Historical Perspective: Earlier Ensembles, Forecasting Forecast Skill, and Global Leading Lyapunov Vectors

Eugenia Kalnay
University of Maryland

With many thanks to E. Lorenz, J. Charney, E. Epstein, C. Leith, R. Hoffman, W. Ebisuzaki, Z. Toth, R. Wobus, B. Hunt, M. Peña, S-C. Yang, A. Norwood, Cheng Da, T. Yoshida

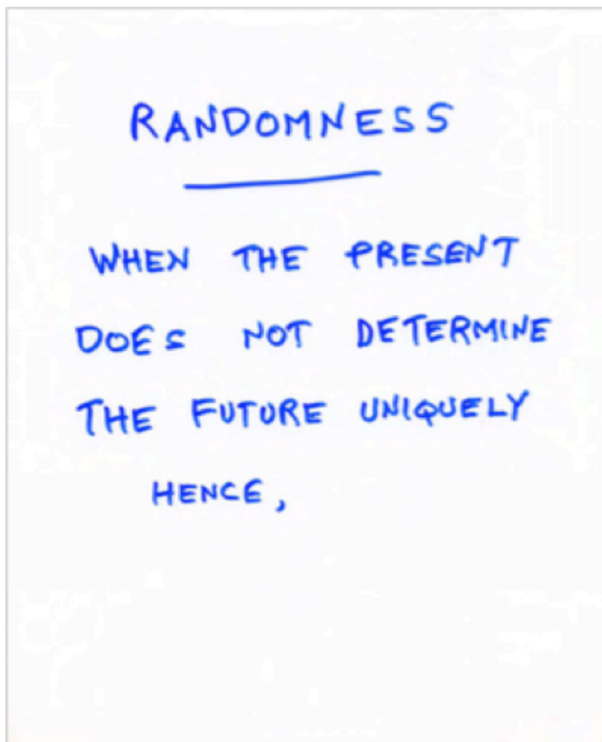
Ed Lorenz (1963, 1965) started it all, with the "limits of predictability"

- It gave impetus to the new science of chaos.
- The realization that the chaotic behavior of the atmosphere requires replacing "deterministic forecasts" with "ensemble forecasts" with perturbed initial conditions (and models).
- This led to the introduction of operational ensemble forecasting in December 1992 at both ECMWF and NCEP.

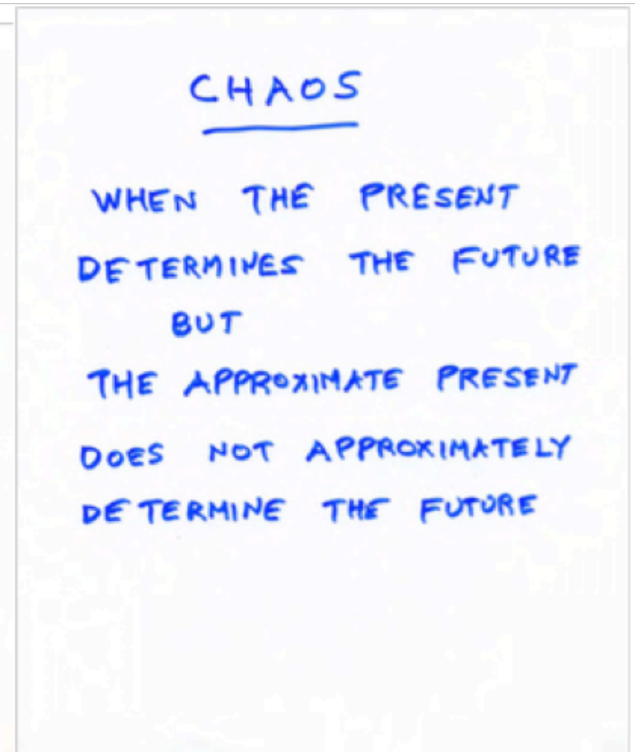
Transparencies from Lorenz in a 2006 workshop in Tallahasse, when he was 89. (courtesy of M. Zupanski)


CHAOS

BEHAVIOR THAT
IS NOT RANDOM
BUT LOOKS RANDOM


RANDOMNESS

WHEN THE PRESENT
DOES NOT DETERMINE
THE FUTURE UNIQUELY
HENCE,


CHAOS

WHEN THE PRESENT
DETERMINES THE FUTURE
BUT
THE APPROXIMATE PRESENT
DOES NOT APPROXIMATELY
DETERMINE THE FUTURE

Stochastic-dynamic forecasting

- Epstein (1969) introduced stochastic-dynamic forecasting, derived continuity equation for probability density φ : $\frac{\partial \varphi}{\partial t} + \nabla_D . (\dot{X}\varphi) = 0$
- "no ensemble member can be created or destroyed".
- Tested it with Lorenz (1963). It worked but…
- It was far too expensive: he developed a "true" probability density φ with 500 L63 members and then tested his approximation of φ

Monte Carlo Forecasting (Leith, 1974)

- Leith proposed ensemble forecasting with *m* members instead of the single (deterministic) forecast.
- Why? He predicted the anomaly wrt climatology

$\mathbf{u}_0$: true state of the atmospheric anomaly

$\hat{\mathbf{u}}$: forecast prediction of the atmospheric anomaly

$[(0-\mathbf{u}_0)(0-\mathbf{u}_0)^T]=[\mathbf{u}_0\mathbf{u}_0^T]=\mathbf{U}$ error of climatological forecast

$[(\hat{\mathbf{u}}-\mathbf{u}_0)(\hat{\mathbf{u}}-\mathbf{u}_0)^T]=[\hat{\mathbf{u}}\hat{\mathbf{u}}^T+\mathbf{u}_0\mathbf{u}_0^T-\hat{\mathbf{u}}\mathbf{u}_0^T-\mathbf{u}_0\hat{\mathbf{u}}^T]=2\mathbf{U}$ twice the error!

Monte Carlo forecasting (Leith, cont.)

- **A single long forecast asymptotes to twice the error of forecasting climatology.**

- Regressing towards climatology is expensive and complicated:
  $\mathbf{A} = [\hat{\mathbf{u}}^T \hat{\mathbf{u}}]^{-1} [\hat{\mathbf{u}}^T \mathbf{u}_0]$ regression matrix: huge!

- Leith proposed instead an ensemble of *m* forecasts:

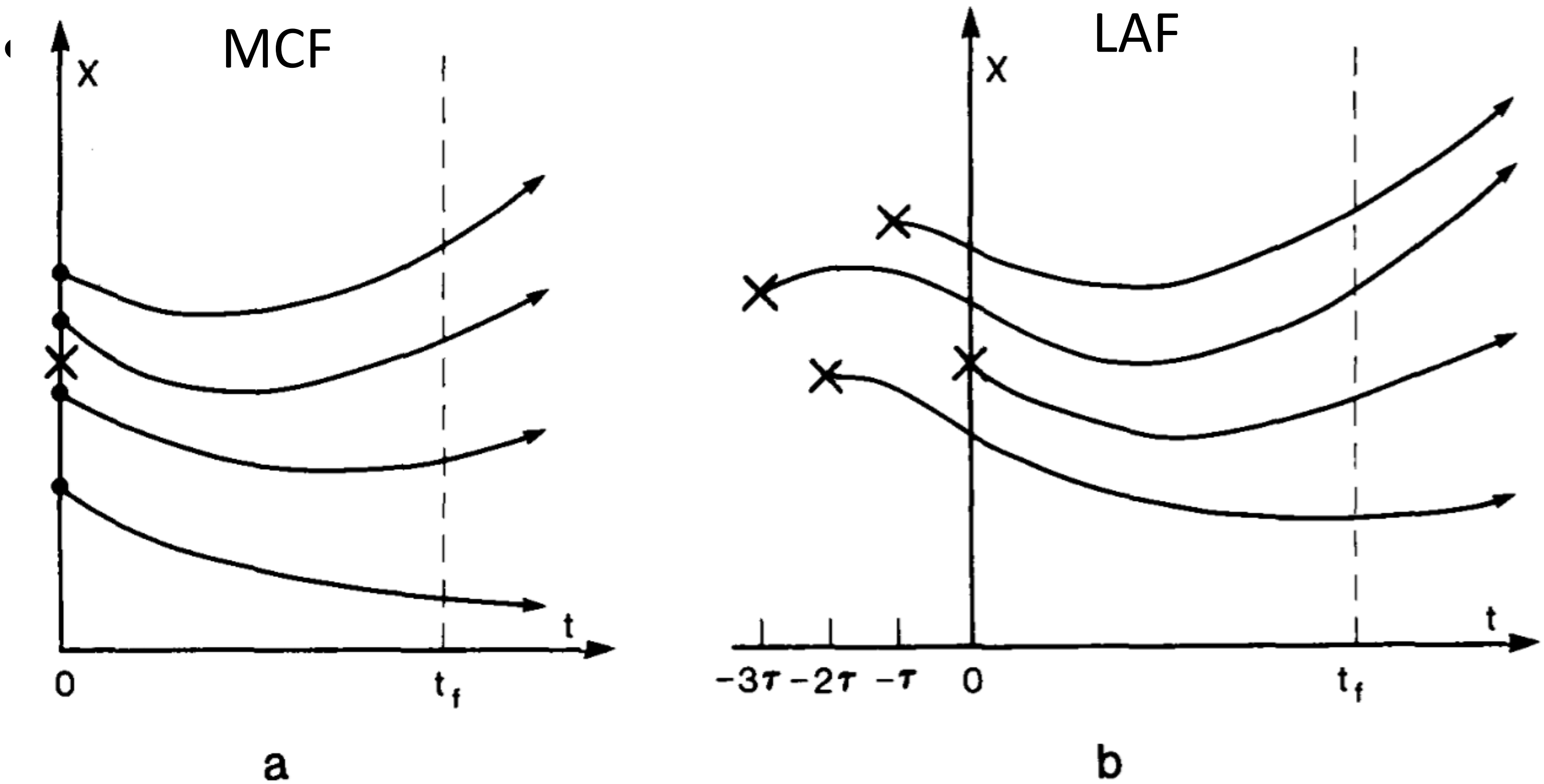
- The mean forecast error at long times goes to
  $$\lim_{t \to \infty} [(\bar{\mathbf{u}} - \mathbf{u}_o)(\bar{\mathbf{u}} - \mathbf{u}_o)^T] = \left(1 + \frac{1}{m}\right) \mathbf{U} \quad \text{without regression!}$$

- Leith suggested m=8 would be enough to make a good estimate of the mean, but the estimation of forecast errors would require many more members

- **Epstein used m=O(1000), Leith suggested m=8 may be enough!**

Lagged Average Forecasting

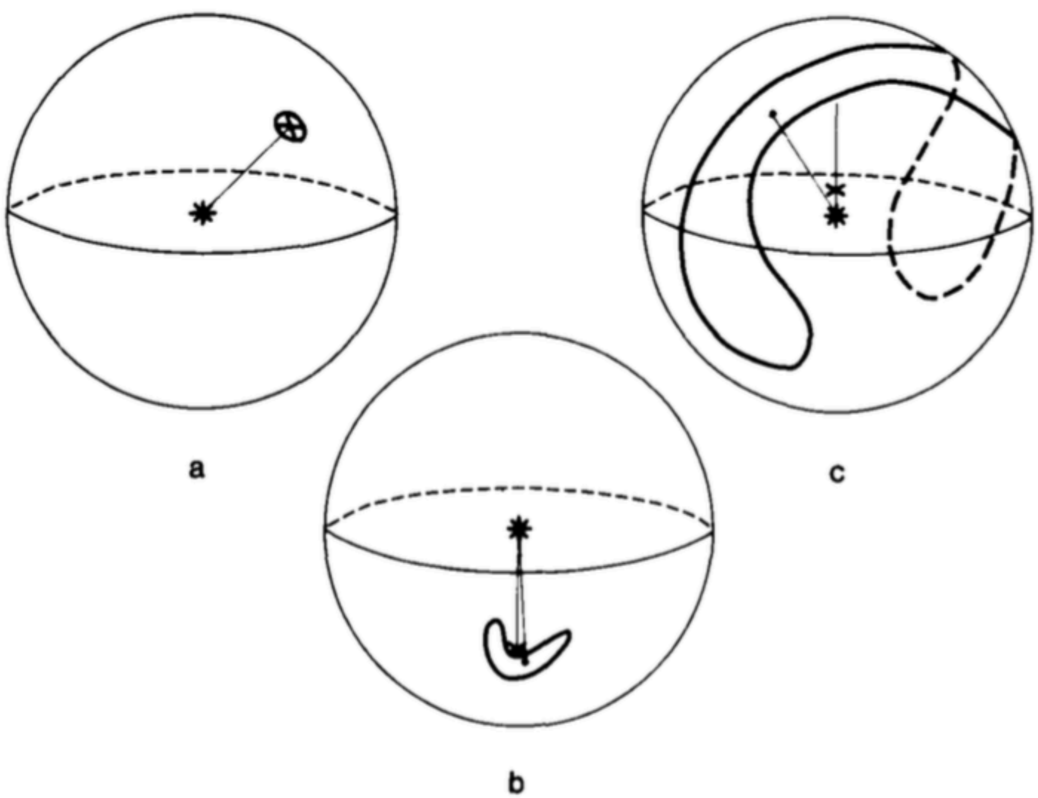
Hoffman and Kalnay, 1983



Use the forecasts already available in operations!

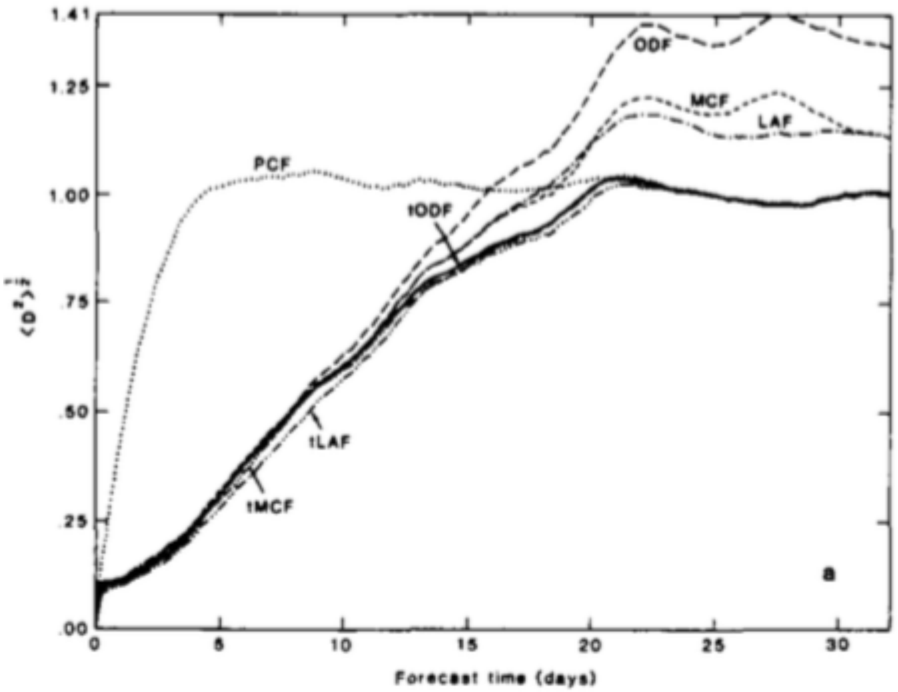
LAF testing

- Nature Model: 2-layer PE spectral model
- Forecast model: QG model (not identical twins)
- Compared ODF (single forecasts), MCF, LAF, N=8

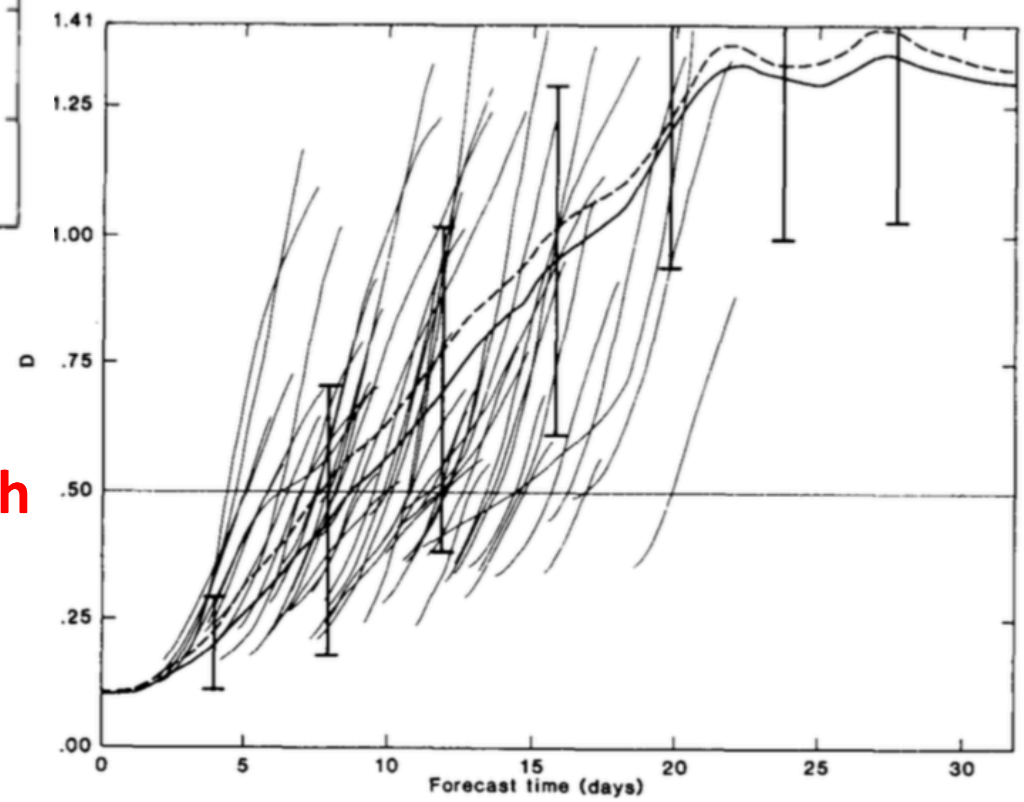


- Ensembles all hedge towards climatology
- ODF, MCF, LAF were all also “tempered” towards climatology

LAF - Results



LAF average errors only slightly better than MCF



There is a wide range in predictability (time at which error distance crosses 0.5)

LAF RESULTS (prediction of skill)

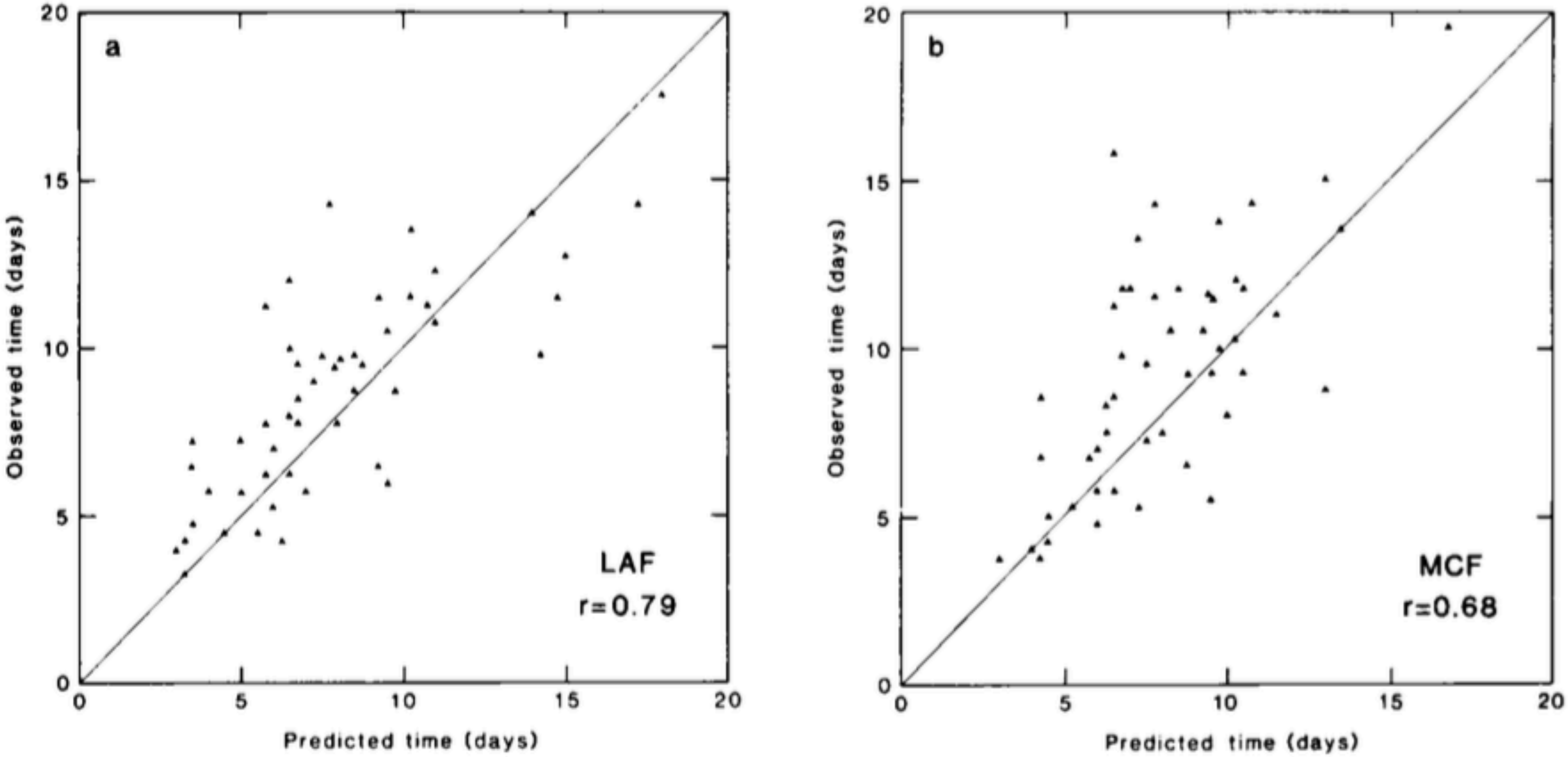


Fig. 8. Scatter plots of predicted versus observed time for $\tilde{D}$ to reach a value of 0.5 for the LAF (a) and the MCF (b).

LAF was significantly better than MCF in predicting the time for D=0.5, at the maximum growth

LAF RESULTS (prediction of skill)

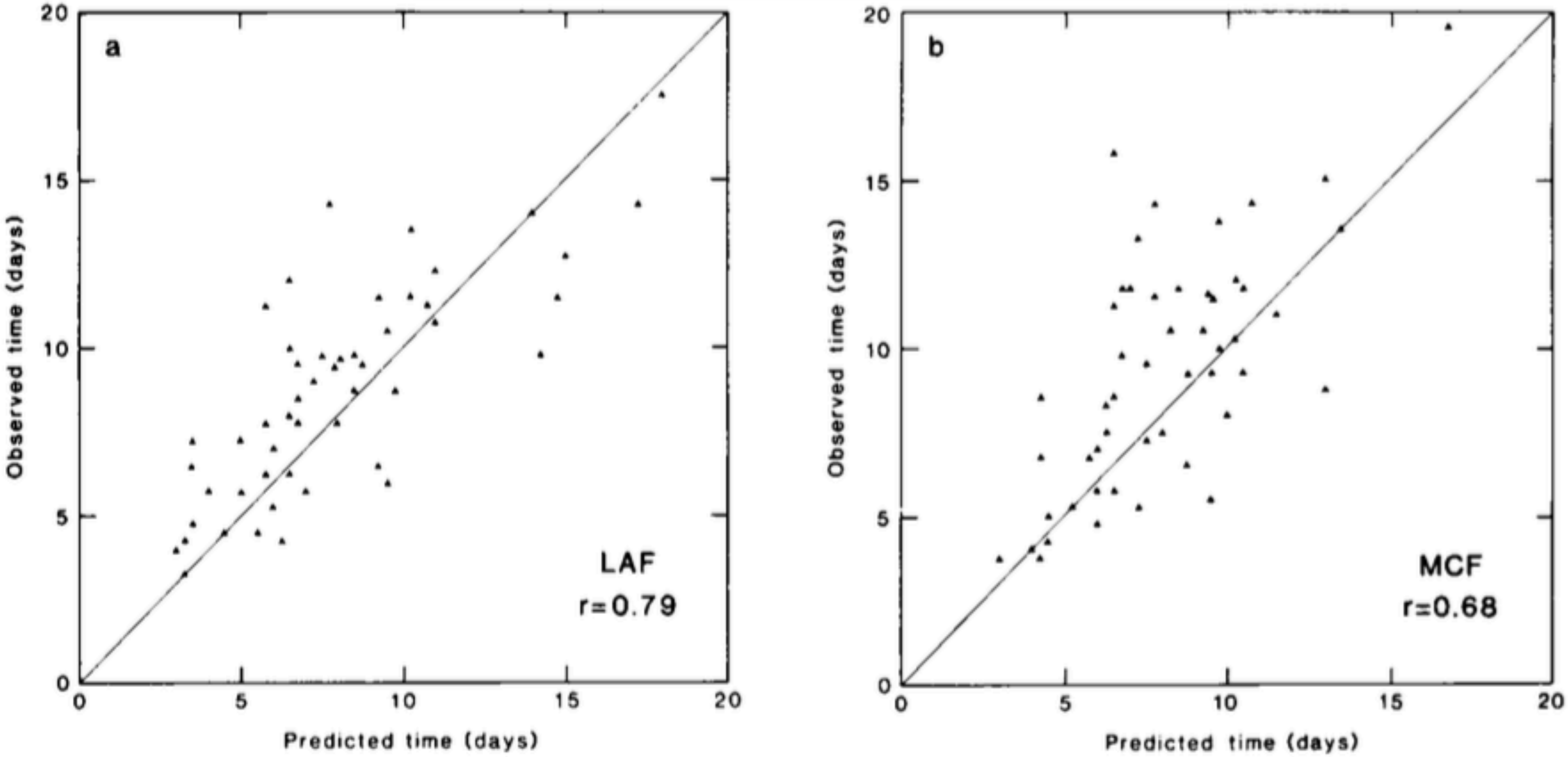


Fig. 8. Scatter plots of predicted versus observed time for $\tilde{D}$ to reach a value of 0.5 for the LAF (a) and the MCF (b).

LAF was significantly better than MCF in predicting the time for D=0.5, maximum growth

Why?

LAF RESULTS (prediction of skill)

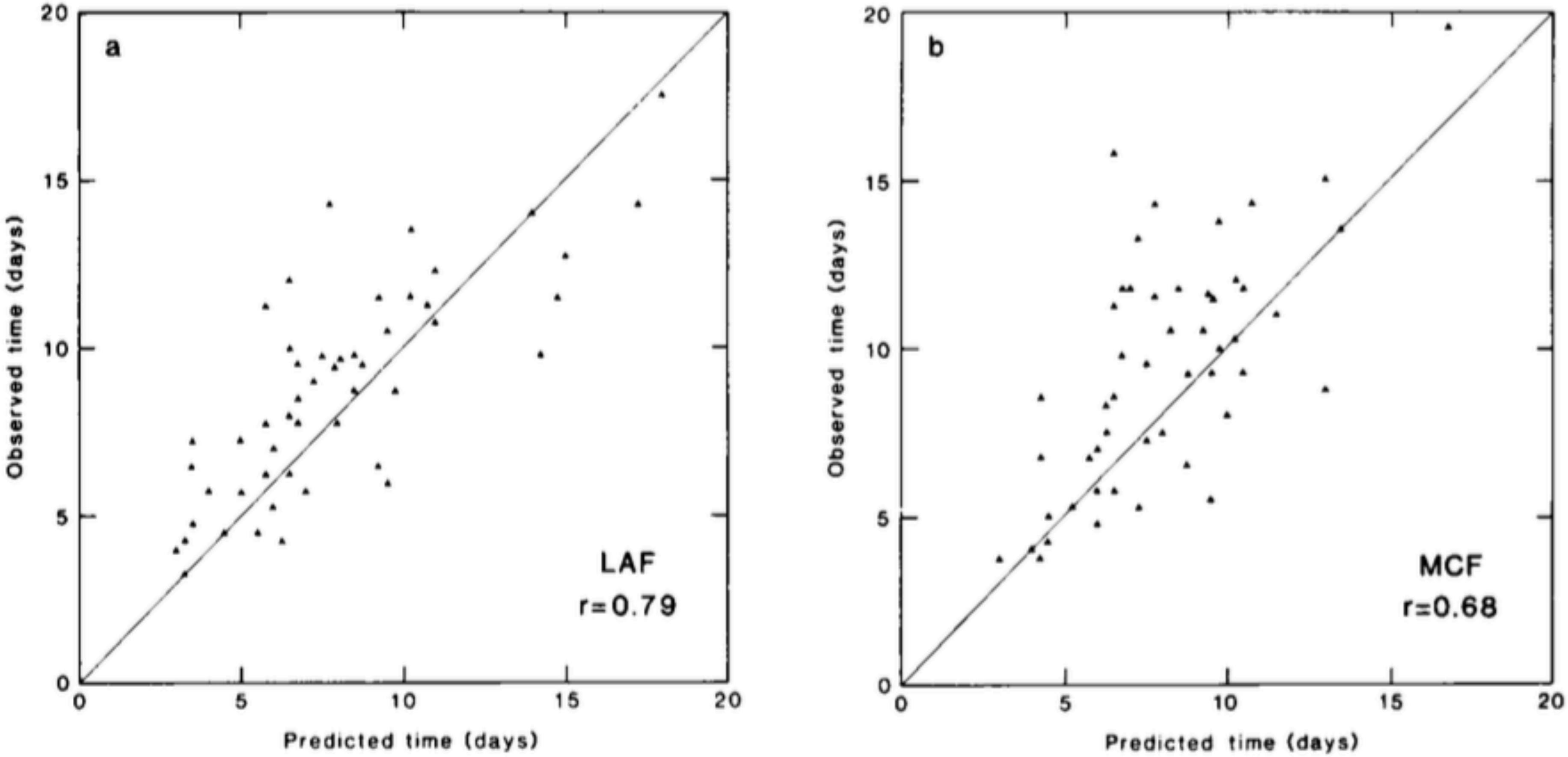


Fig. 8. Scatter plots of predicted versus observed time for $\tilde{D}$ to reach a value of 0.5 for the LAF (a) and the MCF (b).

LAF was significantly better than MCF in predicting the time for D=0.5, maximum growth

Why?

I think because LAF knows about the "errors of the day"!

SLAF, Ebisuzaki and Kalnay, 1991

- Disadvantages of LAF:
  - "Older" forecasts dominate the error average;
  - they should be given less weight.
  - LAF does not sample +/- errors
- **Scaled LAF (SLAF)**:
  - Divide the forecast errors by the difference between the analysis and the forecasts started at $t = -j\Delta t$, by j
  - This reduces by half the needed length of the forecasts, and reduces the amplitude of the "older errors" assuming linear growth.
  - Experimental results were clearly better than LAF. Easy to implement BCs on regional models (Hou et al., 2001)

“Good” and “Bad” Ensembles

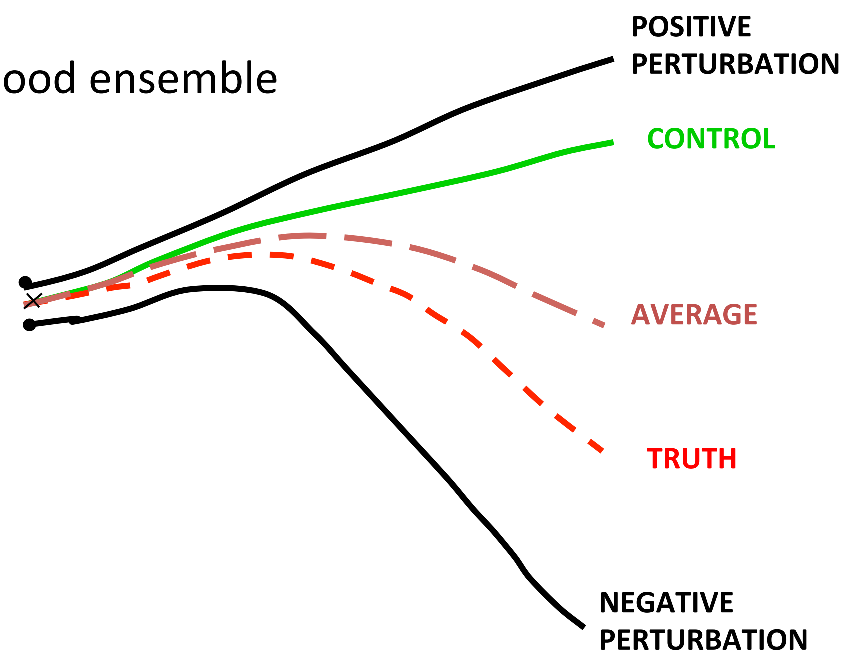
An ensemble forecast starts from initial perturbations to the analysis…

<u>In a good ensemble “truth” looks like an member of the ensemble</u> (Toth, 1992)

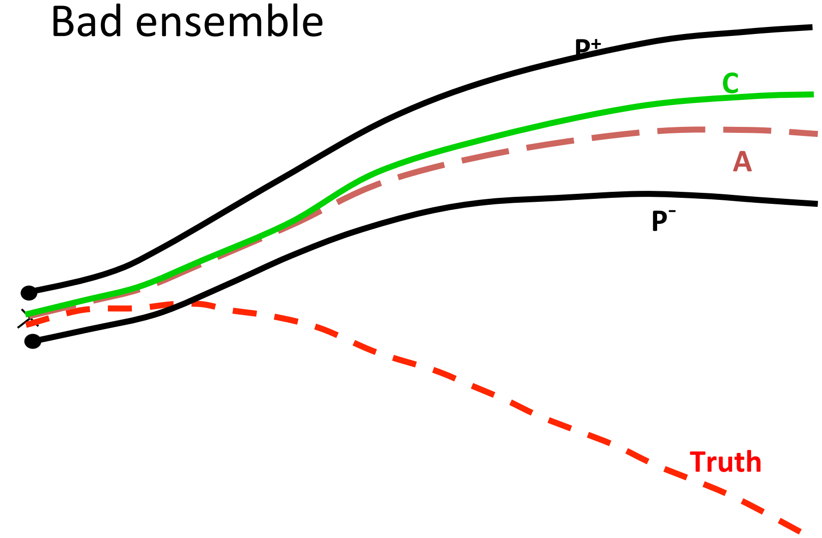
The initial perturbations should reflect the analysis “errors of the day”

A bad ensemble is still useful (implies there is a bug in the system)

Good ensemble



Bad ensemble



Forecasting Forecast Skill

Tennekes et al (1986): “No forecast is complete without a forecast of forecast skill”

- Gronaas (1983): Pilot study on prediction of medium range forecast quality. ECMWF Tech. Memo.
- Kalnay and Dalcher (1987): “Forecasting forecast skill”
- Wobus & Kalnay (1995): 3 years of operational fcsting skill
- Palmer & Tibaldi (1988)... On the Prediction of fcst skill
- Molteni & Palmer (1991): A realtime scheme pred fcst skill
- Arribas et al., (2004): A Poor woman EPS (PEPS competitive!)
- **Basic assumption: Ensemble agreement ⇔ forecast skill**

3 years of Operational Prediction of Fcst Skill

Wobus and Kalnay, 1995

- Barker (1991) had used a perfect model, 100 member ensemble experiments with discouraging results: the correlation between ensemble spread and fcst rms errors was 0.0 at t=0, and only 0.35 after one day 1, and was (.35-.55) in the first 10 days.
- At NCEP we started in 1988 to explore using other operational forecasts (ECMWF, UKMO, JMA) that we received routinely. The medium range run at NCEP started at 00Z, and at the other 3 centers at 12Z. So we forecasted from 12Z using 12hr older forecasts, plus an NCEP average fcst from today and yesterday at 00z, i.e., **the mean of four 5.5 days forecasts from 12Z**.

Method

- Multiple regression, 60 days training, computed daily for each 30x60 region and each forecast length, with 3 predictors and one predictand, the MRF at 12Z:

1) Fcst Agreement (AGR): Regional AC between MRF fcst, and each of the other 4 forecasts, averaged.
2) Fcst RMS Anomaly Amplitude (RMSA)
3) Fcst persistence (PERS): Regional AC between MRF and initial MRF analysis.

AGR is selected ~95%, RMSA ~70%, PERS ~45%

60 days of training; fcsts for 0.5 days to 5.5 days

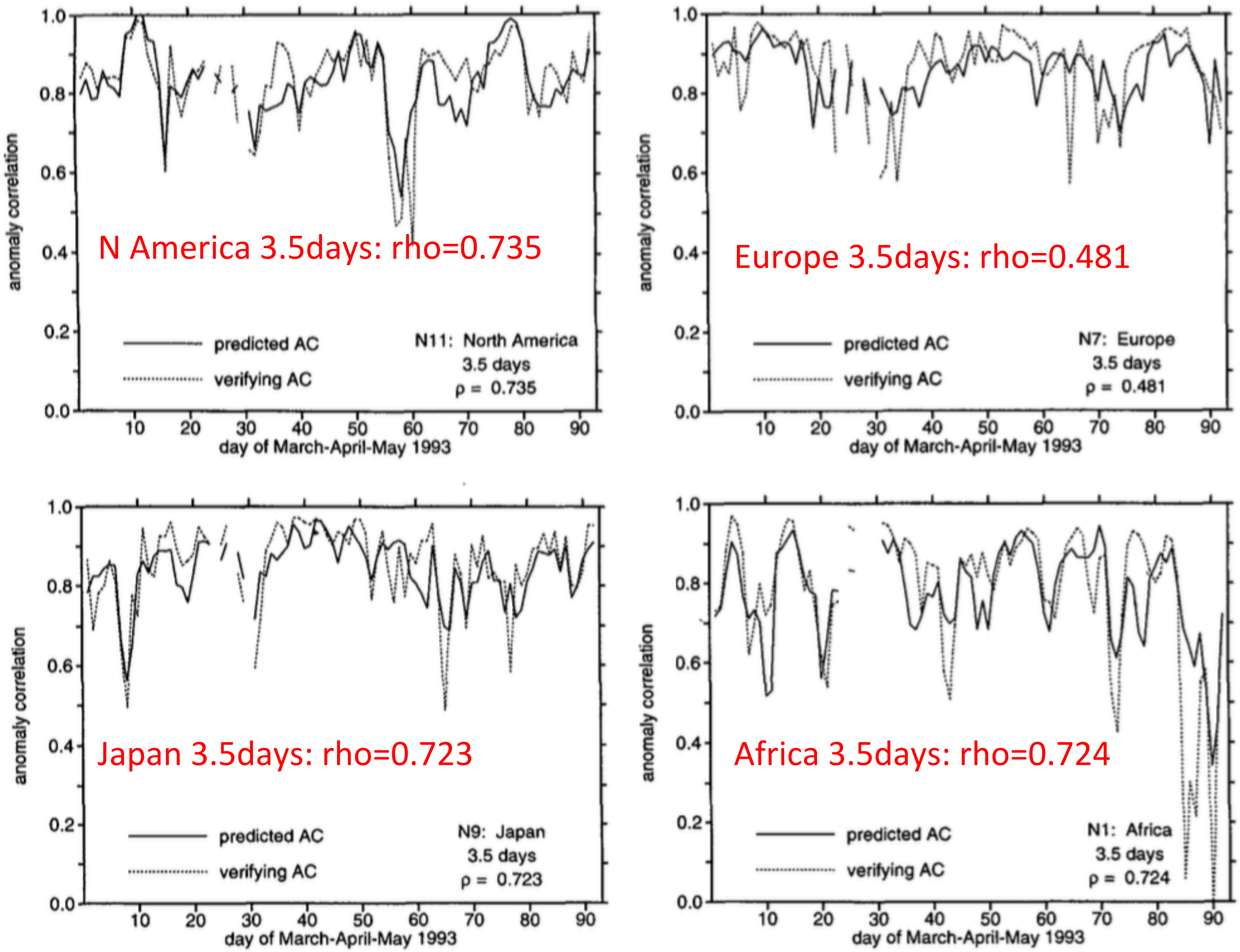
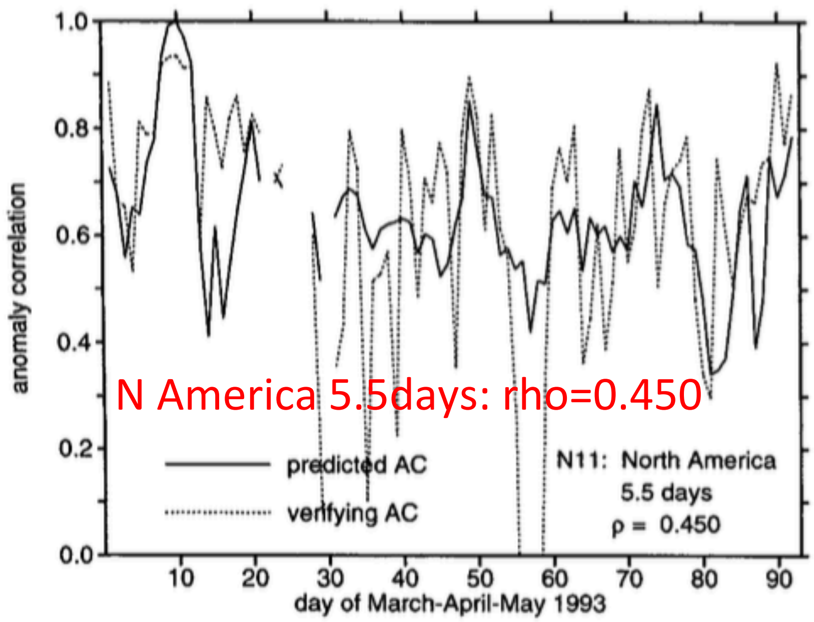
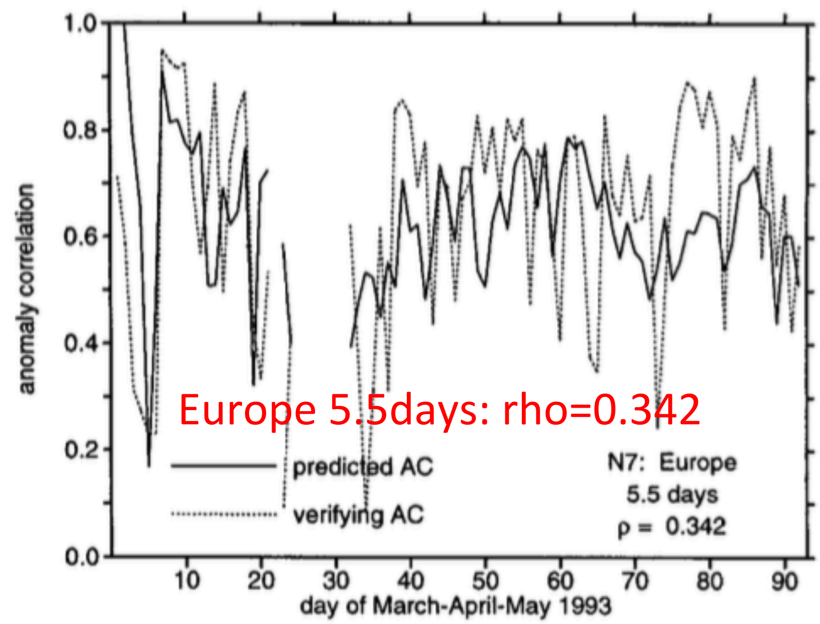


FIG. 4. Predicted (solid) and verifying (dashed) anomaly correlation for 3.5-day MRF forecasts for (a) region N11 (North America), (b) region N7 (Europe), (c) region N9 (Japan), and (d) region N1 (Africa) verifying during March, April, and May 1993.





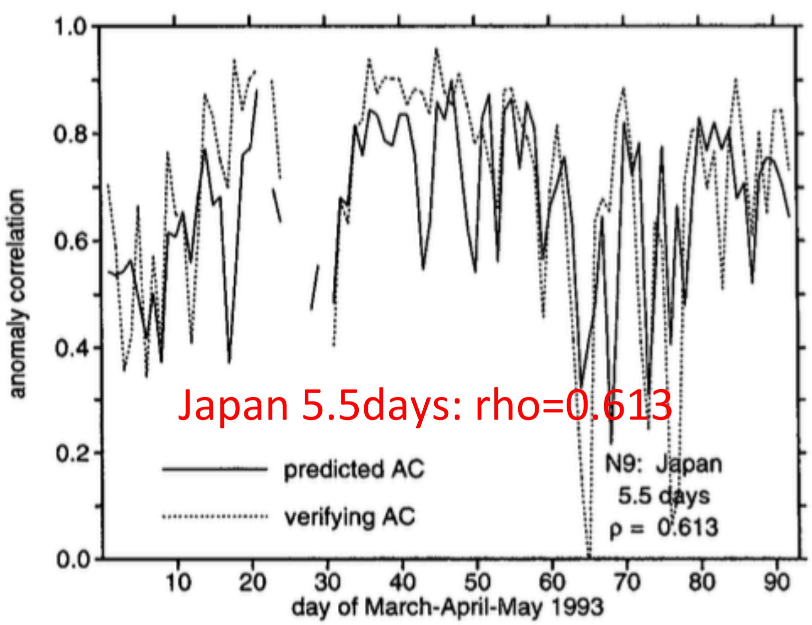


FIG. 6. Predicted (solid) and verifying (dashed) anomaly correlation for 5.5-day MRF forecasts for (a) region N11 (North America), (b) region N7 (Europe), (c) region N9 (Japan), verifying during March, April, and May 1993.

Good in the tropics and in the SH!

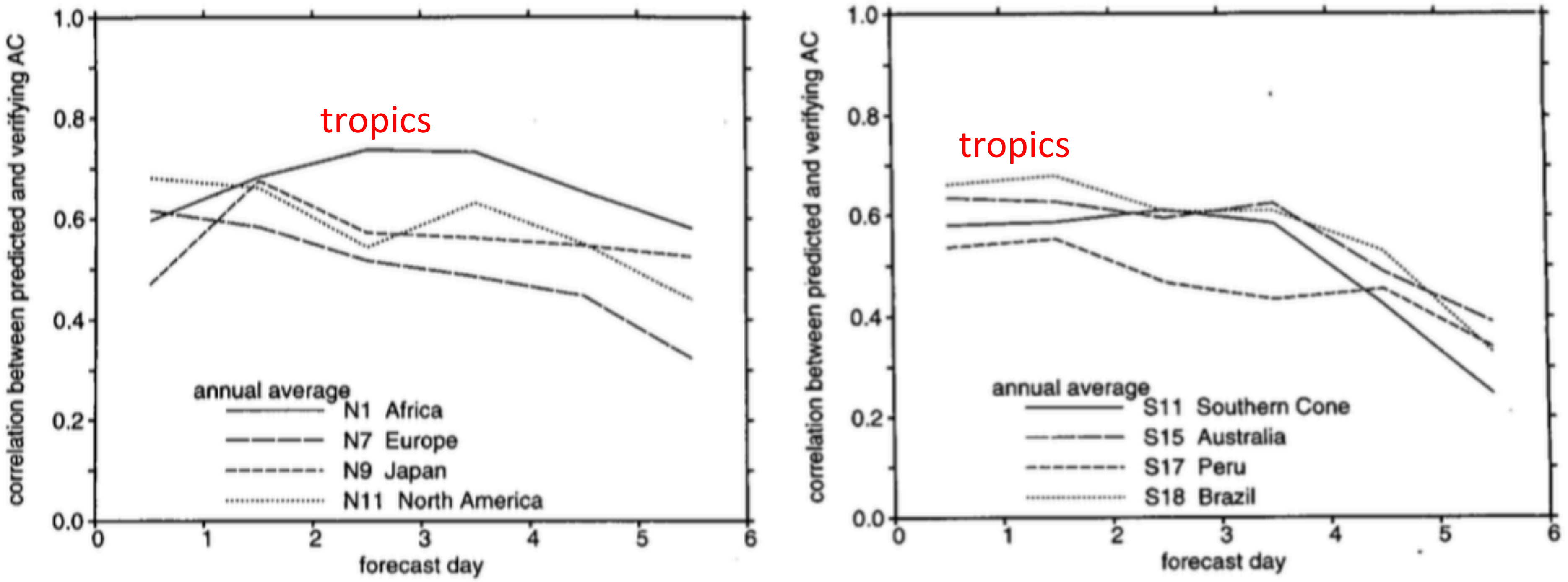


FIG. 12. Correlation between predicted and verifying AC of MRF forecasts for independent sample, annual average, by region for (a) Northern Hemisphere and (b) Southern Hemisphere.

How do these results compare with our current ensembles? Should we combine them?

Have you ever seen a Lyapunov Vector in an NWP model?

- I had never even heard of LVs until after Zoltan and I invented breeding.
- I thought BVs were similar to SV, until Zoltan made an error (factor of 4) in defining KE, and then we found that **BVs**, like **leading LVs**, were **independent of the norm**!
- Then we discovered with different space and time scales BVs could capture different types of instabilities. (Peña and Kalnay 2004).
- Talagrand told us that **there is a *single* global fastest growing LV (Ocedelec theorem)**. But I didn't believe it: it didn't seem to make sense.

Adrienne Norwood's thesis (1)

- She computed BVs, SVs and LVs for the Lorenz (1963) model and for the toy coupled ocean-atmosphere model (9 variables) and compared them, using Wolff and Samelson's method to obtain the LVs from the SVs.
- One discovery was that BV did not stick to LV1! Whenever LV2 grew faster than LV1, the BVs joined LV2, and only returned, when LV1 grew faster.
- Similar to BV capturing fastest growing instabilities in different parts of the world.
- She then computed the BVs for a QG model. It was easy to get the leading LV from breeding because the QG model has a single type of instabilities.
- But with her computer resources she couldn't get the SVs to converge, and thus get a complete set of LVs.

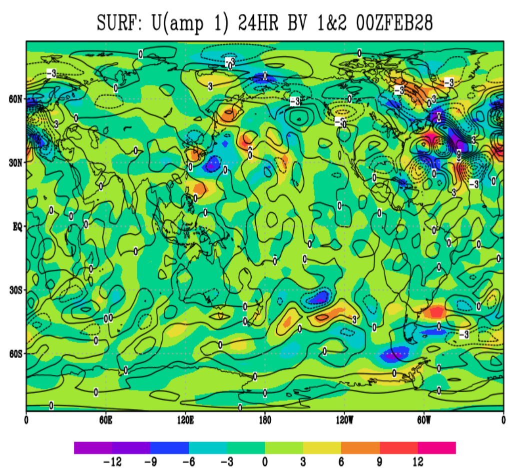
Adrienne Norwood's thesis (2)

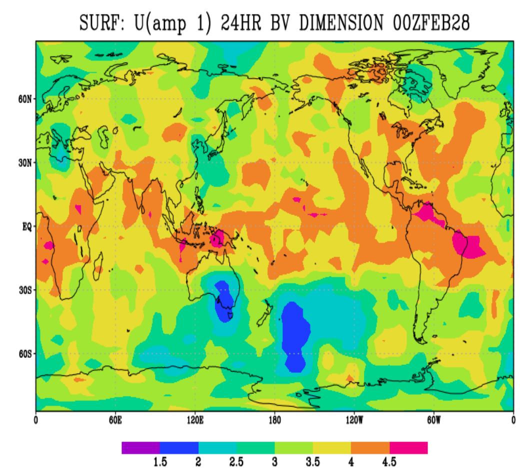
- She then used the SPEEDY model, a realistic GCM with all types of waves and instabilities (e.g., baroclinic waves, inertia Lamb waves, convection).
- We were hoping to find the leading LV with breeding… but BVs did not converge: In regions of baroclinic instability different BVs would lie on top of each other in unstable regions, (as in T and K, 1997) and Patil et al (2006), but not globally.
- So we chose the smallest time scale possible, a time step (40 minutes) and a wind scale very small: 1mm/sec! And then…

What happened with SPEEDY?

- Five BVs were computed. With an amplitude of 1 m/s and integration window of 24 hours, this BV targets baroclinic instabilities, stronger in the winter hemisphere than in the summer hemisphere.
- They do NOT converge to a leading LV!
- But, as in Toth and Kalnay, 1997 and Patil et al., 2001, in regions of baroclinic instability, the BVs tend to locally align with each other, with low E-dimension.

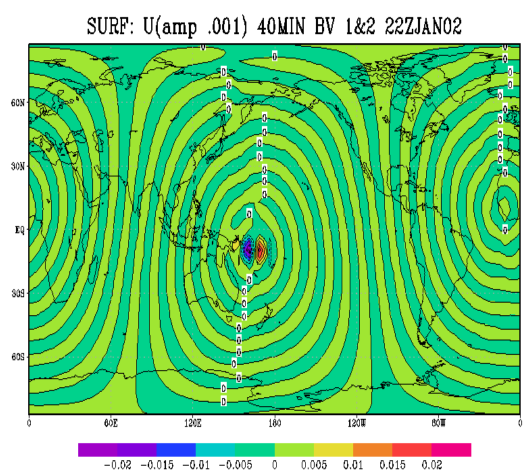
Two different BVs on top of one another.

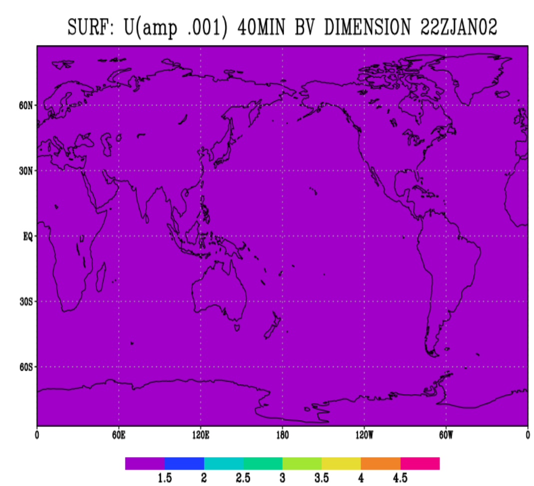




SPEEDY with very small amplitudes and very small rescaling intervals

- If we take a very small amplitude (1 mm/s) and the shortest rescaling window (40 minutes), we obtain a leading LV corresponding to a global Lamb Wave probably triggered by convection in the Warm Pool. It is forced by convective instabilities, with a signal that propagates globally through Lamb waves (horizontal sound waves).
- This is the first time a global leading LV has been found for a full atmospheric model!
- But it may be useless for the creation of ensemble members.

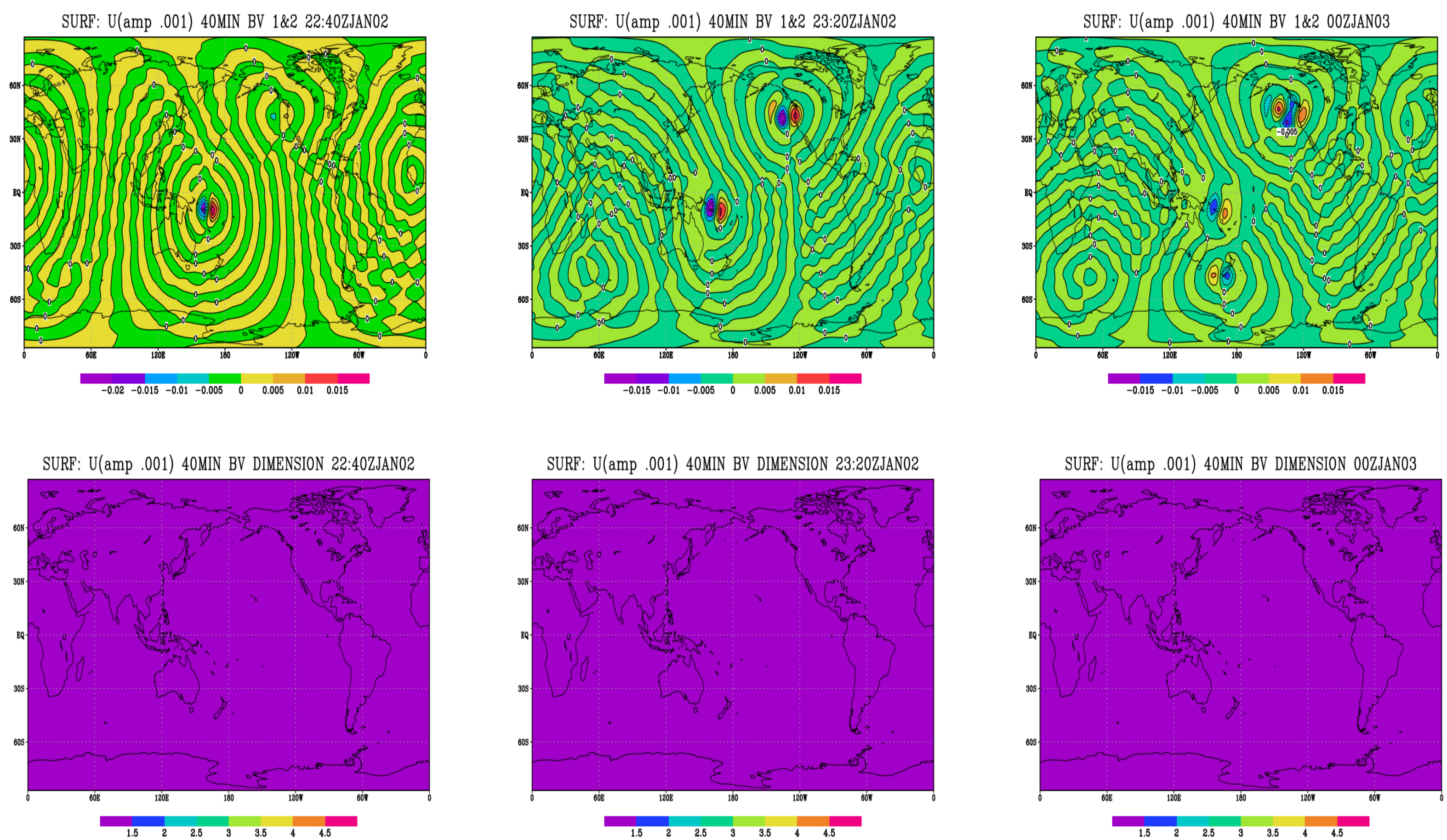




ED=1!

SPEEDY global LV

The leading Lyapunov Vector is not stationary. It is forced by deep convection, but the 5 BV are identical!

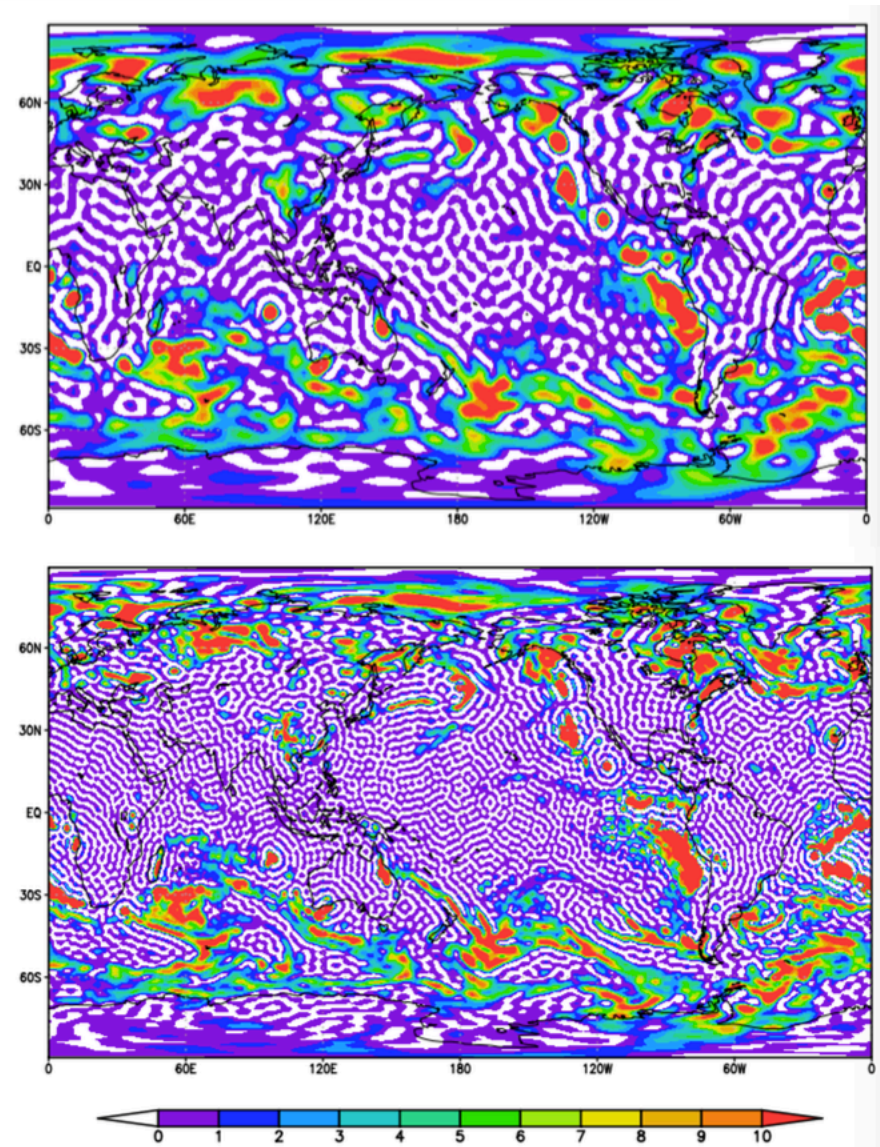


SPEEDY Conclusions (Norwood)

- BVs can target different modes of growth within the model: baroclinic, and external gravity waves forced by convection.
- If we rescale the BV at the shortest possible Δt , and a very small amplitude, we can construct LVs, and they converge to the fastest growing LV.
- There is a leading global LV, as "promised" by Oseledec's Theorem (1968), found through the use of global BVs.
- This is the first time a LLV has been seen in a full atmospheric model.
- BVs associated with baroclinic and convective instabilities do not converge to a single vector because there is no leading global LV associated with these types of instabilities.
- **Are these observations relevant to operational NWP models?**

The GFS assimilating GOES-15 11μm brightness temperature generates a leading Lyapunov Vector just like the SPEEDY model! (Courtesy of Cheng Da, Fuqing Zhang)

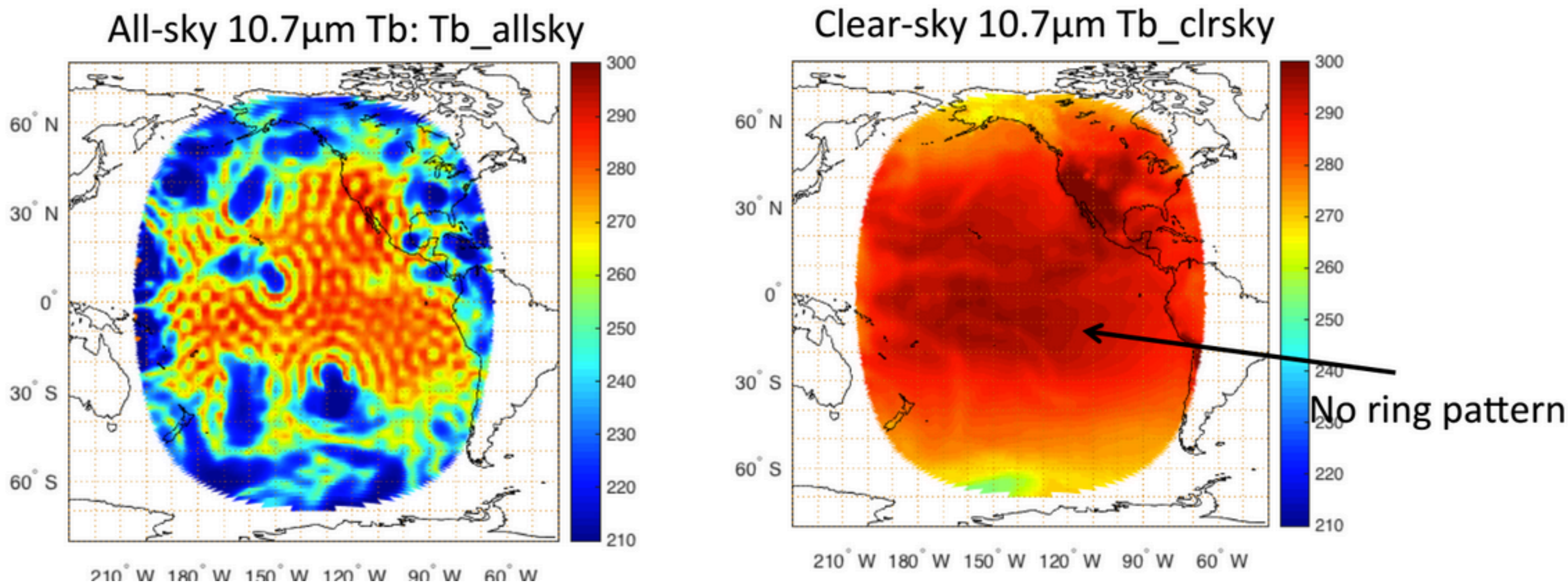
Why? Because "Data Assimilation is like Breeding".
The Oseledec theorem "promises" that there is a global Leading Lyapunov Vector!
The LLV does exist when you do data assimilation, but it is totally irrelevant!



T62

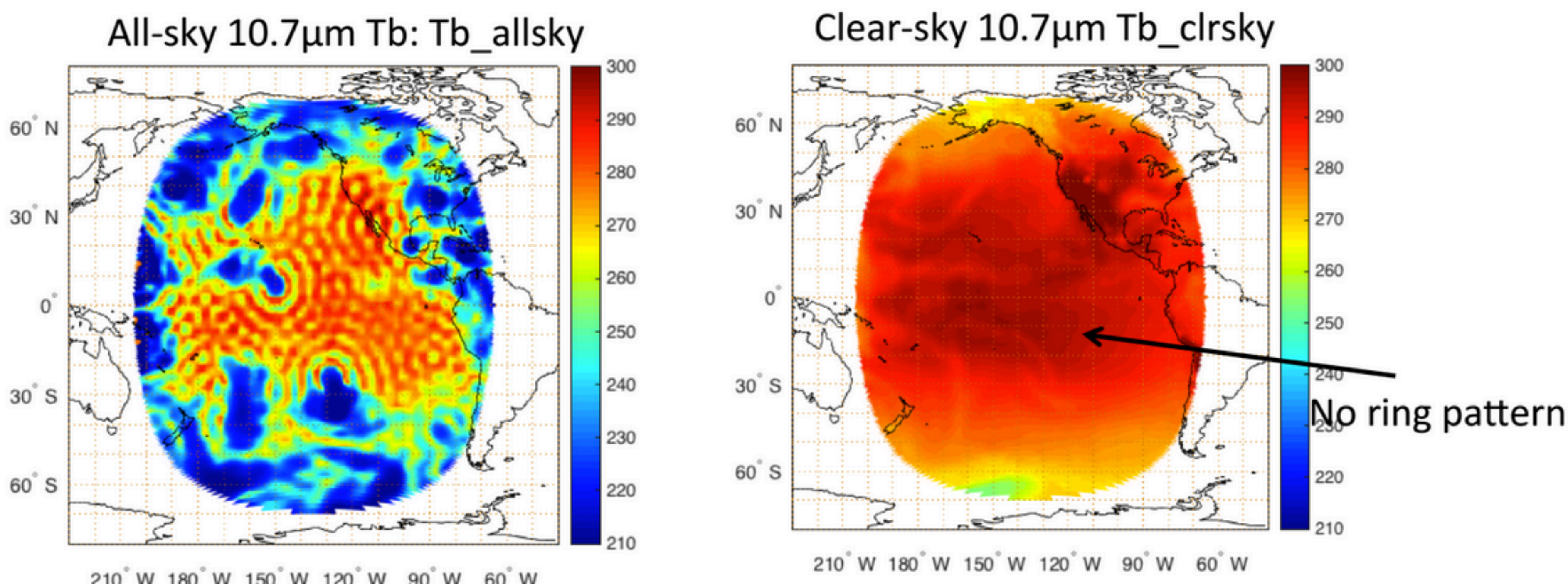
T126

Assimilated All-Sky and Clear-Sky at 10.7μm



As found by Adrienne Norwood, with the SPEEDY model, global leading LV are generated by doing DA (~ breeding), and triggered by tropical convective cells. Again, global leading Lyapunov Vectors exist, but are irrelevant!

Assimilated All-Sky and Clear-Sky at 10.7μm



As found by Adrienne Norwood, with the SPEEDY model, global leading LV are generated by doing DA (~ breeding), and triggered by tropical convective cells. Again, global leading Lyapunov Vectors exist, but are irrelevant!

THANKS!!