AN EMPIRICAL BENCHMARK FOR DECADAL FORECASTS OF GLOBAL SURFACE TEMPERATURE ANOMALIES

Matt Newman
University of Colorado/CIRES and NOAA/ESRL/PSD

Newman, M., 2013: An empirical benchmark for decadal forecasts of global surface temperature anomalies. *J. Climate*, in press, doi 10.1175/JCLI-D-12-00590.1

"Multivariate Red Noise"

- Noise/response is non-local: patterns matter
  - For example, SST sensitive to atmospheric gradient
  - use multivariate ("patterns-based") red noise:

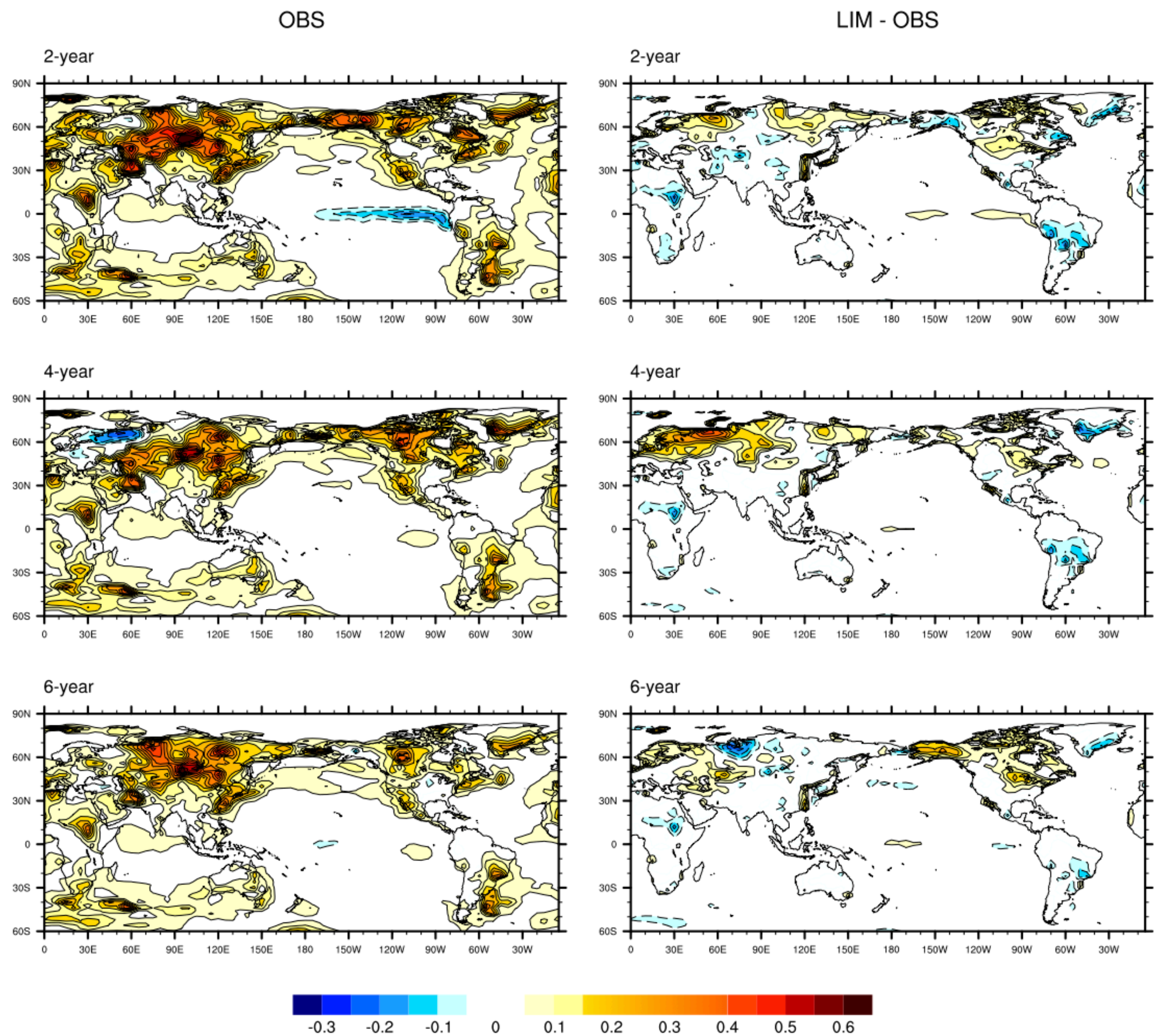
$$d\mathbf{x}/dt = \mathbf{B}\mathbf{x} + \mathbf{F}_s$$

where $\mathbf{x}(t)$ is a series of maps, $\mathbf{B}$ is stable, and $\mathbf{F}_s$ is white noise (maps)

  - Analogous to "univariate red noise", BUT: If **B** is nonnormal (not symmetric), transient anomaly growth is possible even though exponential growth is not
- Determine **B** and $\mathbf{F}_s$ using "Linear Inverse Model" (LIM)
  - ***x*** is ***SST/Land (2m) temperature***, 12-month running mean anomalies, 1901-2009 (HadISST/CRU 3.10)
    - Same used for verification
  - prefiltered in reduced EOF space (17 dof)
  - LIM determined from specified lag (12 months) as in AR1 model

Test of LIM

How well does operator (trained on 1-yr lag) reproduce lag-covariance at longer lags?



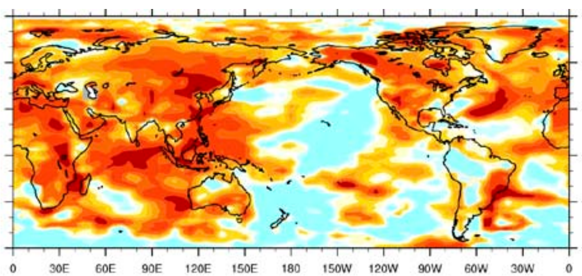
Decadal hindcast skill for forecasts initialized 1960-2000

LIM has clearly higher skill than damped persistence, comparable skill to 3 CMIP5 CGCM decadal hindcasts

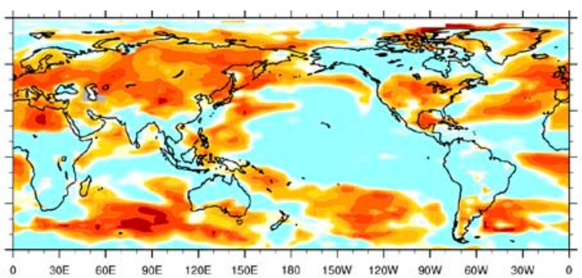
[LIM forecast skill assessed by cross-validation: remove 10% of data, recompute mean/anomalies/EOFs/LIM from remaining 90%, and use this to make forecasts from withheld 10%]

Decadal hindcast skill (Years 2-5) for forecasts initialized 1960-2000, other CMIP5 hindcasts

LIM



CCSM4



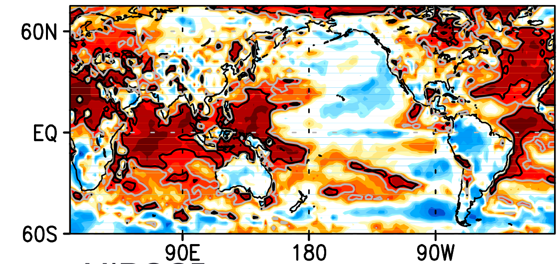
MME

HadCM3

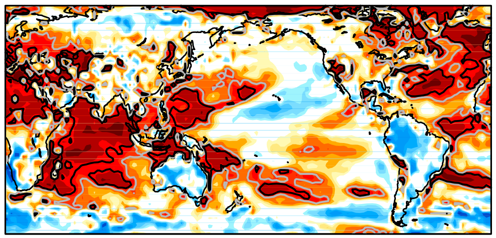
CanCM4

CNRM

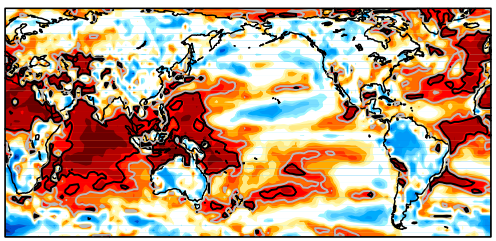
MIROC4h Year02−05



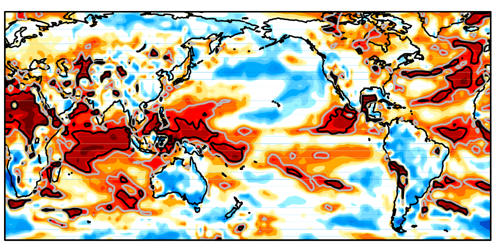
MIROC5



MRI



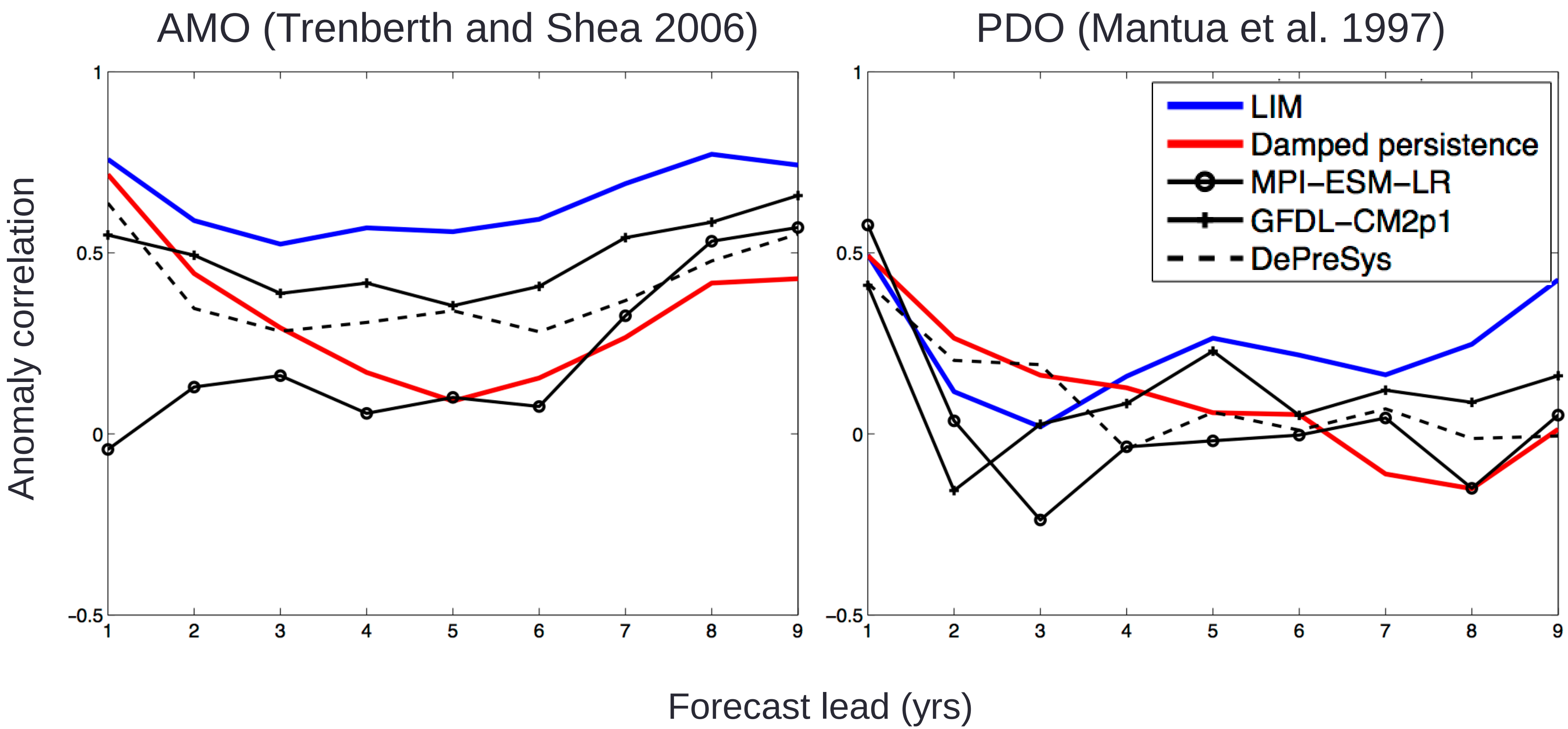
CFSv2



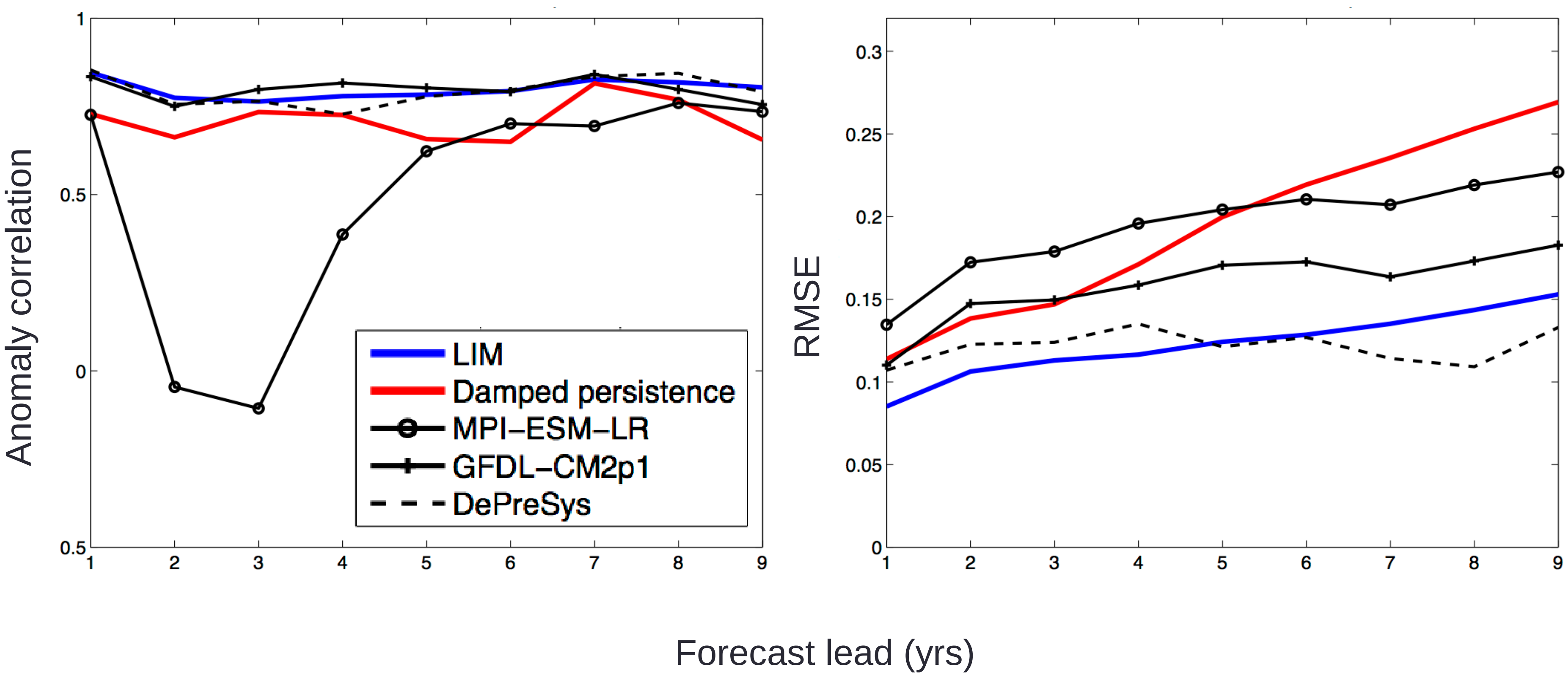
Kim et al., GRL (2012)
ERA40/ERAI verification

AMO and PDO hindcast skill

AMO skill generally higher than PDO skill

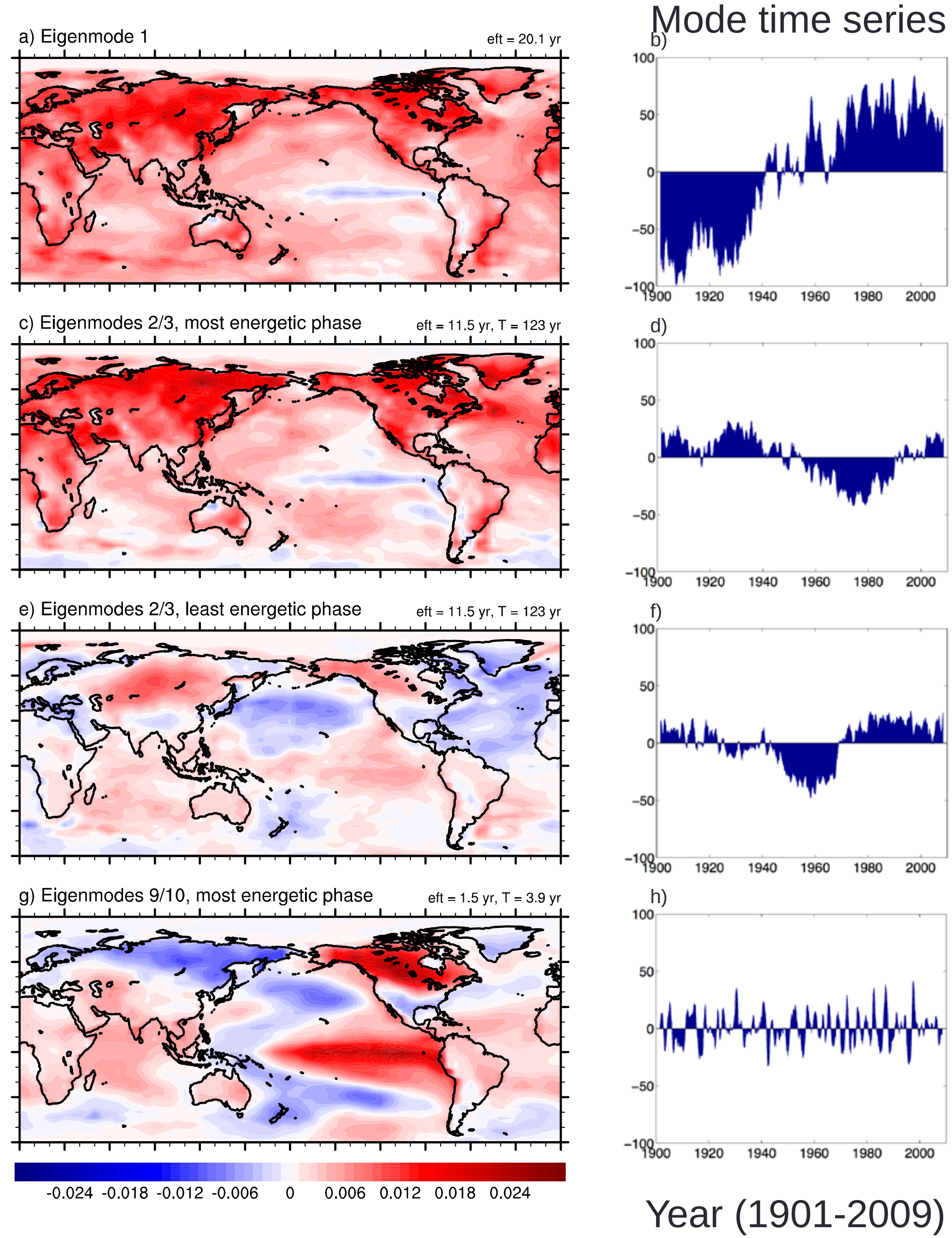


Global mean temperature hindcast skill



Leading eigenmodes of **B**

Not shown: Decadal "ENSO" eigenmode, which has decorrelation time scale of under about 2 years



1901-2009 hindcast skill is mostly due to the leading three eigenmodes

Years 2-5

Years 6-9

LIM

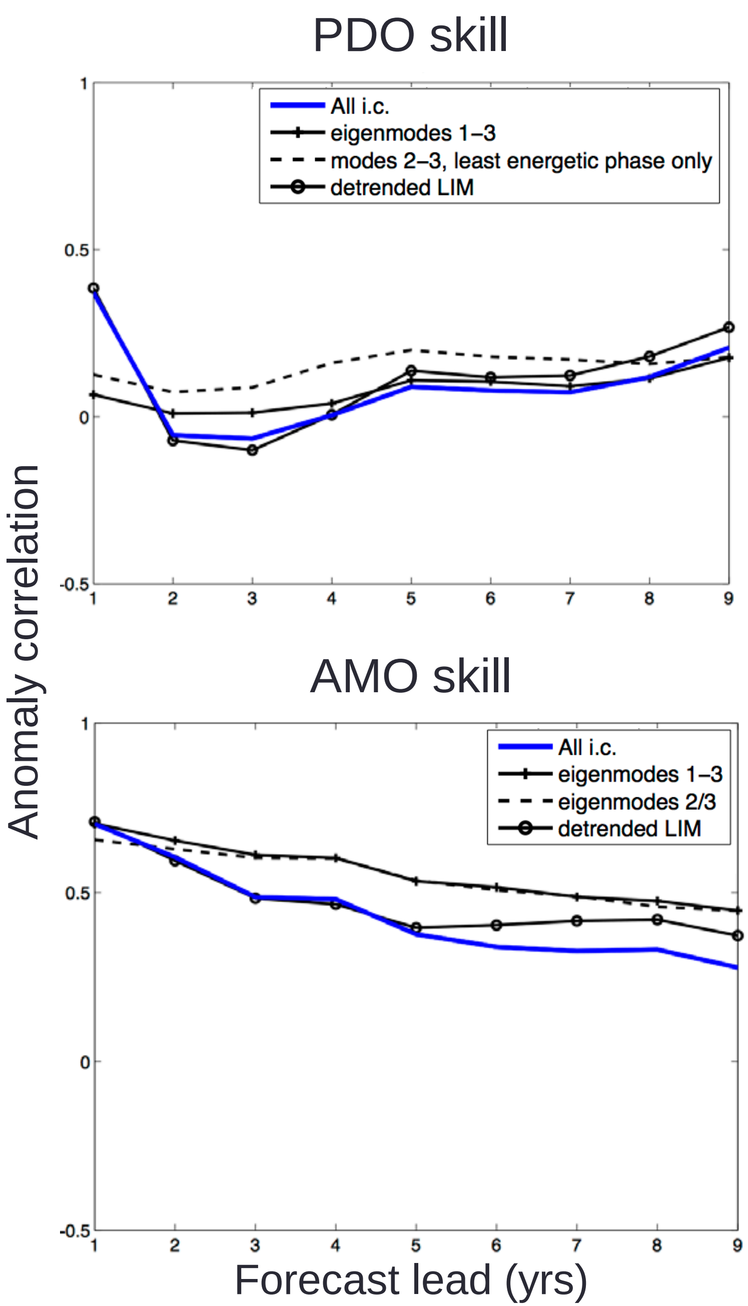
LIM
–
"eigenmodes 1-3"
LIM

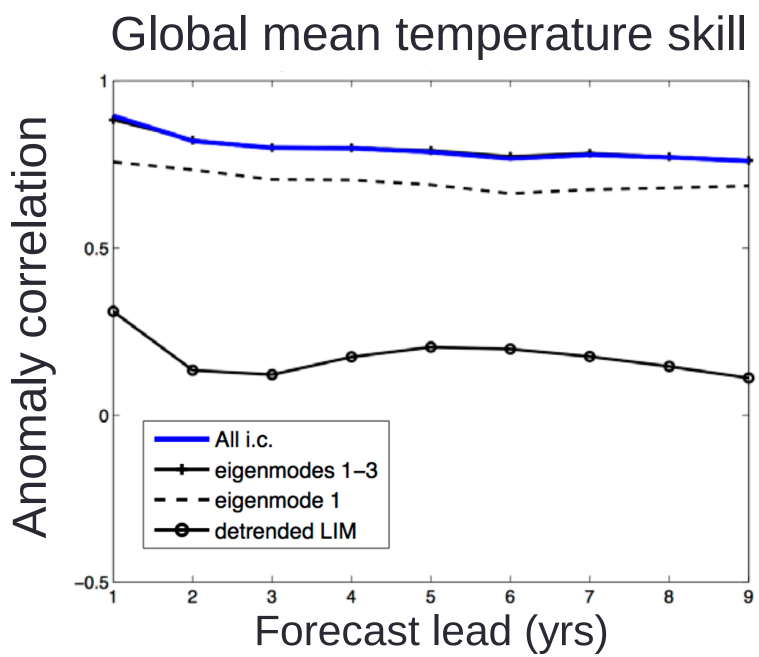
Detrended LIM

Eigenmode dependence of climate index hindcast skill

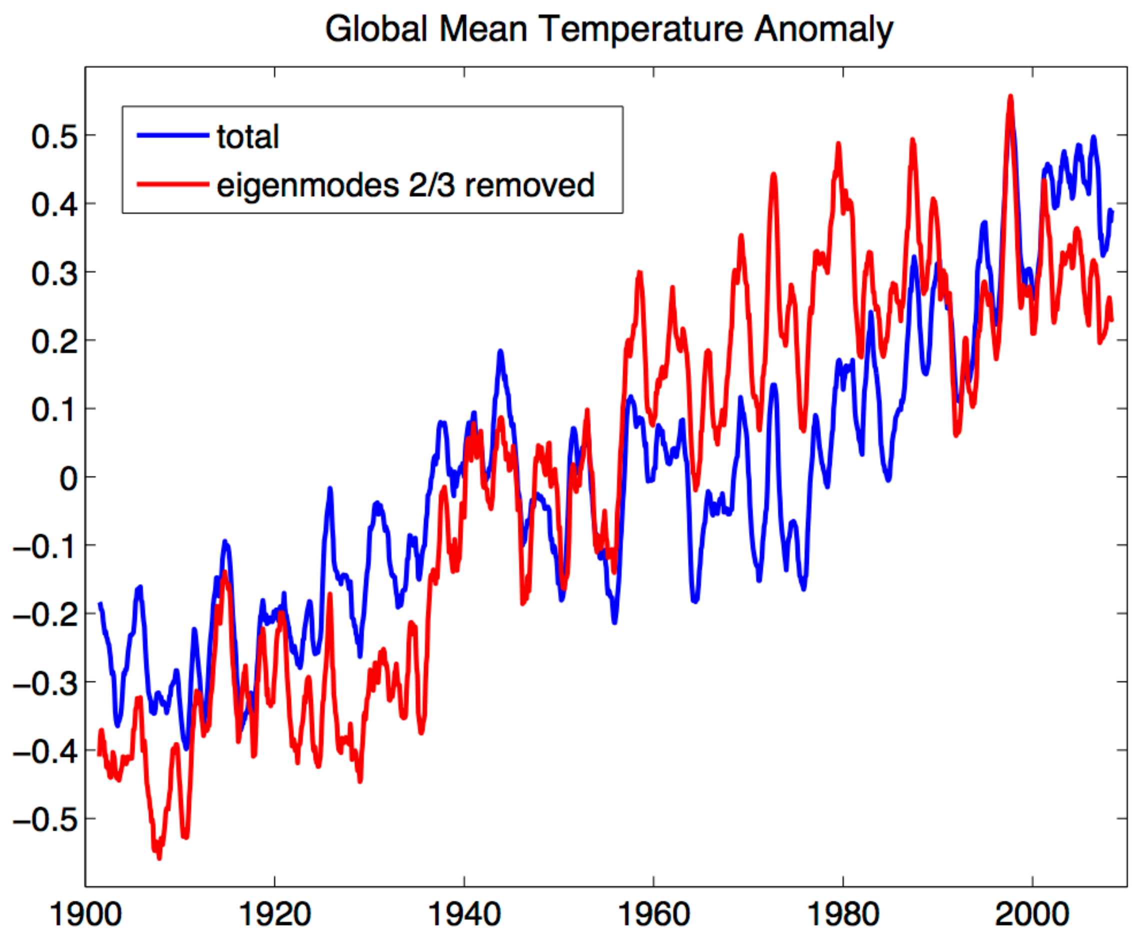
Global mean temperature: eigenmode 1
AMO/PDO: most/least energetic phase of eigenmode pair 2/3

Trend has little impact on AMO, PDO skill





Eigenmodes 2/3 (energetic phase) impacts global mean trend

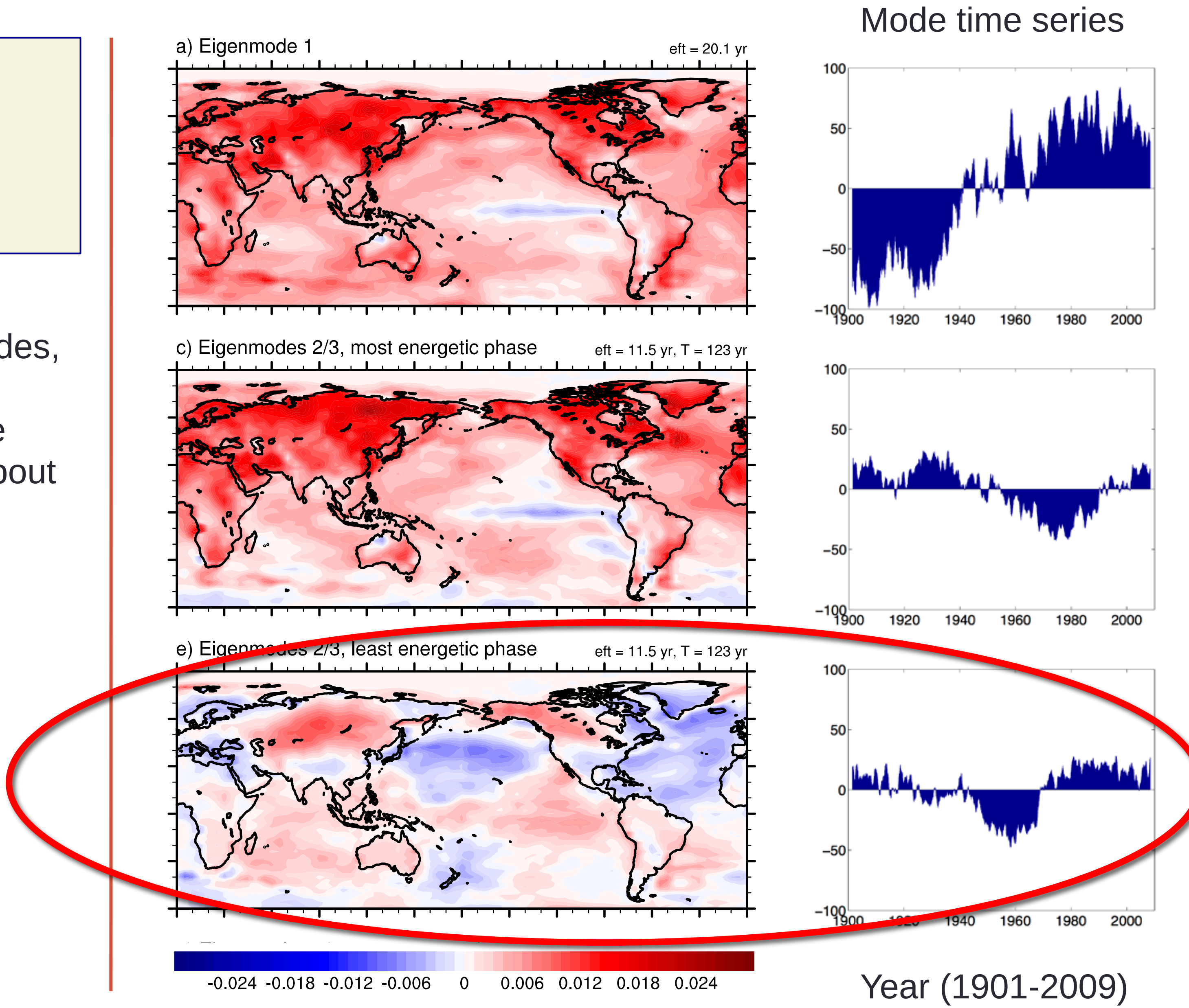


Leading eigenmodes of **B**

Not shown:

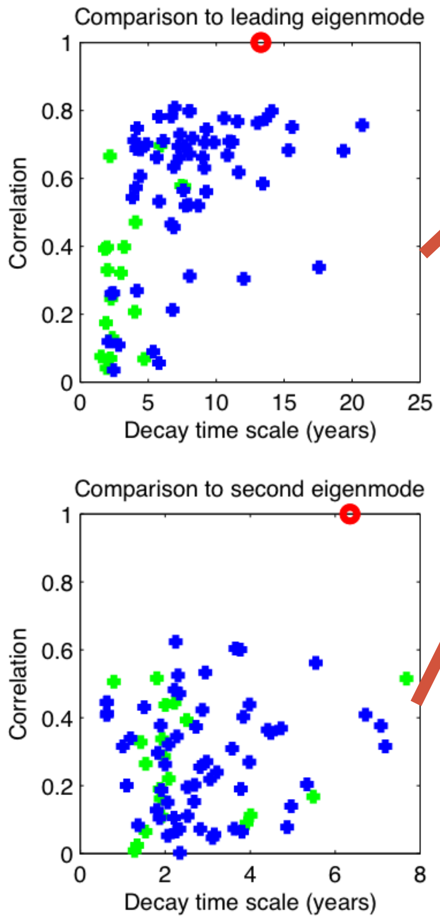
"ENSO" eigenmodes, which have decorrelation time scale less than about 2 years

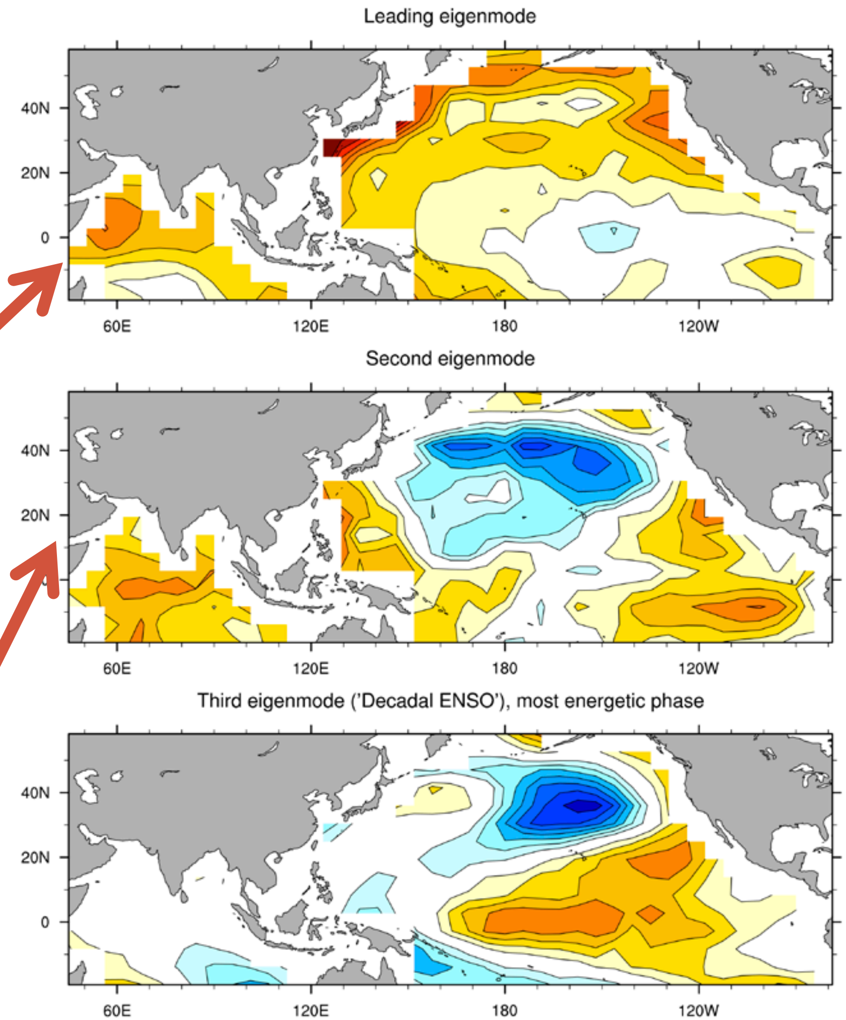
What's left of the PDO after ENSO influences are removed; poorly captured in Pacific by CMIP3 models (Newman 2007; Solomon et al. 2011)



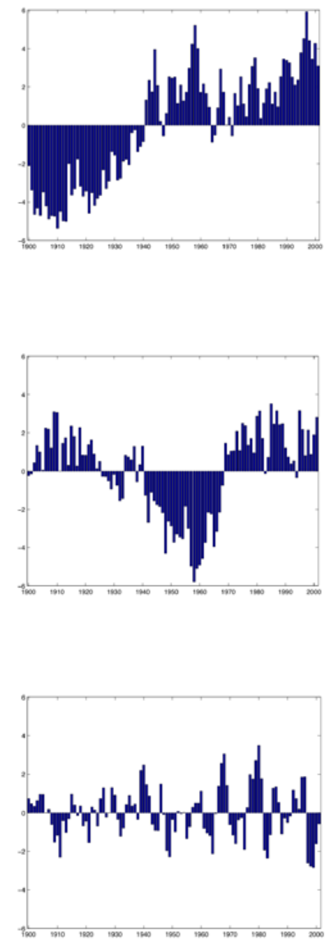
Eigenmodes of Pacific-SST **B** and the PDO

Second eigenmode poorly captured in CMIP3





Mode time series

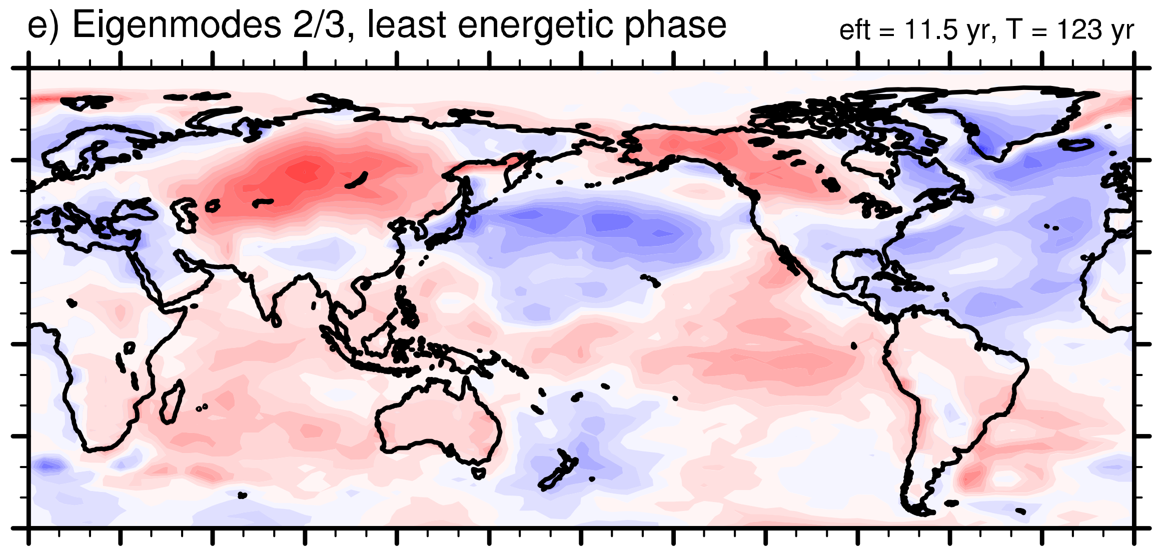


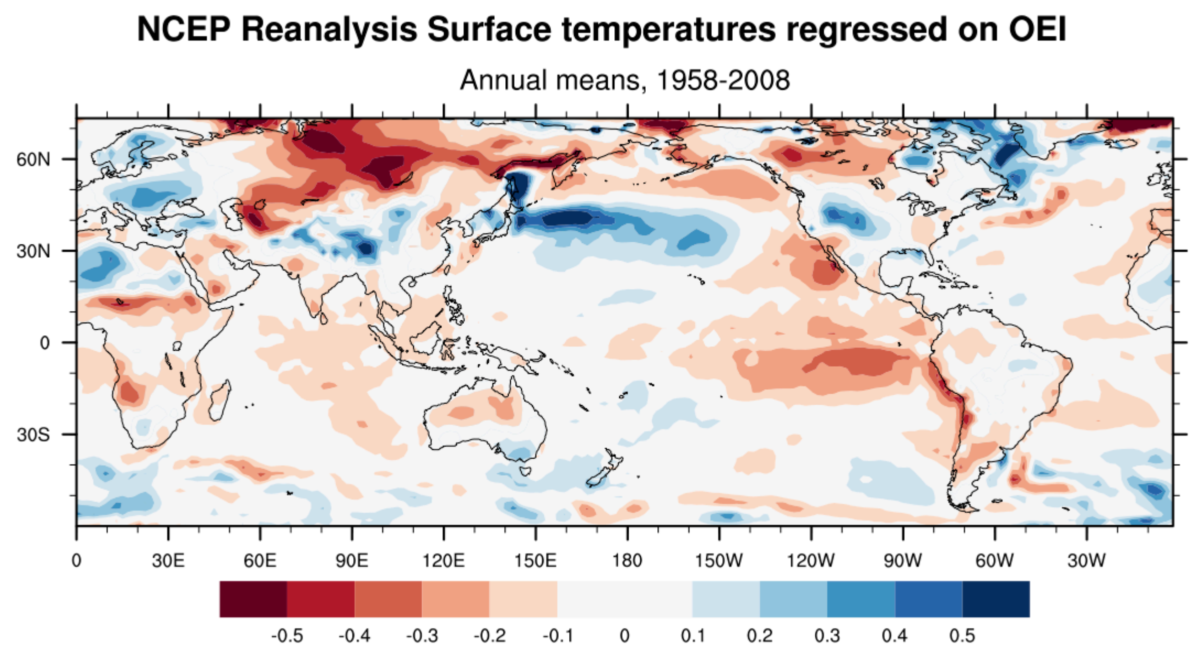
Year (1901-2001)

Solomon et al. 2011

Second eigenmode looks similar to “Oyashio-Extension Index” (OEI) pattern

OEI: leading PC of the latitude of the maximum meridional SST gradient between 145º and 170ºE (Frankignoul et al. 2011)





Conclusions

- Multivariate red noise (determined empirically by LIM) is a useful benchmark for forecasts of global surface temperature anomalies
  - Patterns of skill similar between LIM and CGCMs
- Long-range hindcast skill is:
  - represented by three nonorthogonal patterns with relatively long decorrelation time scales
  - degraded by ENSO, which acts as noise in tropics/North Pacific
  - largely *but not entirely* due to trend
- Decadal CGCM forecasts need to beat LIM forecasts
  - But they don't yet (so LIM forecasts coming soon)
  - Better empirical forecasts possible but severely limited by data
  - Potential skill in Pacific (maybe related to Oyashio Extension) as yet unrealized by CGCMs?