC corr: like PC but shifted 5 h (spin-up time for convective variability)

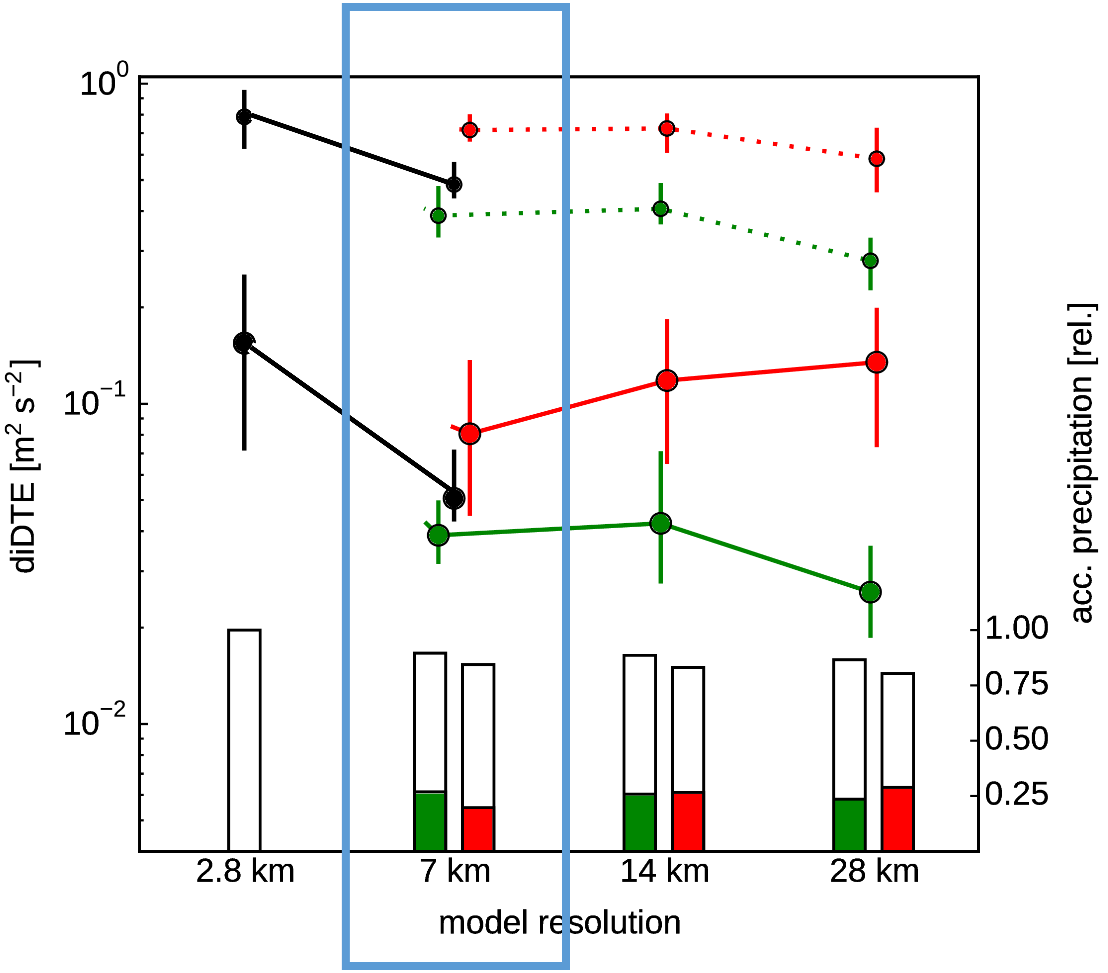
Selz and Craig *GRL* 2015

Scale invariance

- DTE on medium (dashed) and large (solid) scales after 60 h perturbation growth

- **No parameterization** – growth damped at low resolution
- **Default Tiedtke scheme** – too little growth
- **Plant-Craig stochastic** – realistic growth independent of resolution

Resolution 7 km – gray zone



Selz and Craig *GRL* 2015

Criteria for a stochastic parameterisation

1. Is the the scheme stable and well-behaved in the full model? (e.g. resolution dependence)
2. Is the variability contributed by the scheme significant? (compared to initial condition uncertainty, etc.)
3. Is the forecast skill superior to that obtained with a deterministic scheme? (on some score!)
4. Are there nontrivial interactions with the resolved flow? (Could the same skill be obtained by postprocessing output of model with deterministic scheme?)
5. Could the same skill be achieved with an inexpensive ad hoc scheme?

Conclusions

Summary

- Parameterizations well-posed if scale separation
- Otherwise intrinsically stochastic (Schär – multiple gray zones)
  - Cumulus convection (~10 km)
  - Convective boundary layer (~1 km)
- Generic theory for variability as sum of random processes
- e.g. Stochastic convection scheme
  - Improved precipitation variability
  - Improved upscale error growth

Issues

- Interaction with “climatological variance“ (convection like S2S?)
- Interaction with numerical artifacts “gray zone“