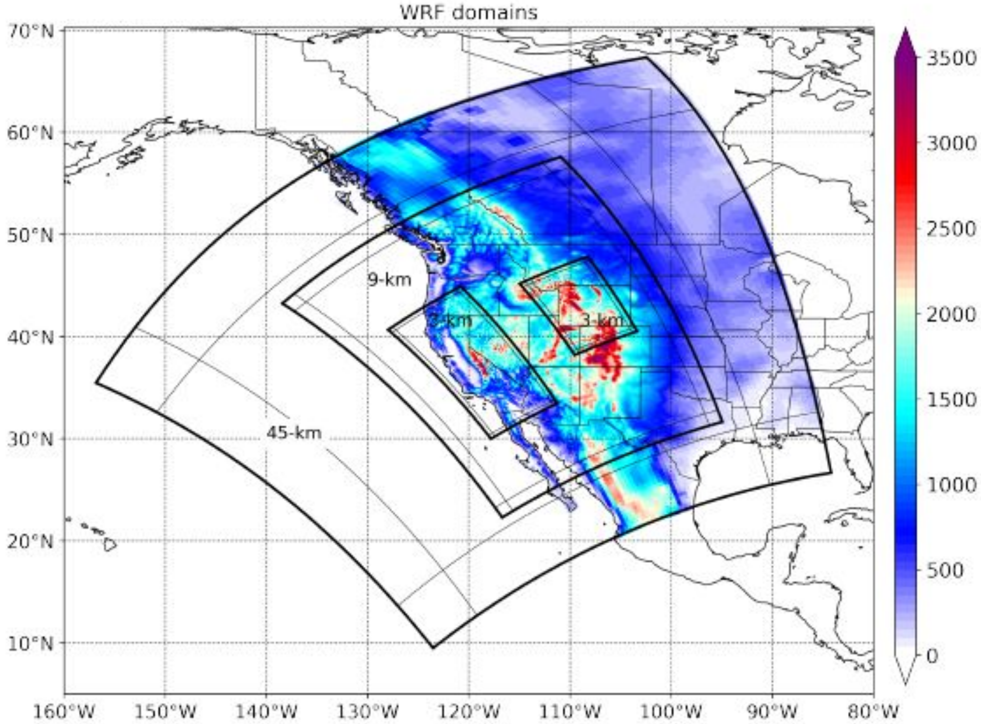
Finding a subset of a large ensemble to maximize variance for downscaling

Naomi Goldenson
with contributions from Andy Rhines (now at Netflix)

This study: California and western US

Ongoing dynamical downscaling at UCLA using these domains was the motivation here:



Rahimi et al., 2022

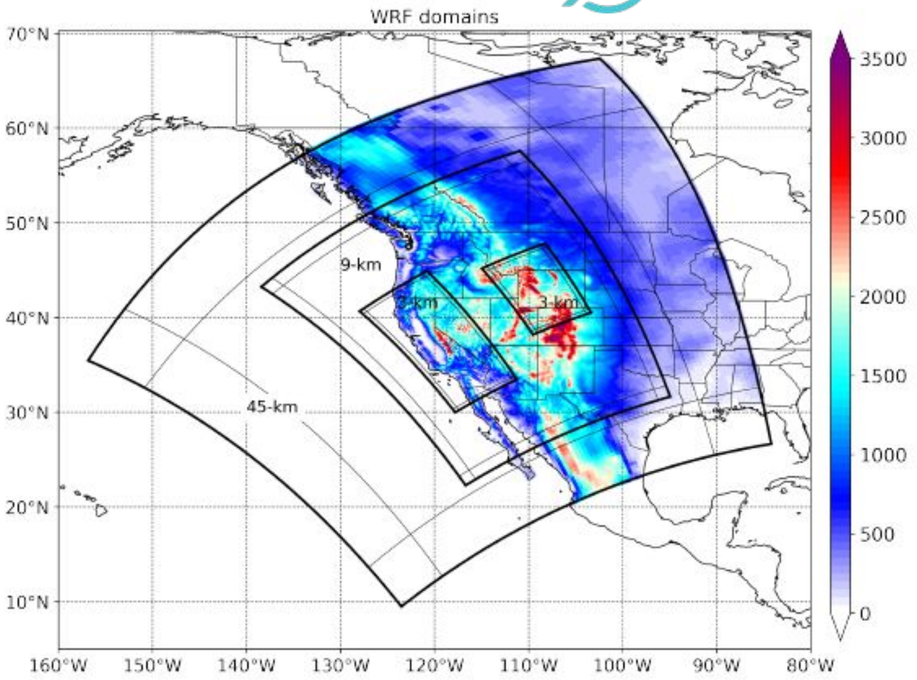
We simulate regional climate for multiple reasons

Science: better understand regional processes

Model development: evaluate and improve simulations

Planning: data applied to climate impacts assessments

UCLA Center for Climate Science



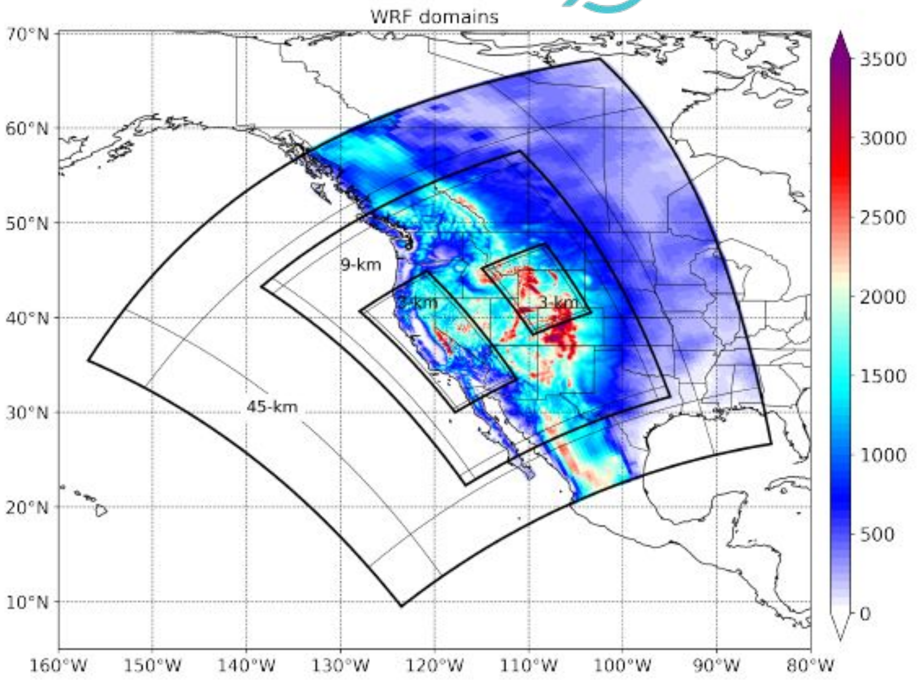
Rahimi et al (2021, in prep)

When people come looking for data...

Science: better understand regional processes

Model development: evaluate and improve simulations

Planning: data applied to climate impacts assessments

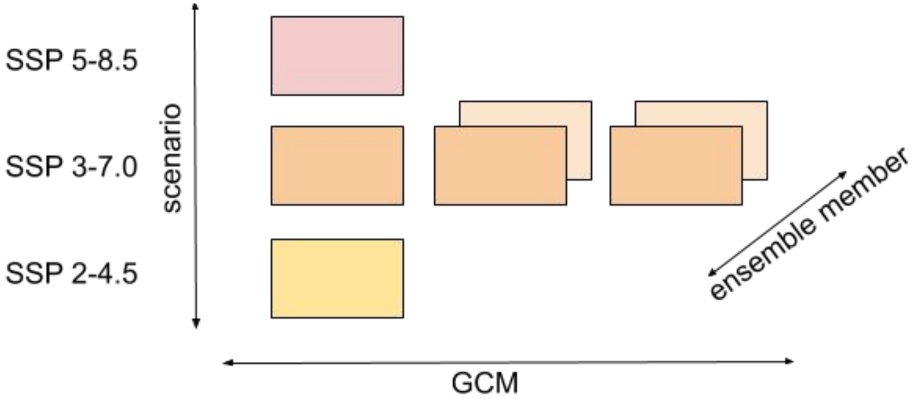


Rahimi et al (2021, in prep)

Building an ensemble of downscaled simulations

Focus today is on the multiple ensemble member axis

– can we do better than randomly sampling internal variance?



Two reasons to select ensemble members thoughtfully:

1. Making sure to include extreme events / seasons around which to build storylines (aka plausible timeseries that you could use for scenario planning)

Climatic Change (2018) 151:555–571
https://doi.org/10.1007/s10584-018-2317-9

CrossMark

Storylines: an alternative approach to representing uncertainty in physical aspects of climate change

Theodore G. Shepherd[1] · Emily Boyd[2] · Raphael A. Calel[3,4] · Sandra C. Chapman[5,6] · Suraje Dessai[7] · Ioana M. Dima-West[8] · Hayley J. Fowler[9] · Rachel James[10,11] · Douglas Maraun[12] · Olivia Martius[13] · Catherine A. Senior[14] · Adam H. Sobel[15] · David A. Stainforth[4,5] · Simon F. B. Tett[16] · Kevin E. Trenberth[17] · Bart J. J. M. van den Hurk[18,19] · Nicholas W. Watkins[4,5,6,20] · Robert L. Wilby[21] · Dimitri A. Zenghelis[4]

Received: 14 November 2017 / Accepted: 19 October 2018 / Published online: 10 November 2018

Abstract
As climate change research becomes increasingly applied, the need for actionable information is

PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Check for updates

Cite this article: Shepherd TG. 2019 Storyline approach to the construction of regional climate change information. *Proc. R. Soc. A* **475**: 20190013.
http://dx.doi.org/10.1098/rspa.2019.0013

Received: 9 January 2019

Storyline approach to the construction of regional climate change information

Theodore G. Shepherd

Department of Meteorology, University of Reading, PO Box 243, Earley Gate, Whiteknights, Reading RG6 6BB, UK

TGS, 0000-0002-6631-9968

Climate science seeks to make statements of confidence about what has happened, and what will happen (conditional on scenario). The approach is effective for the global, thermodynamic aspects of climate change, but is ineffective when it

Two reasons to select ensemble members thoughtfully:

1. Making sure to include extreme events / seasons around which to build storylines (aka plausible timeseries that you could use for scenario planning)
2. Enable storylines to be tied to something physically meaningful – e.g. explain the source of the regional variability from large-scale modes

Goal:

- Choose *n* large ensemble members that span the range of trends and variance, and include some extreme extremes.

Approach:

1. Cluster members across *m* axes chosen based on physical reasoning.
2. Find all sets of *n* that are in different clusters <u>from each other</u> for all *m* items.
3. *also* limited to the members that are downscalable
4. Add constraints for more extreme extremes until there is one set left

Goal:

- Choose *n* large ensemble members that span the range of trends and variance, and include some extreme extremes.

n=3

Approach:

1. Cluster members across *m* axes chosen based on physical reasoning.
2. Find all sets of *n* that are in different clusters from each other for all *m* items.
3. *also* limited to the members that are downscalable
4. Add constraints for more extreme extremes until there is one set left

m=3 also

10/100 from CESM2

Large uncertainties in California winter precipitation:

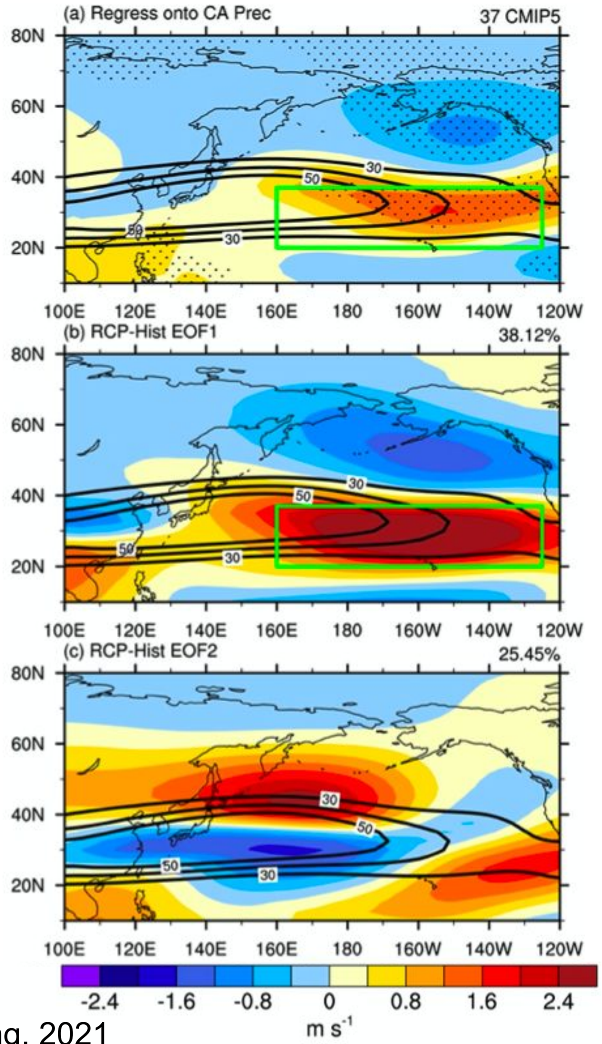
Mostly unavoidable internal variance, but reason to believe some explanatory power comes from **ENSO**.

Therefore, some of my metrics will use the phasing of ENSO variability.

Large uncertainties in California winter precipitation:

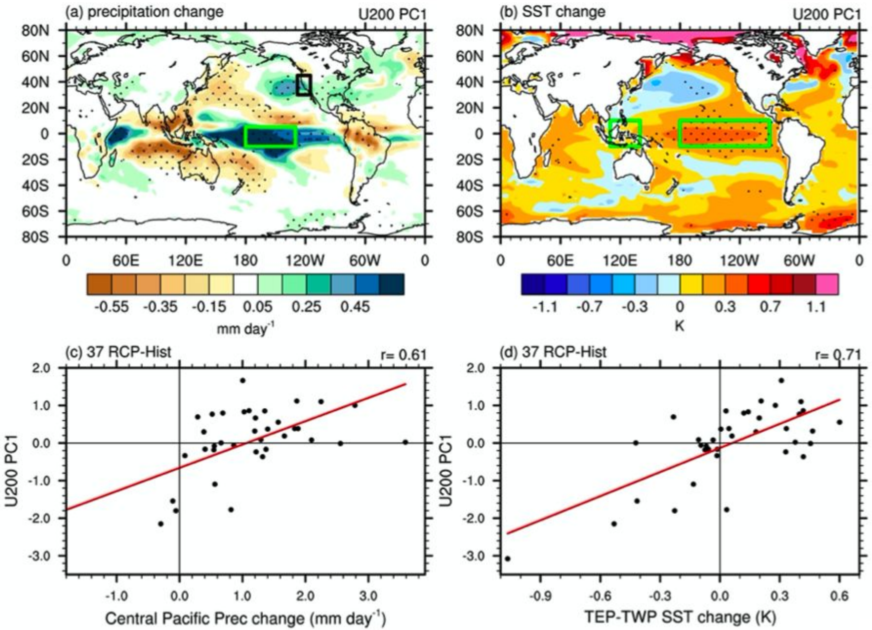
Mostly unavoidable internal variance, but reason to believe some explanatory power comes from **ENSO**.

Strength of jet extension mode well-correlated with cross-model CA precipitation changes.

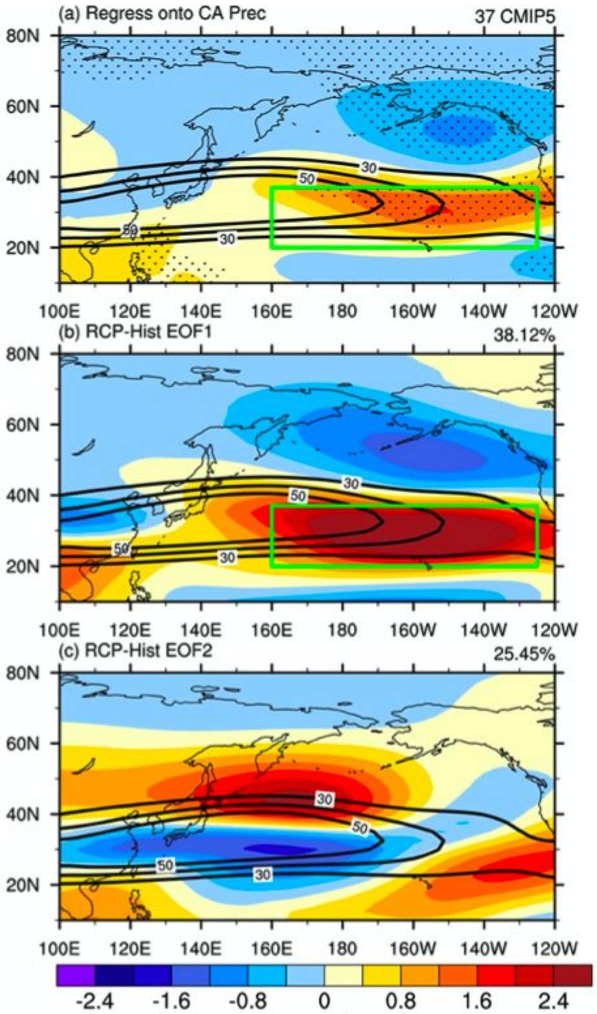


Dong & Leung, 2021

And this jet-extention mode strength is related to the extent to which a GCM becomes more El Niño-like in the future.



Strength of jet extension mode well-correlated with cross-model CA precipitation changes.



Dong & Leung, 2021

The axes across which to cluster

1. Decadally-smoothed ENSO variability phasing
2. Trend in El Niño-like pattern
3. Local temperature and precipitation trends

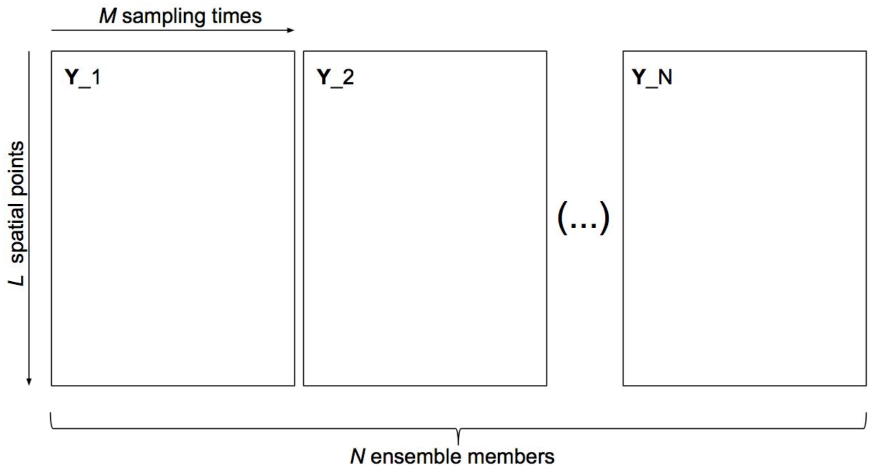
BONUS:

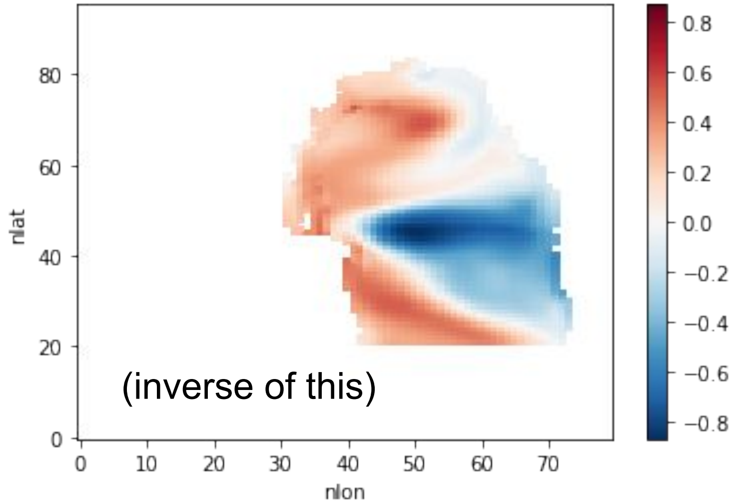
4. Local temperature and precipitation extremes

Use these to narrow it down until there's one combination left.

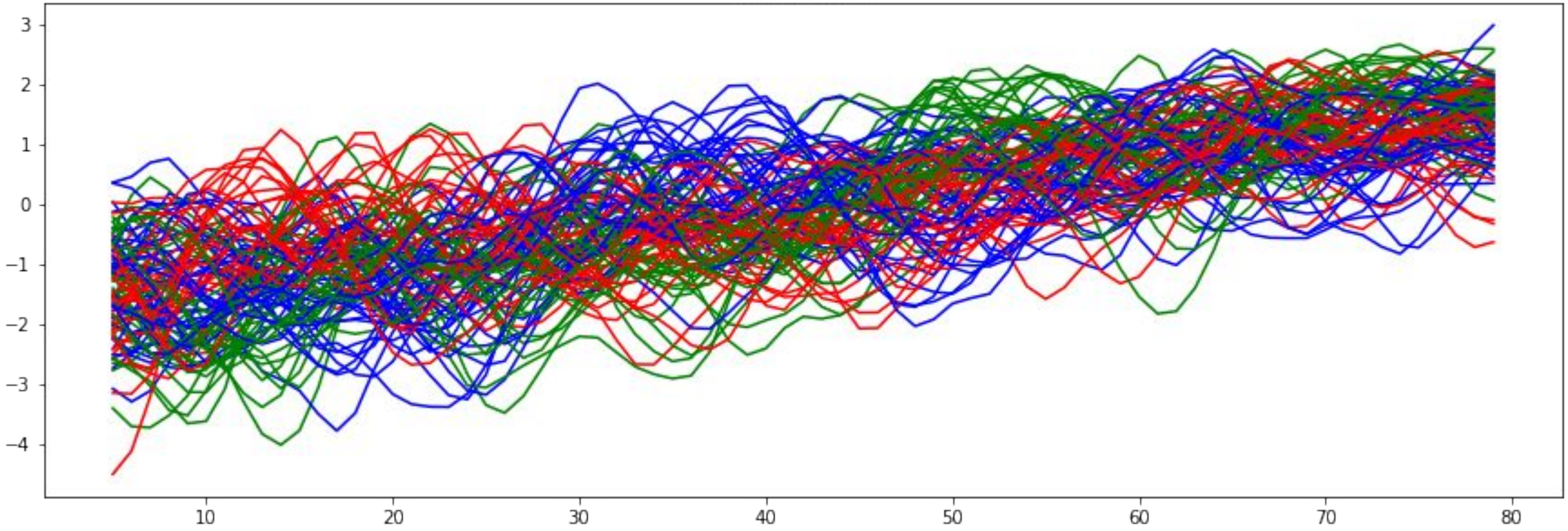
ENSO methods

- First we need to define the PC time-series for all ensemble members with respect to one consistent loading pattern for ENSO.
- We use the first ensemble EOF – calculated based on all 100 ensemble members SSTs stacked together at once.
- This is the stable pattern that reflects this GCM's ENSO behavior.



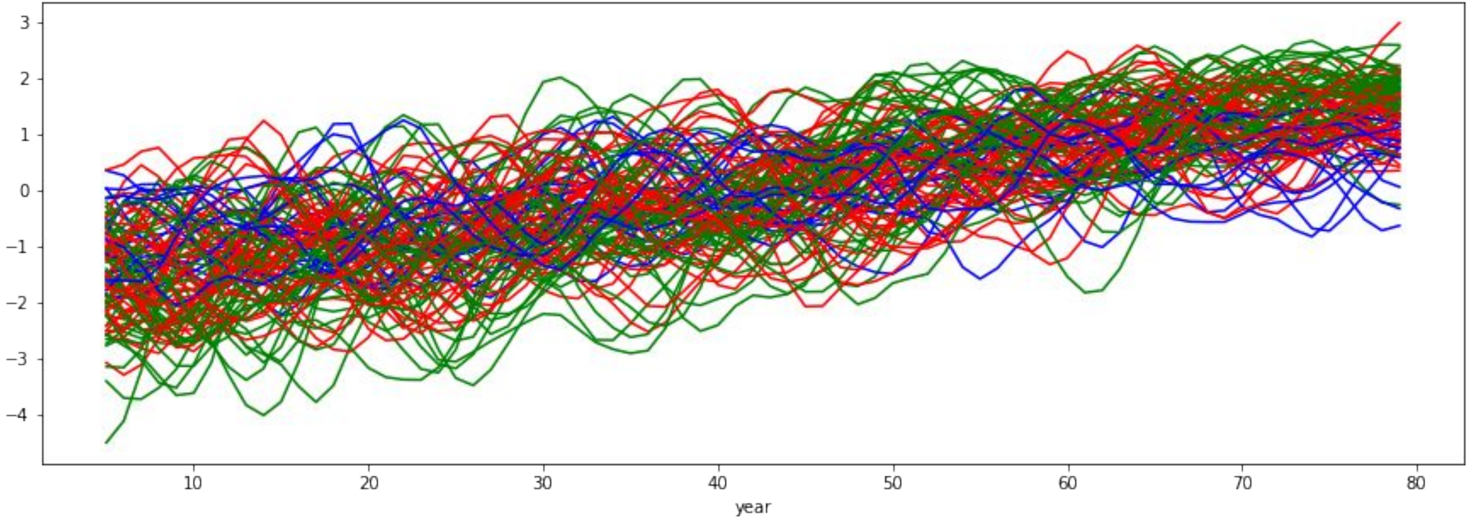


Three clusters on the smoothed PC time-series separate those with a lot of El Niño years in the earlier, mid, and mid-late parts of the time-series.



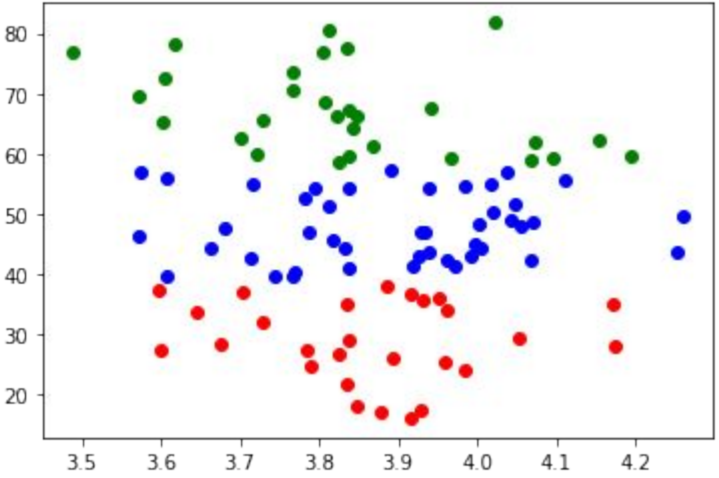
Years (from 2015)

2. How strong a trend to a more El Niño-like future?



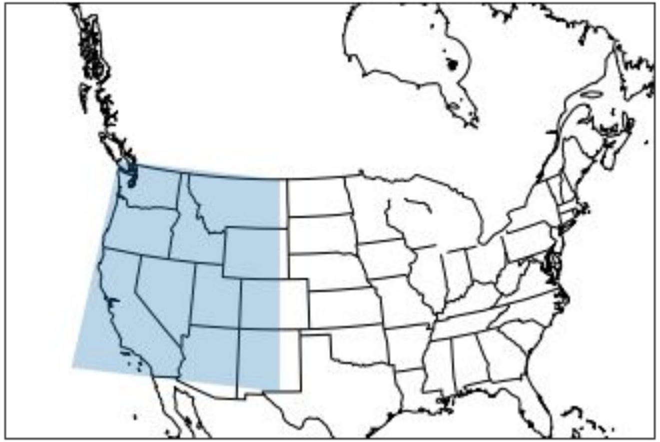
Note: none of the 10 downscalable ensemble members fall in the blue cluster. So we'll require that we have at least one from each of the other two clusters for this measure.

3. Local precipitation trend



Δprecipitation (mm/yr)

Δtemperature (˚C)



(2070-2099 minus 1950-1979)

4. Extremes:

Further limit our list of possible combinations of three ensemble members so that **at least one of the three must contain**:

a) >= 4 wet winters at least as wet as the 99th percentile historical 90-day running maximum for the Sep-Aug year

27/100 ensemble members

b) at least two single annual max hot days that exceed the historical 99th percentile

23/100 ensemble members

c) >= 4 instances of a 10-year rolling mean below the 1st historical percentile water year total

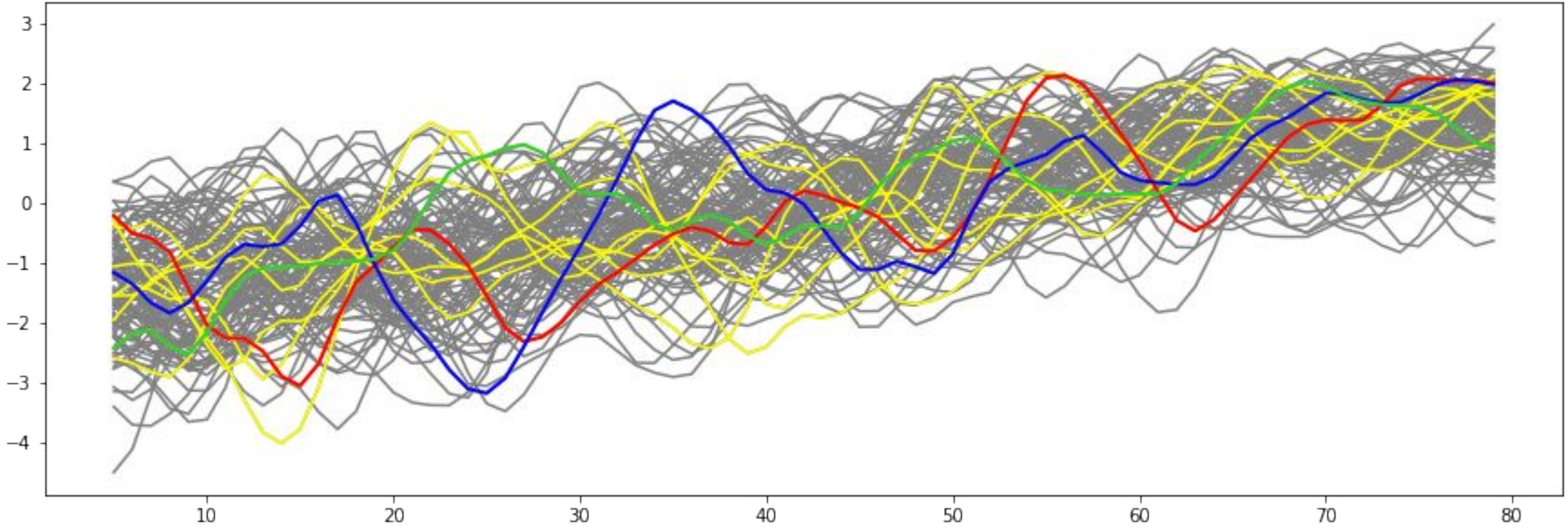
47/100 ensemble members

Having applied all these constraints:

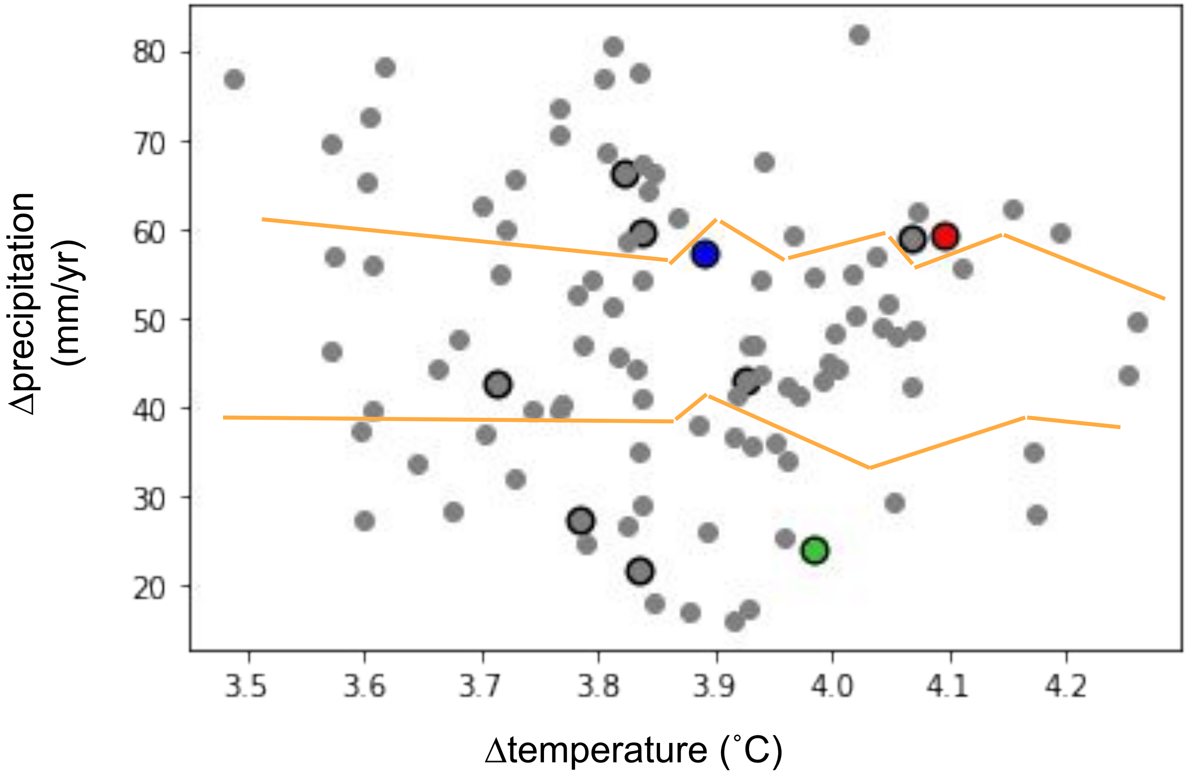
['smbb.LE2-1151.008', 'smbb.LE2-1051.003', 'smbb.LE2-1011.007']

Could easily tighten/loosen the extremes specified until there's only one possible set left.

Highlighting selected models

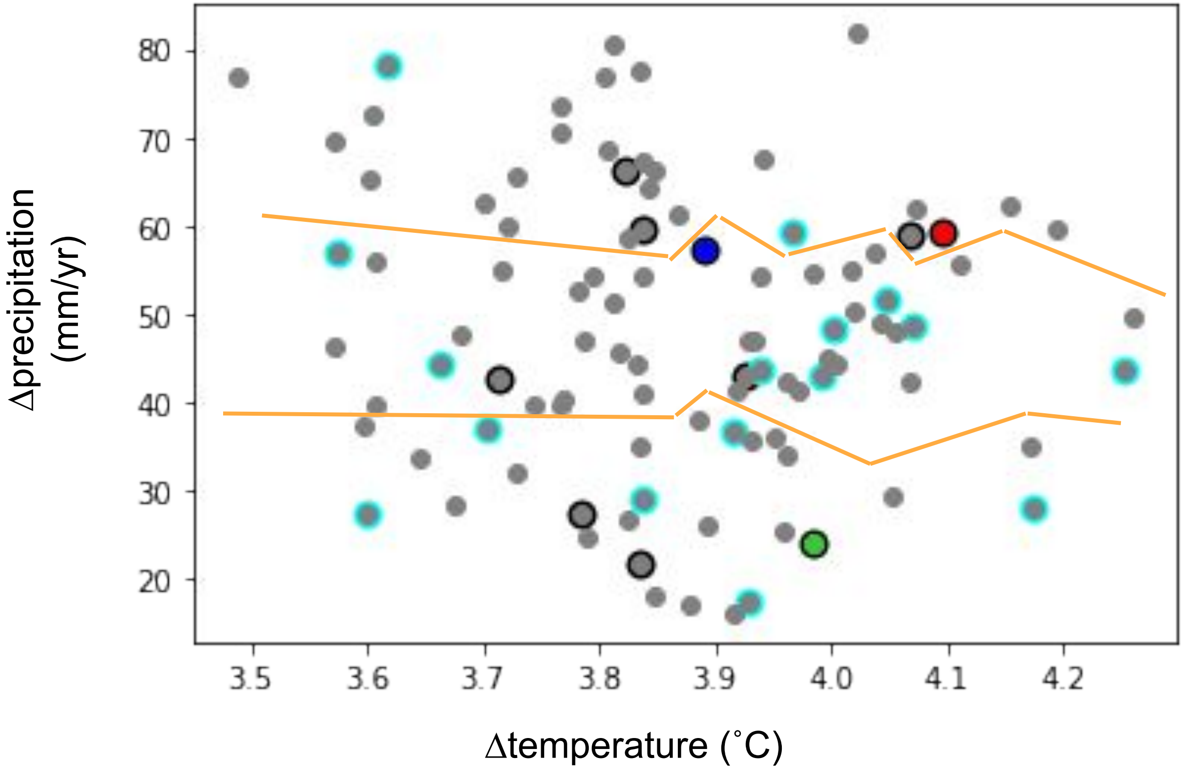


downscalable



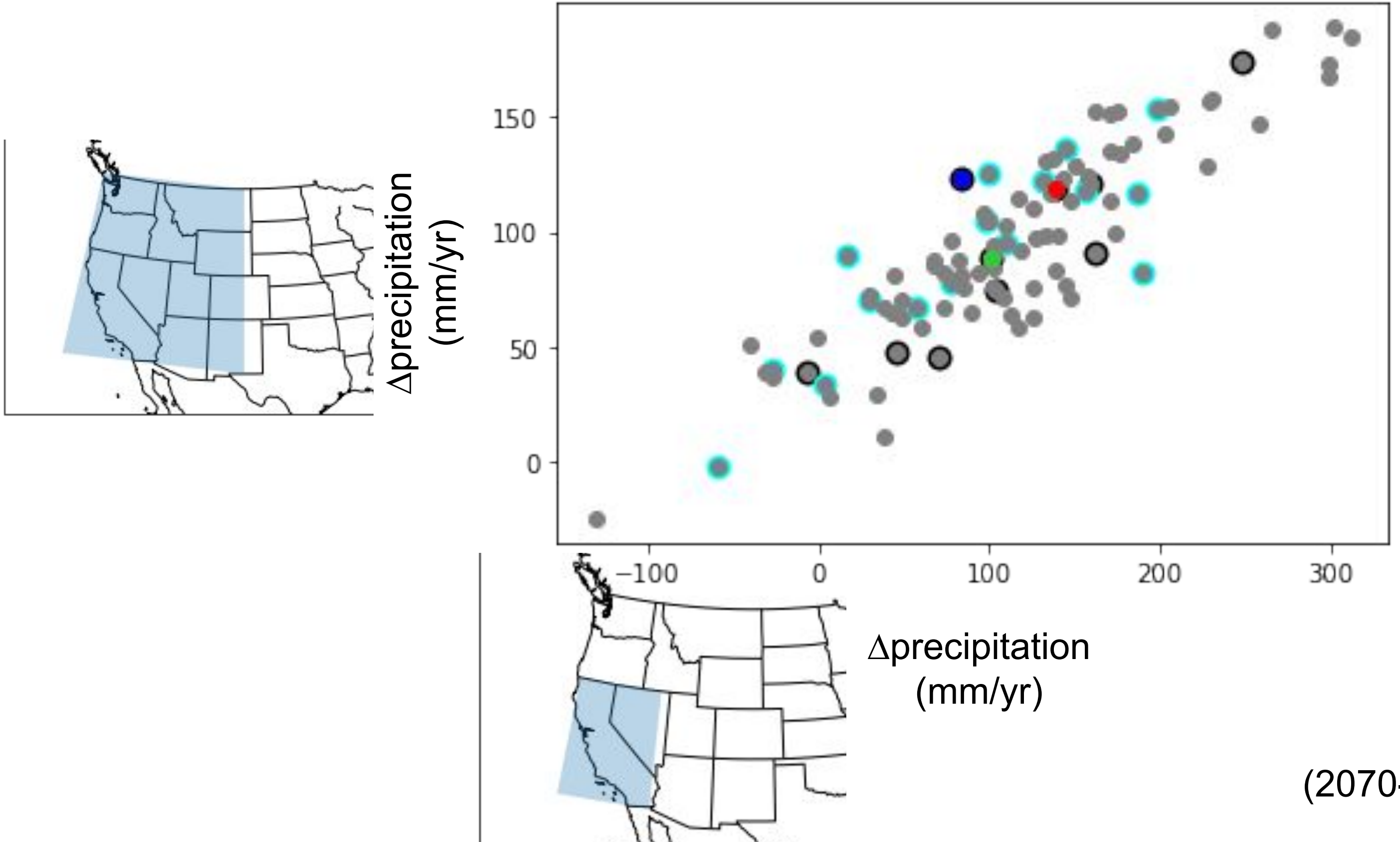
(2070-2099 minus 1950-1979)

Weaker trend in El Niño-like SST pattern does not yield any particular regional response

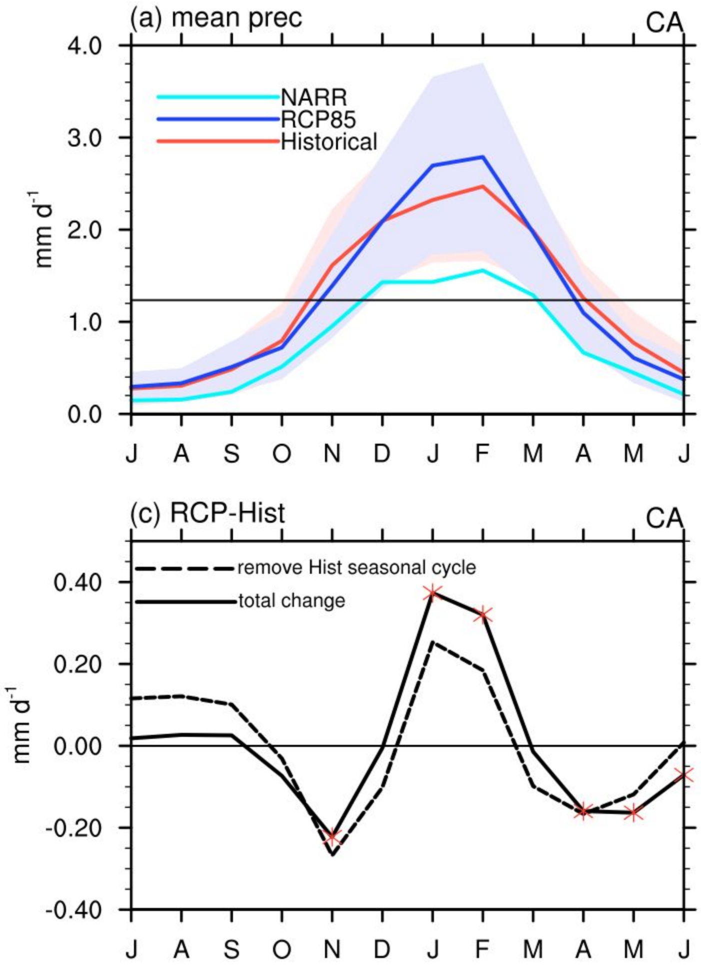


(2070-2099 minus 1950-1979)

Weaker trend in El Niño-like SST pattern does not yield any particular regional response, even if limited to DJF



(2070-2099 minus 1950-1979)



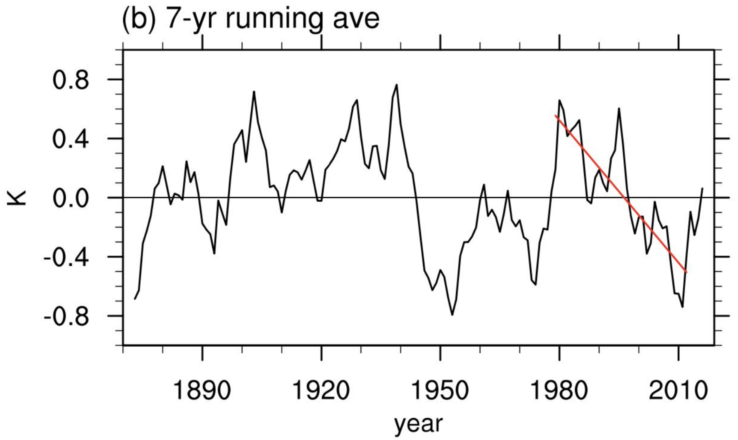
Expected a stronger signal over California in just DJF because of the documented sharpening of the seasonal cycle we see in simulations.

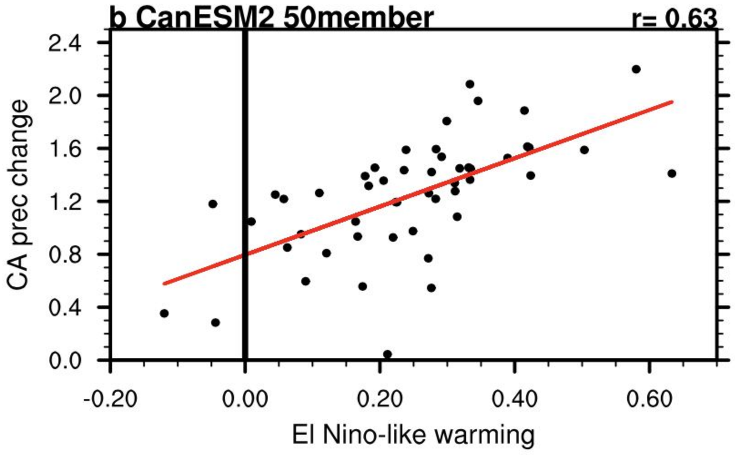
Mechanisms for an Amplified Precipitation Seasonal Cycle in the U.S. West Coast under Global Warming

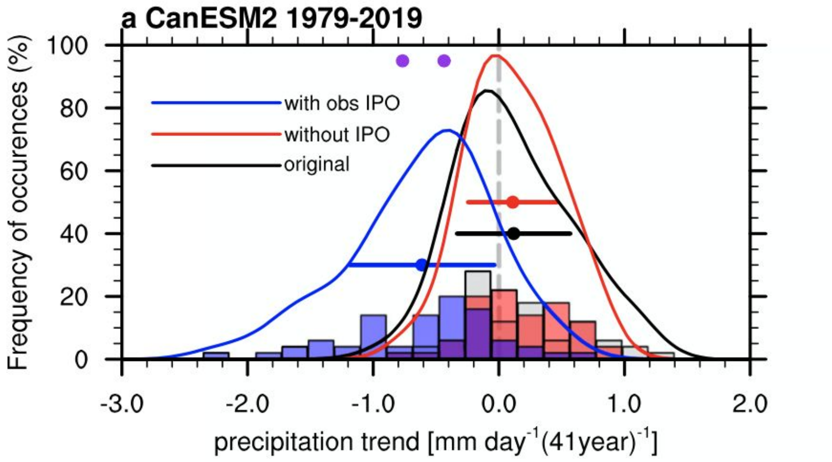
LU DONG, L. RUBY LEUNG, JIAN LU, AND FENGFEI SONG

Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington

Positive-to-negative phase transition of IPO in historical observations:

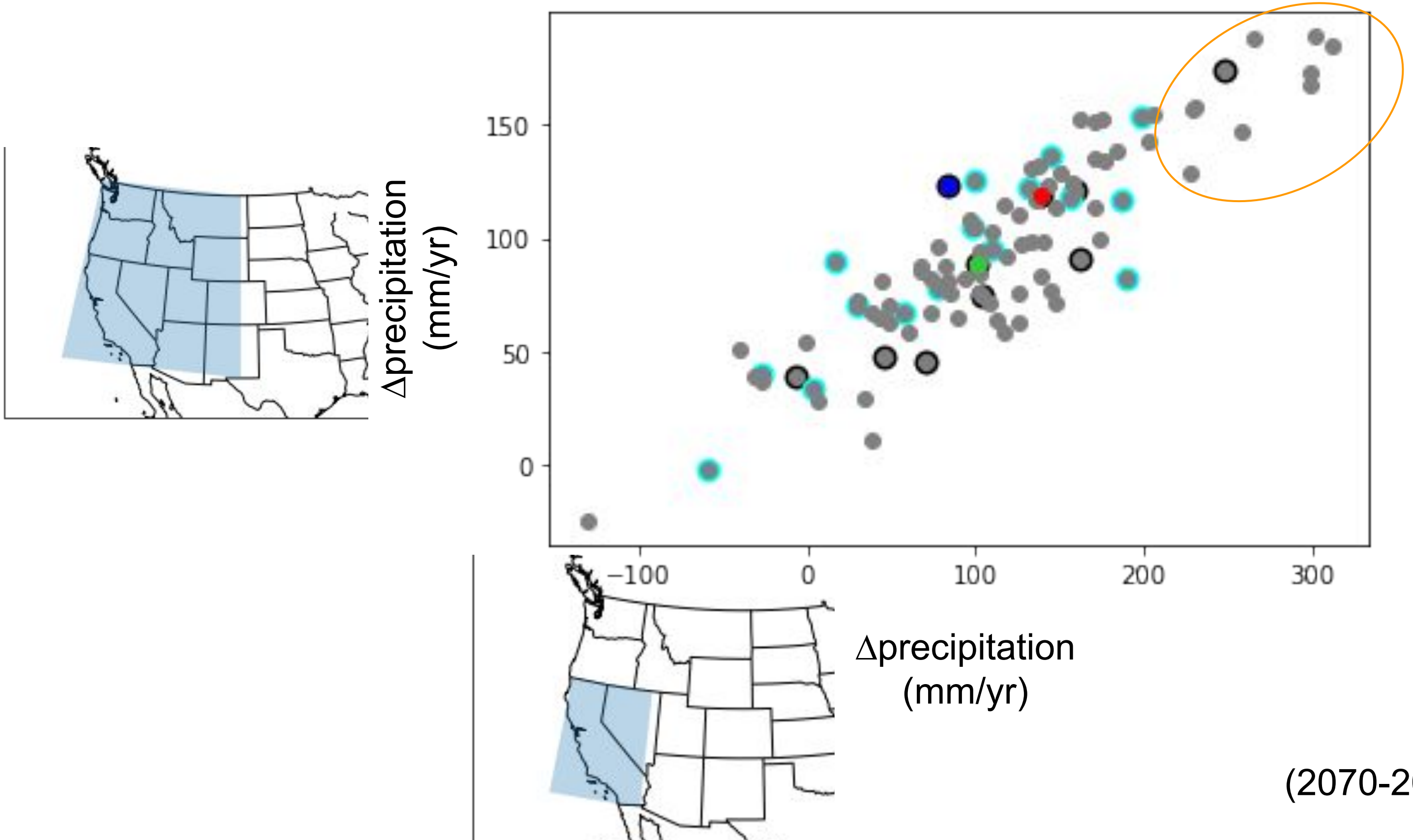






Dong et al., 2021

Not meaningful for interpretation, even if something similar could be seen to be statistically significant



(2070-2099 minus 1950-1979)

Summary

Objective method shared for choosing a subset of large ensemble members to downscale dynamically.

The down-scalable subset of the CESM2-LE does not adequately span tropical trends in the larger LE.

Including large-scale factors, like tropical variability, might not have the expected influence on regions like the western US if teleconnections are not properly represented

but could still be helpful if a different region were being downscaled.

Extra Slides

Downscalable members:

['LE2-1111.006', 'LE2-1051.003', 'LE2-1091.005', 'LE2-1071.004', 'LE2-1151.008', 'LE2-1131.007', 'LE2-1011.001', 'LE2-1171.009', 'LE2-1031.002', 'LE2-1191.010']

ENSO variability can be quite different for long periods even in one long-control run of the same GCM

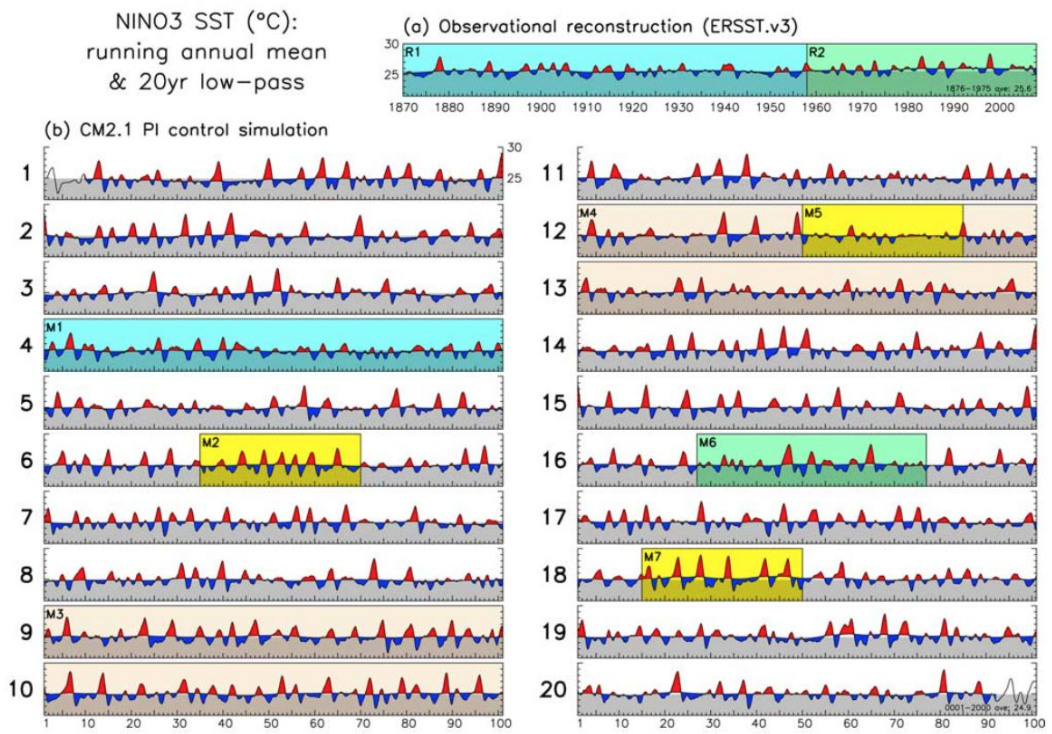
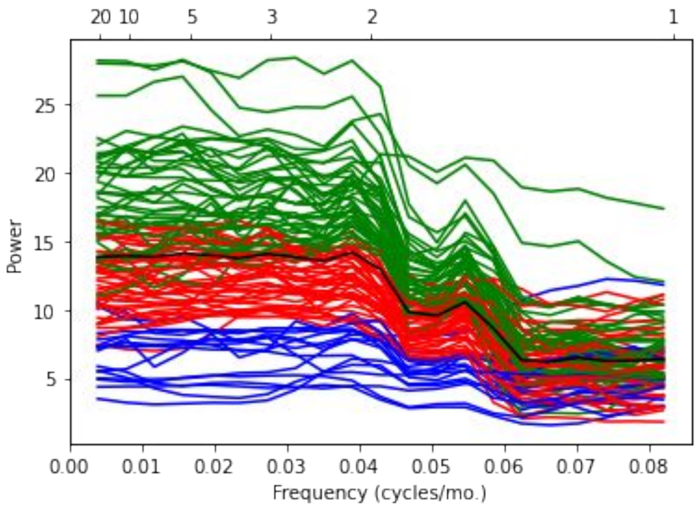
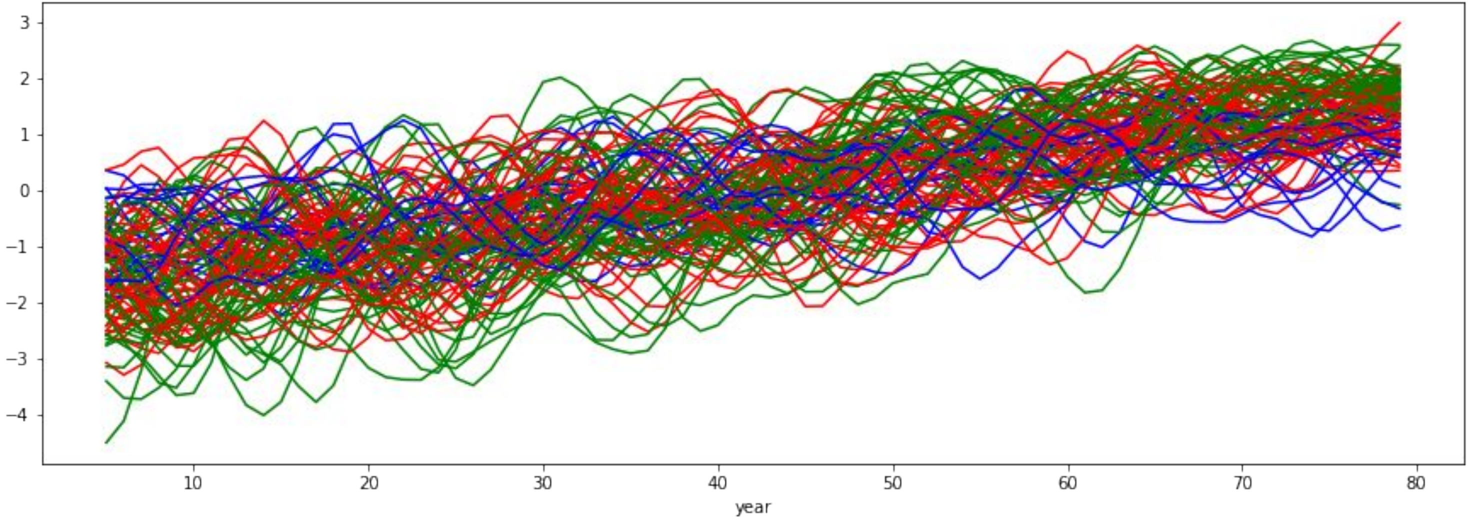


Figure 1. SST (°C) averaged over the NINO3 region (150°W–90°W, 5°S–5°N), for (a) the ERSST.v3 historical reconstruction of *Smith et al.* [2008], and (b) the 20 consecutive centuries (numbered) from the CM2.1 pre-industrial control run. Red/blue shading highlights departures of the running annual-mean SST from the multidecadal background state, where the latter is obtained via a 211-month triangle smoother which transmits (25, 50, 75)% of the time series amplitude at periods of (15, 20, 30) yr. Unshaded time series ends in Figure 1b indicate the half-width of the triangle smoother; ends of the observed time series in Figure 1a are zero-padded prior to smoothing. The top of the gray bar is the long-term mean, indicated at the bottom right of each plot. Labeled epochs are discussed in the text.

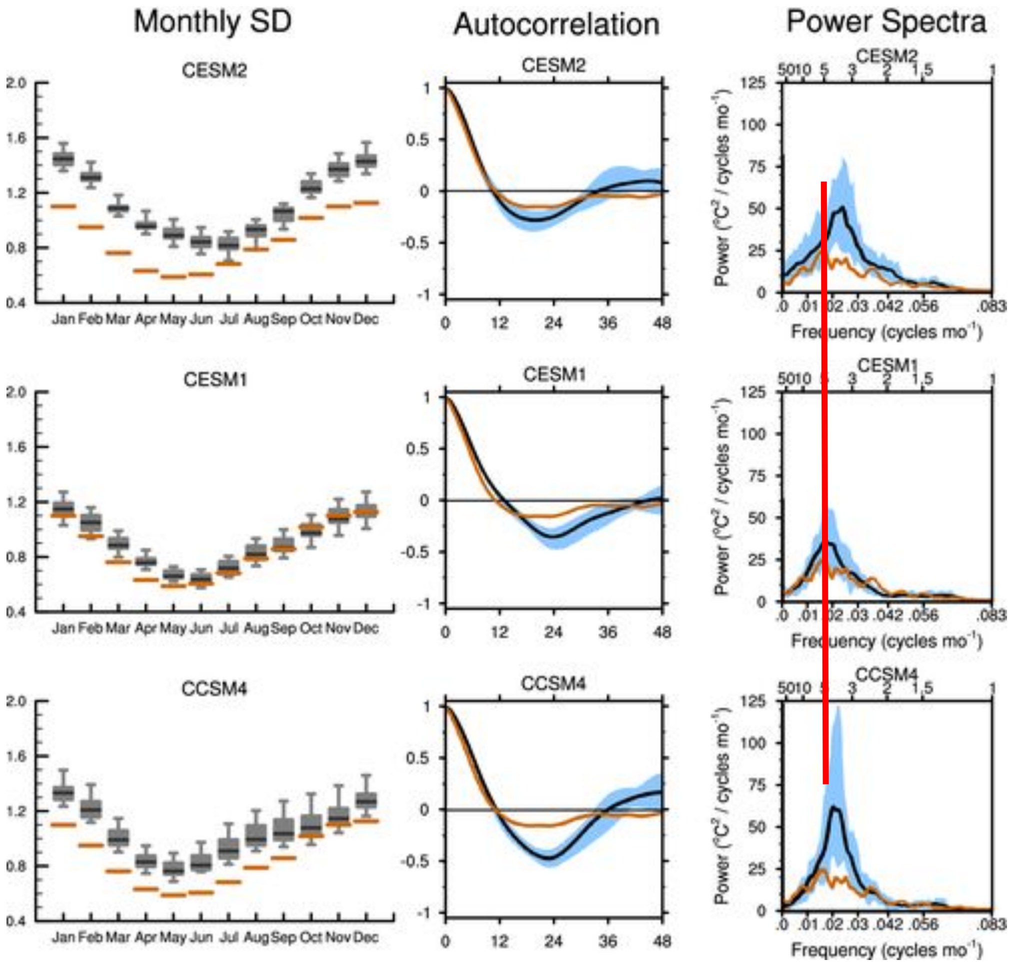
Wittenberg et al. (2009)

2. How strong a trend to a more El Niño-like future?



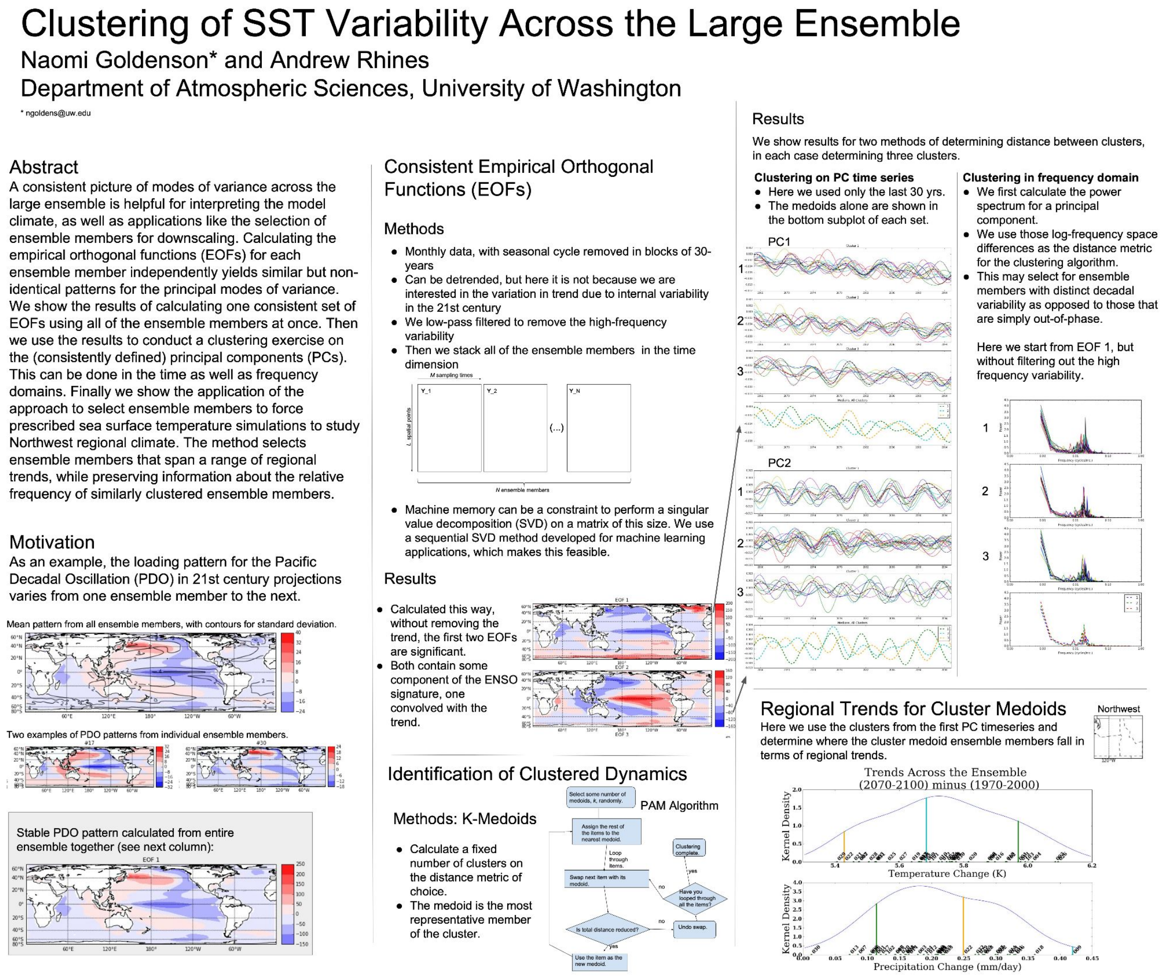


Note: none of the 10 downscalable ensemble members fall in the blue cluster. So we'll require that we have at least one from each of the other two clusters for this measure.



Capotondi et al (2020), Figure 3

This 2016 CVC Working Group poster introduced clustering across PCs, which forms one of the metrics described for distinguishing ensemble members.



Clustering of SST Variability Across the Large Ensemble

Naomi Goldenson* and Andrew Rhines
Department of Atmospheric Sciences, University of Washington
* ngoldens@uw.edu

Abstract

A consistent picture of modes of variance across the large ensemble is helpful for interpreting the model climate, as well as applications like the selection of ensemble members for downscaling. Calculating the empirical orthogonal functions (EOFs) for each ensemble member independently yields similar but non-identical patterns for the principal modes of variance. We show the results of calculating one consistent set of EOFs using all of the ensemble members at once. Then we use the results to conduct a clustering exercise on the (consistently defined) principal components (PCs). This can be done in the time as well as frequency domains. Finally we show the application of the approach to select ensemble members to force prescribed sea surface temperature simulations to study Northwest regional climate. The method selects ensemble members that span a range of regional trends, while preserving information about the relative frequency of similarly clustered ensemble members.

Motivation

As an example, the loading pattern for the Pacific Decadal Oscillation (PDO) in 21st century projections varies from one ensemble member to the next.

Mean pattern from all ensemble members, with contours for standard deviation.

Two examples of PDO patterns from individual ensemble members.

Stable PDO pattern calculated from entire ensemble together (see next column):

Consistent Empirical Orthogonal Functions (EOFs)

Methods

- Monthly data, with seasonal cycle removed in blocks of 30-years
- Can be detrended, but here it is not because we are interested in the variation in trend due to internal variability in the 21st century
- We low-pass filtered to remove the high-frequency variability
- Then we stack all of the ensemble members in the time dimension
- Machine memory can be a constraint to perform a singular value decomposition (SVD) on a matrix of this size. We use a sequential SVD method developed for machine learning applications, which makes this feasible.

Results

- Calculated this way, without removing the trend, the first two EOFs are significant.
- Both contain some component of the ENSO signature, one convolved with the trend.

Identification of Clustered Dynamics

Methods: K-Medoids

- Calculate a fixed number of clusters on the distance metric of choice.
- The medoid is the most representative member of the cluster.

PAM Algorithm

Results

We show results for two methods of determining distance between clusters, in each case determining three clusters.

Clustering on PC time series

- Here we used only the last 30 yrs.
- The medoids alone are shown in the bottom subplot of each set.

Clustering in frequency domain

- We first calculate the power spectrum for a principal component.
- We use those log-frequency space differences as the distance metric for the clustering algorithm.
- This may select for ensemble members with distinct decadal variability as opposed to those that are simply out-of-phase.

Here we start from EOF 1, but without filtering out the high frequency variability.

Regional Trends for Cluster Medoids

Here we use the clusters from the first PC timeseries and determine where the cluster medoid ensemble members fall in terms of regional trends.

Northwest

Trends Across the Ensemble (2070-2100) minus (1970-2000)