

Interpreting Correlation and Multiple Linear Regression Analyses of Solar Cycle Impacts

A. K. Smith, NCAR

Z. K. Mullen, Univ. of Colorado

basics

- Variability of the sun can affect variability in the atmosphere.
- Finding the part of the atmospheric variability that is due to solar variability is straightforward in some cases (upper atmosphere, upper part of middle atmosphere)
- In other cases, identifying the solar response is not simple.

Sources of my concern with attribution analysis and what we can do to address it

Observational data record is short.

- *Use comprehensive interactive numerical models to explore cause and effect pathways.*

There is mounting evidence from numerical models that analysis finds decadal or solar response signals that are actually the result of other forced or random variability.

- *Use diagnostic studies with synthetic (model) data to characterize and understand the misleading signals to aid in better interpretation of observations.*

Our scientific community is motivated to find and quantify any/all factors affecting climate change.

- *We should ensure that the analysis we use is suitable to finding a small quasi-periodic signal in a complex system.*

Evidence from numerical models for other variability mapping into solar response

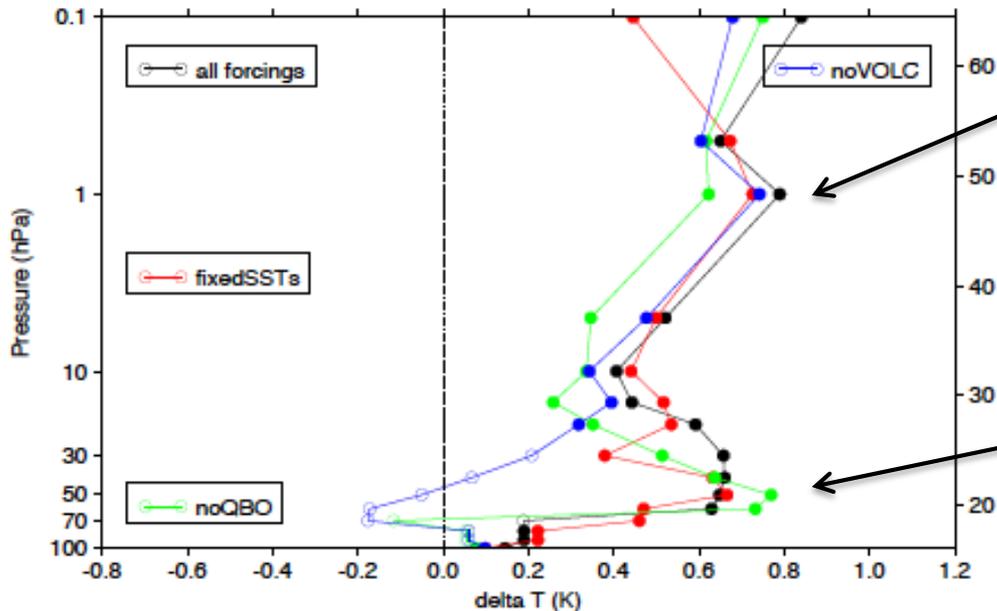
- Decadal timescales:
 - **Volcanoes**: Aerosol from major eruptions directly affects radiative balance in lower stratosphere.
(El Chichon in 1982; Mt. Pinatubo in 1991)
 - **PDO**: Pacific Decadal Oscillation is an atmosphere-ocean cycle that affects temperatures in the north Pacific and elsewhere.
- Shorter timescales:
 - **QBO**: Quasi-biennial Oscillation affects temperature and wind in the tropical and winter middle atmosphere.
 - **ENSO**: El Niño/Southern Oscillation is an atmosphere-ocean cycle that affects temperatures in the tropical Pacific and NH winter.
 - **annual cycle**: pretty much everything in the atmosphere depends on this, often not linearly
- Unknown/random
 - large (unforced?) interannual variability in winter extratropics

Aerosol from major volcanic eruptions

Solar signal (multiple linear regression with improved technique) in tropical stratosphere from WACCM simulations with other forcing turned on/off

Other forcing includes QBO, ENSO, and volcanic aerosol.

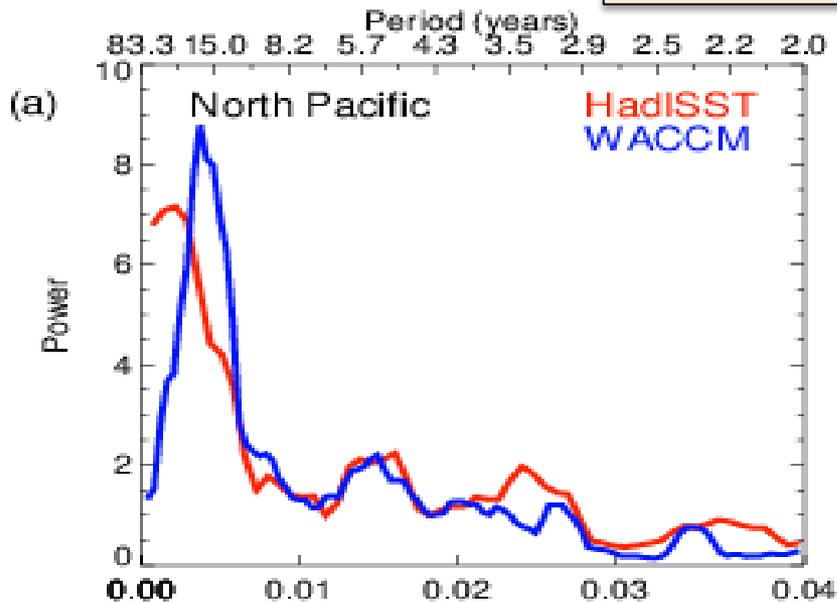
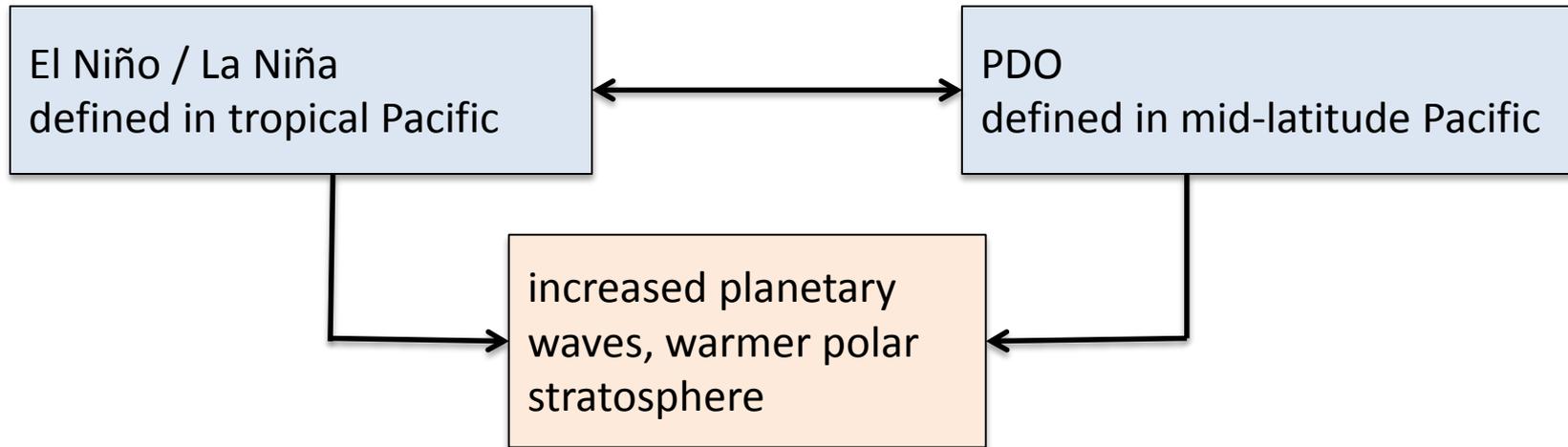
Solar signal in zm T [25N-25S]



Positive signal in tropical upper stratosphere – does not vary when other (non-solar) forcing is omitted.

Positive signal in tropical lower stratosphere is not significant when volcanic aerosol is omitted.

Interannual variability from coupled atmosphere-ocean (ENSO and PDO)

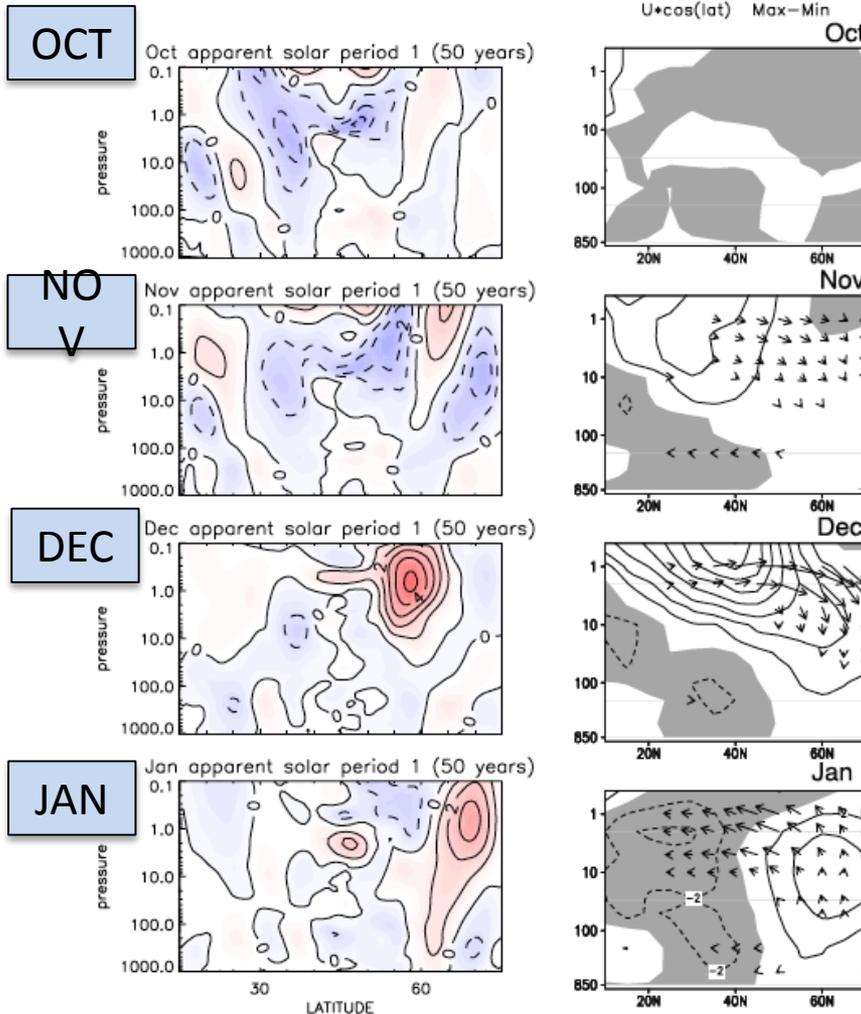


Kren et al., J. Clim., under review

Analysis of 200-year WACCM simulation with coupled atmosphere-ocean that generates these cycles but does not include other non-random sources of interannual variation:

- solar UV variation
- QBO
- greenhouse gas increase
- ozone-depleting chemicals

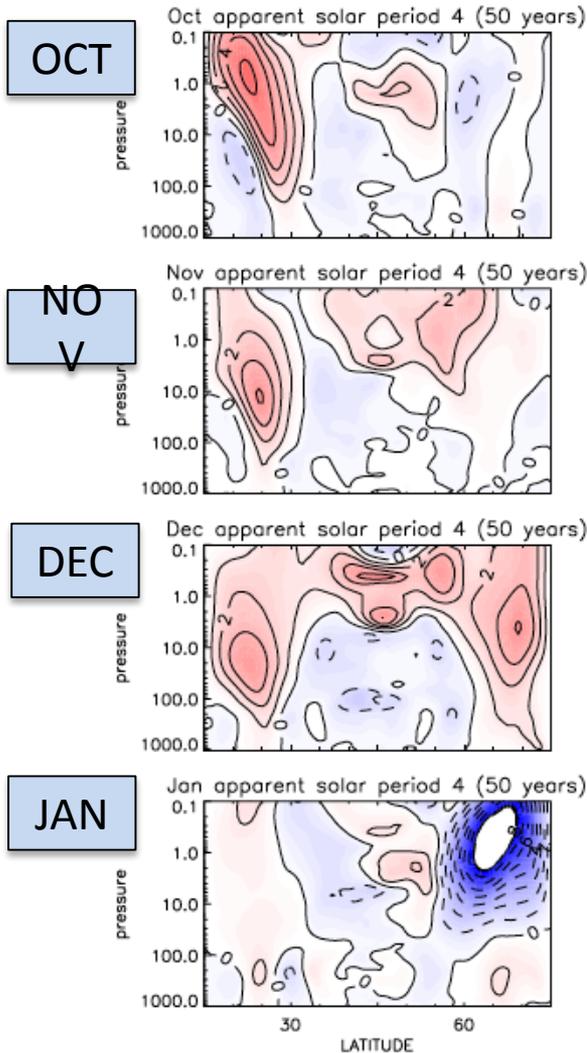
“solar” response in WACCM simulations without QBO, solar cycle, or anthropogenic trends



Analysis of 50-year period of WACCM $u \cdot \cos(\text{latitude})$ for comparison with obs.

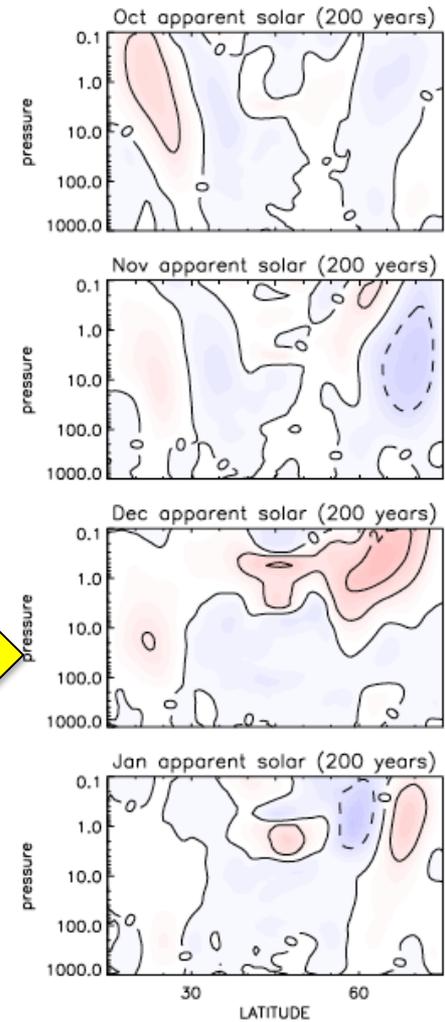
➤ NO actual solar variability

“solar” response in WACCM simulations without QBO, solar cycle, or anthropogenic trends



Another WACCM 50-year period with same analysis gives results that do not look like the Kodera & Kuroda obs.

The apparent solar signal persists even for a period of 200 years.



Where do the misleading “solar” responses originate?

- from the fact that the atmosphere has intrinsic or externally forced variability at frequencies that overlap the solar variability?

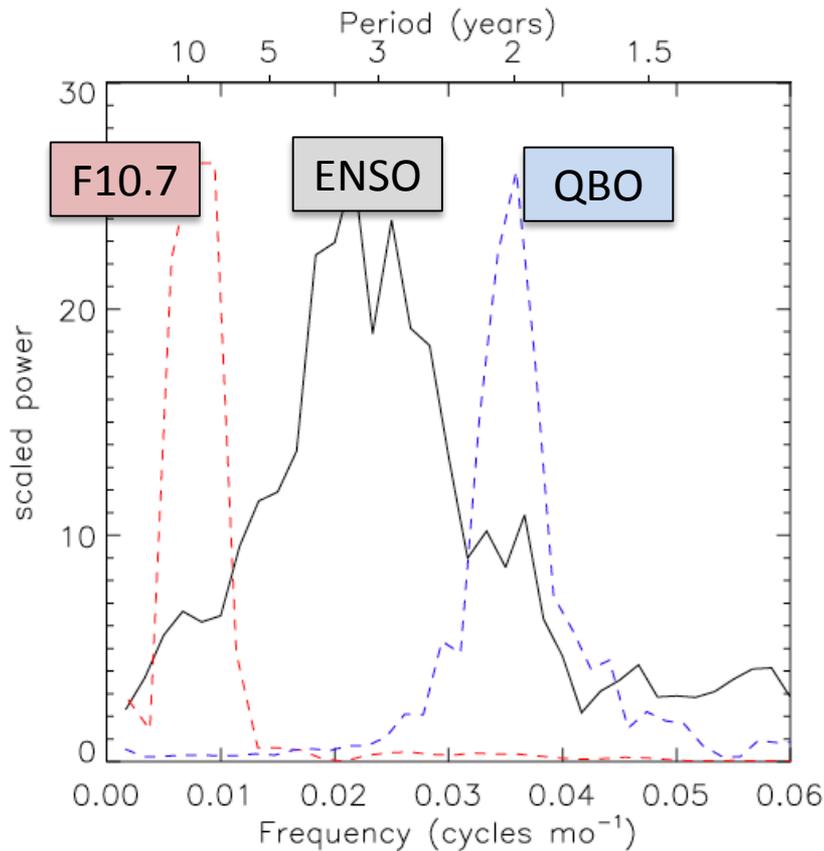
and/or

- from analysis methods that do not adequately separate solar from other forcing?

Can we use information from model analyses to identify and remove misleading signals from observational analysis?

1. The atmosphere has intrinsic or externally forced variability at frequencies that overlap the solar variability.

Intrinsic variability: overlap in periods of “external” forcing



- QBO (Berlin, since 1953)
- F10.7
- ENSO (Niño 3.4 index from WACCM)

- Minimal apparent overlap in periods of QBO and F10.7 indices
- Period of ENSO index has appreciable overlap with both solar and QBO

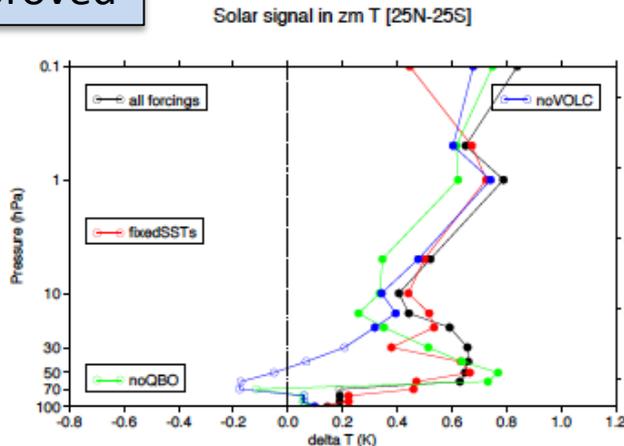
2. Analysis methods do not adequately separate solar from other forcing.

Refine multiple linear regression technique

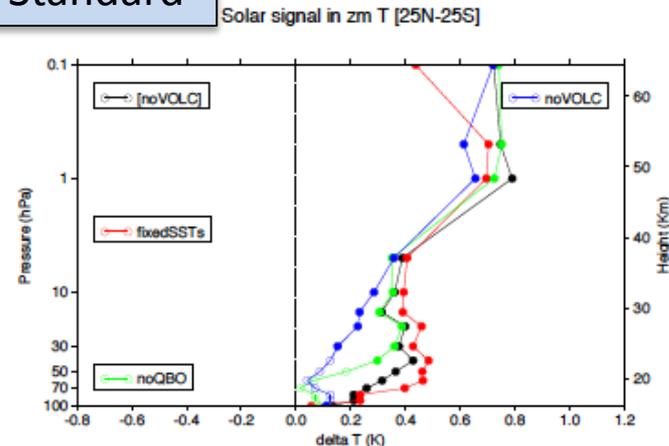
Chiodo et al. (ACP, 2014) used an improved method that:

- removes auto-correlation by pre-whitening
- calculate lags that maximize correlation between field (i.e. temperature) and predictor index
- optimum lags do not bring forcing predictors into phase

Improved



Standard



Improved MLR is still not able to separate volcanic from solar forcing in the tropical lower stratosphere.

The methods give similar results in the upper stratosphere.

A. Maybe the “model” is too simple

$$DT = b_0 + b_1t + b_2Q + b_3V + b_4S + b_5E$$

T = deseasonalized temperature

t = time (trend term)

Q = QBO (often there are two QBO terms)

V = volcanic aerosol

S = solar variability

E = ENSO

The first recommendation from statistics is to include everything we know about interannual variability in the model.

Examples of missing sources of variability

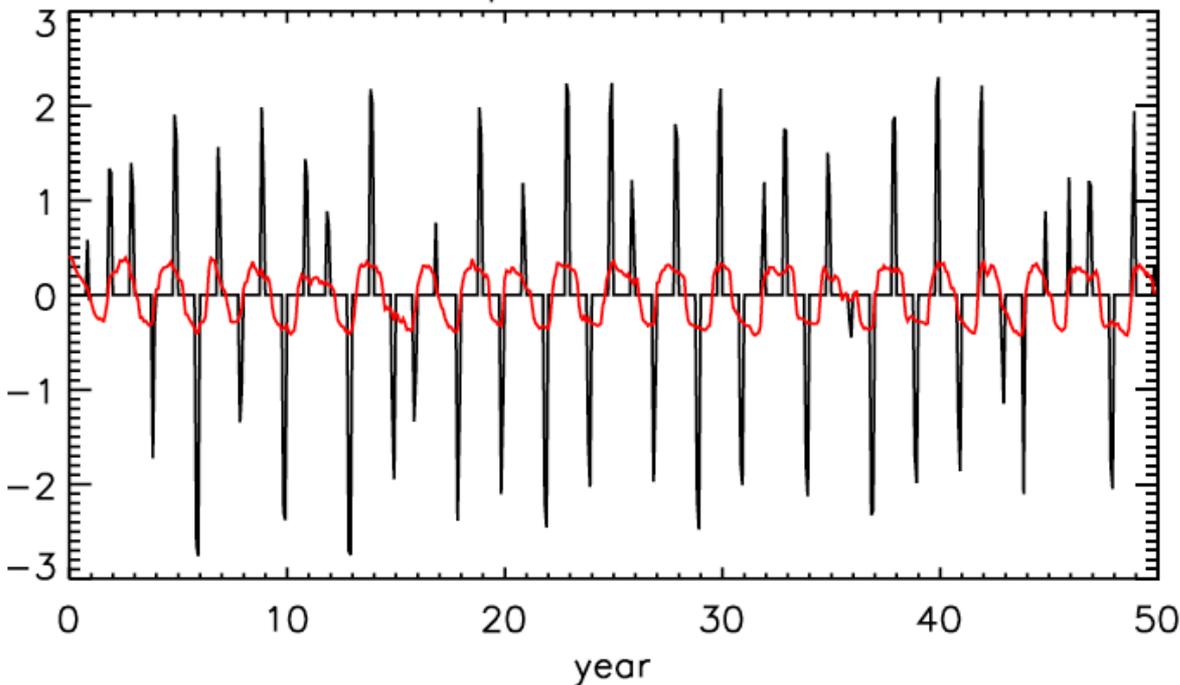
- Response to the QBO and ENSO in mid/high latitudes depends on annual cycle.
- QBO interacts with volcanic aerosol.
- ENSO frequency and phase vary with PDO.
- Trend in ODS (ozone depleting substances) is not linear.
- Lag in volcanic response varies with latitude and hemisphere.

Example: linear regression when atmospheric response depends on annual cycle

idealized case – looking at a single gridpoint

- only external forcing is QBO (based on Berlin data at 30 hPa)
- QBO index is exactly known
- response of the atmosphere to QBO is exactly known
- QBO perturbs timeseries only in Nov-Dec (idealized Holton-Tan effect)
- no other forcing or variability

response to QBO



black: deseasonalized
timeseries

red: reconstructed
from QBO fit

Linear regression
does not capture variability
driven by QBO

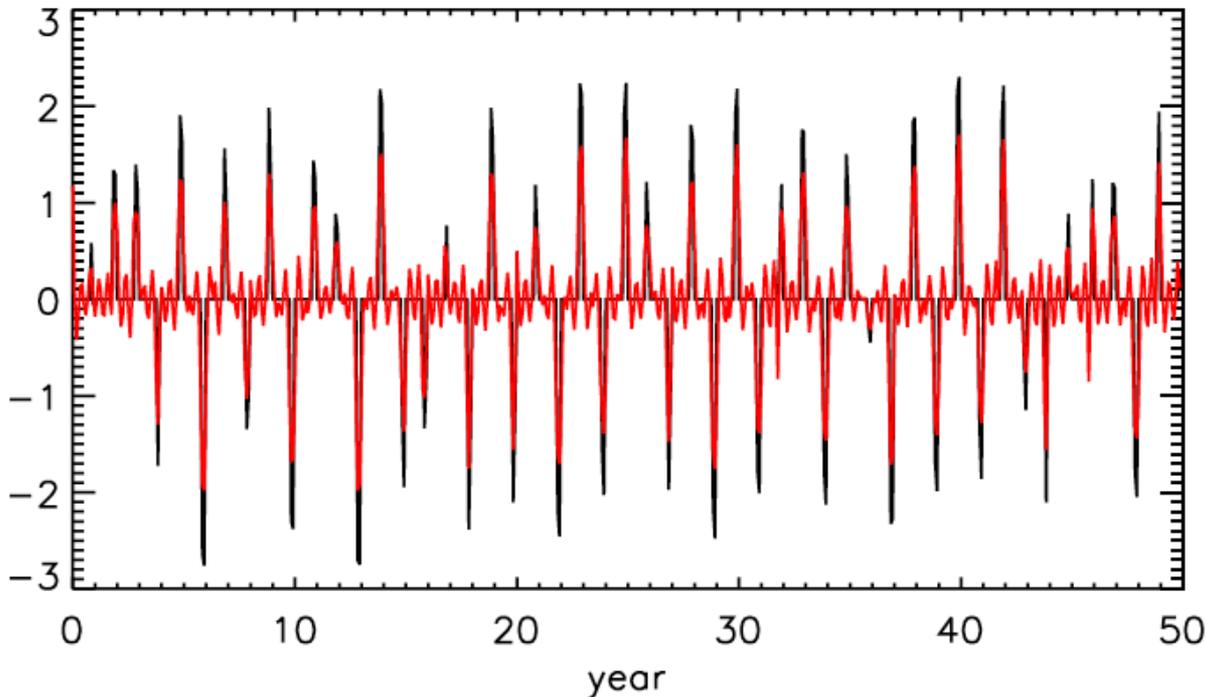
Multiple linear regression using method of Randel and Cobb (JGR, 1994)

This case uses five coefficients for the QBO:

- basic
- two for sinusoidal annual cycle
- two for sinusoidal semiannual cycle

Idealized Holton-Tan response, no added variance

annual, no random var



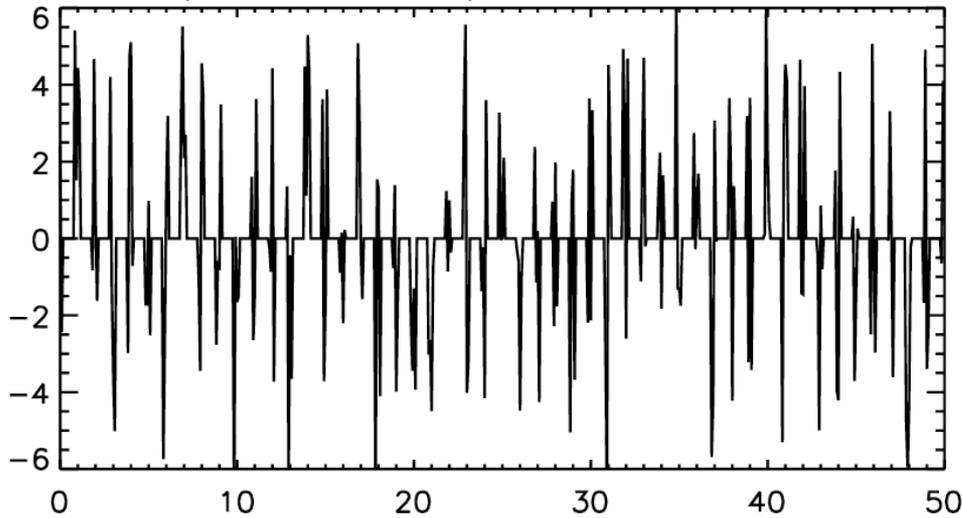
black: deseasonalized
timeseries
(same as before)

red: reconstructed
from QBO fit

MLR method of R&C
does better at capturing
variability driven by QBO

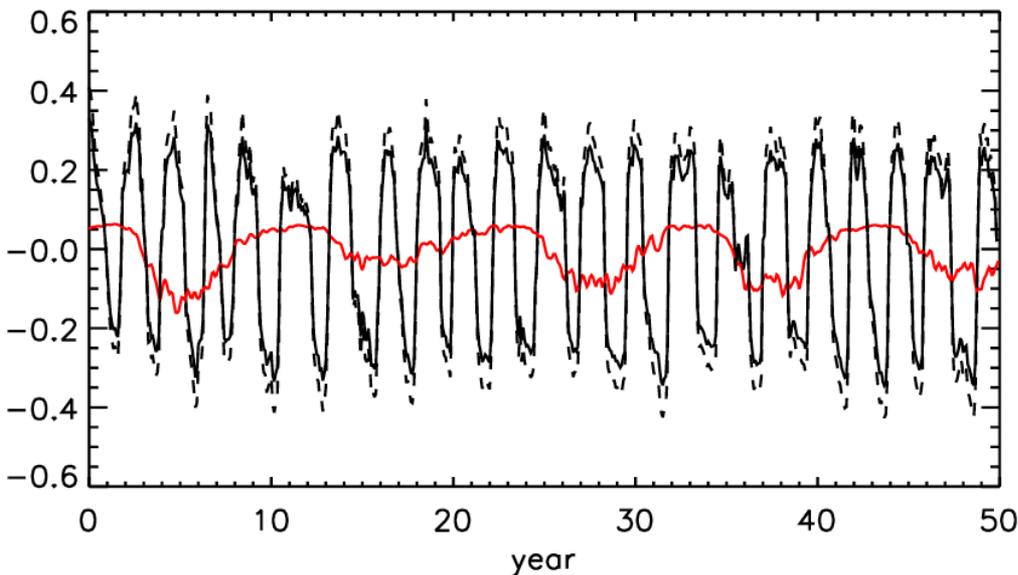
Now consider that there is added random variance in winter months (Nov – Feb) and two predictors

response to QBO plus random variance



deseasonalized timeseries
(as before except added variance)

reconstructed from fit to QBO and "solar"

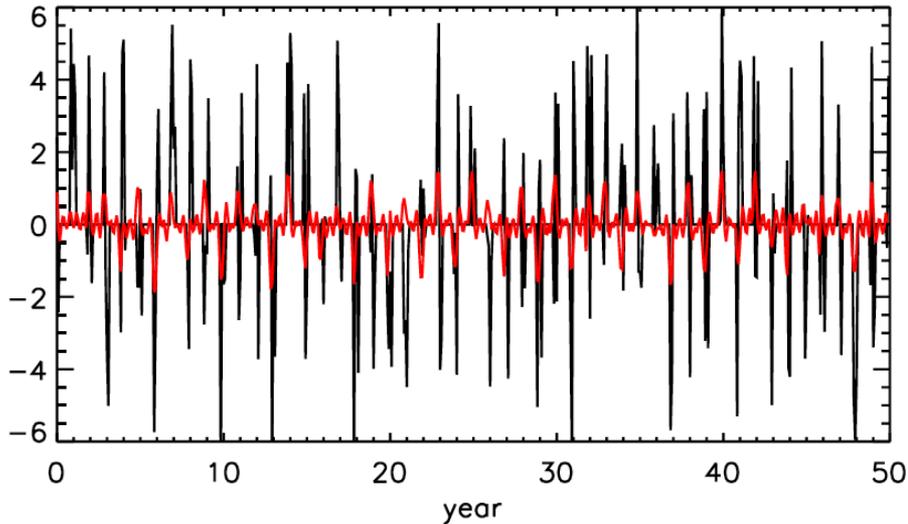


fit to QBO
fit to F10.7

Some of variance projects onto solar predictor (F10.7) even though atmosphere is not affected by solar.

Compare the “solar” response from 2 methods

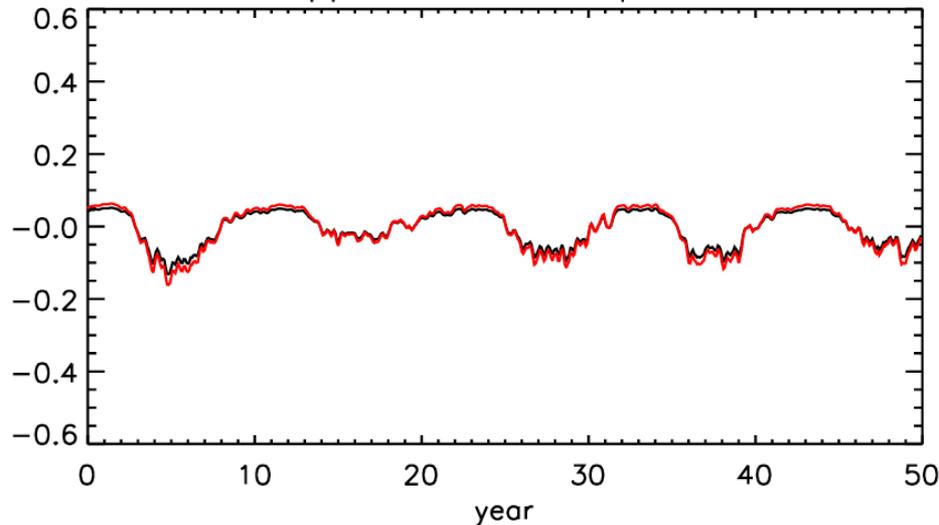
QBO annual, with random var



method of Randel & Cobb still isolates the QBO response

-> QBO signal is almost the same as case with no added variance)

apparent solar response



... but does not improve misleading “solar”

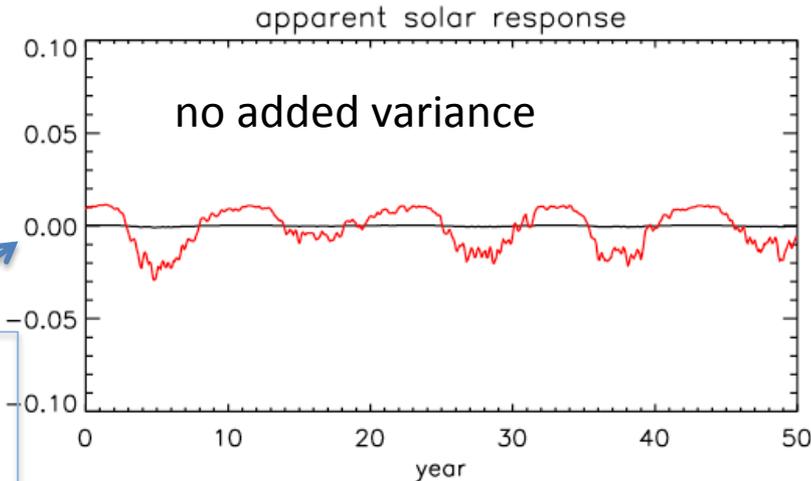
solar fit using conventional method
(2 regression indices: QBO and solar)

solar fit using method of Randel & Cobb

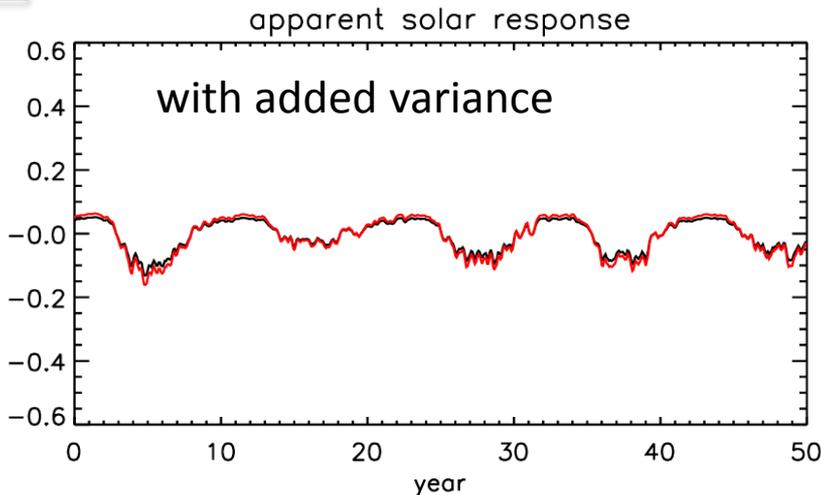
Add even more regression predictors

solar fit using conventional method

solar fit using additional indices



Note
different
scale



INDICES:

- seasonal cycle (4 predictors)
- QBO including annual cycle (5 predictors)
- QBO squared
- solar

Misleading response does not disappear when signal includes random variance.

B. Use care in the interpretation of significance

- T-test on solar term measures the difference in the residual from regression with all predictors included (the “full” model) and regression with the same model but with the solar predictor omitted.
- If there are only a few predictors (say trend, QBO, volcanic aerosol, ENSO), then it is likely that any added predictor in the regression analysis (trend, QBO, volcanic aerosol, ENSO **plus solar**) will account for some “significant” extra variance.
- The above is particularly true if more than one of the predictors is periodic.

simpler model -> more likely that an added predictor will
have a significant response

BACK TO THE QUESTION: Where do the misleading “solar” responses originate?

- ✓ from the fact that the atmosphere has intrinsic variability at frequencies that overlap the solar variability
- and*
- ✓ from analysis methods that do not adequately separate solar from other forcing

Can we use information from model analyses to identify and remove misleading signals from observational analysis?

- ✓ Identify: yes
remove: doesn't look promising

Conclusions

- There is a growing body of evidence from numerical simulations that linear regression analysis finds signals in the atmosphere that are not what they seem.
- This is at least in part the result of overlapping spectra of known externally forced variability.
- Data from model simulations can be used to refine analysis techniques to better isolate the response to weak forcing.
- The first step is to develop analysis techniques that include more of the information that we know about atmospheric variability.