Our shared responsibility as users of statistics and consumers of results

Caveat Venditor, Lector et Emptor (Seller, Reader and Buyer Beware)

13IMSC, 9 June 2016

Francis Zwiers
PCIC, University of Victoria

Photo: F. Zwiers

Outline

- Introduction
- Moving Windows
- EOF analysis
- Extremes
- Detection and Attribution
- Discussion

Photo: F. Zwiers

Introduction

Thanks to Xuebin Zhang

Photo: F. Zwiers

The strange alternative title of this talk …

- It tries to say that everyone bears responsibility in the use of statistical information (producer, user, and casual bystander)
- An important part of that responsibility is to understand the framework within which statistical information is produced
- That framework, including the assumptions that are made, represents a model that provides context for the interpretation of the variability in the data to be analyzed
- The framework does not need to be correct to be useful, but useful interpretation does require that it be understood, and that there is an appreciation of the degree of approximation to the real world that it entails

The strange title …

- The assumptions are often not well understood, stated explicitly or discussed, and their importance is often not appreciated by users and passive observers.
- The concern is that this can result in the over interpretation of statistical findings
- Consider four examples – two where there is a reasonable appreciation of statistical limitations, and two where the statistical foundation seems to be less appreciated


1. Moving window analyses

Methods

- Come in many flavours
- Sliding window correlation is often used to study the stability of links between snow cover and the Indian monsoon

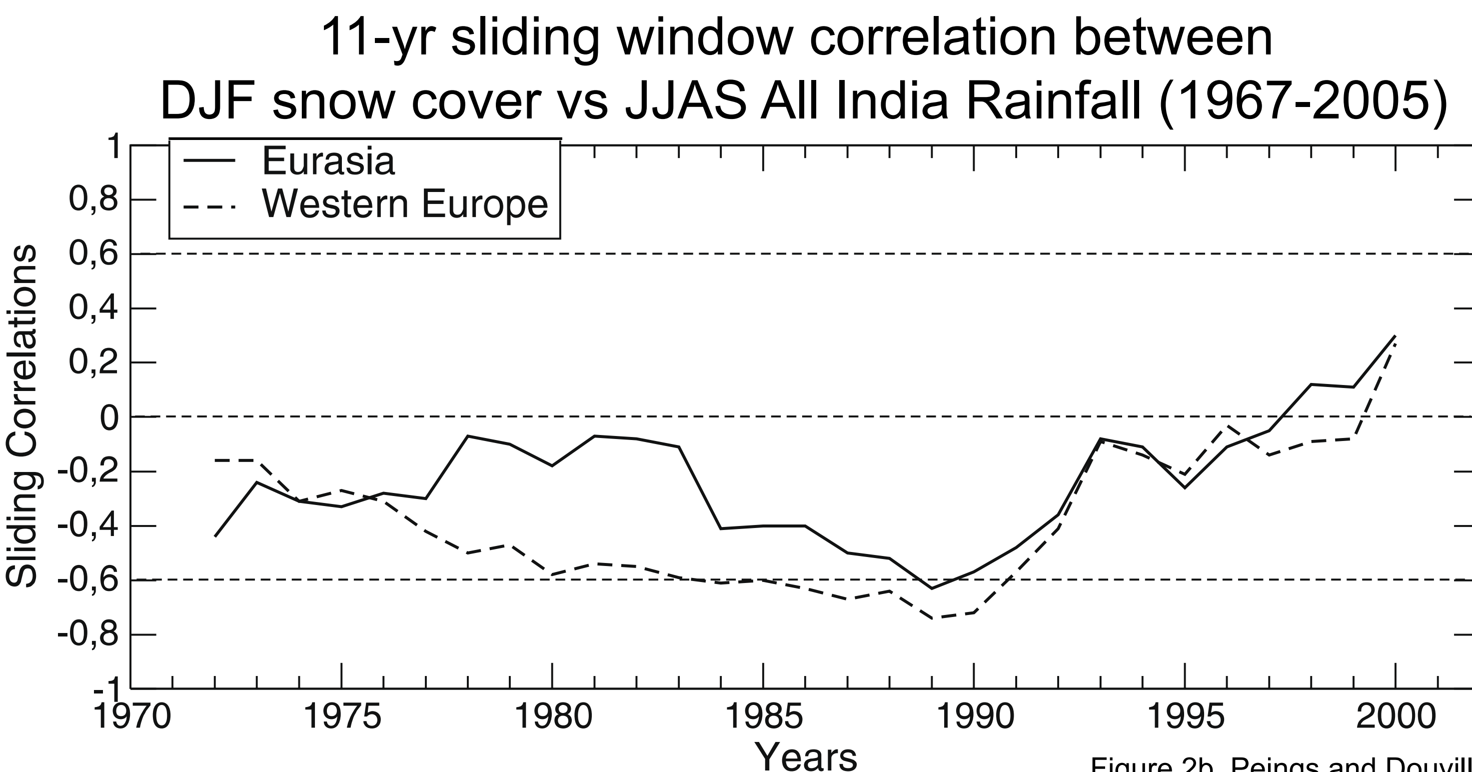


Figure 2b, Peings and Douville, 2009

Some concerns …

- Dependence between individual correlations
- Low power because of small sample size
- Multiple testing
- Physical support for the interpretation of fluctuating correlation

2. Modes of variability

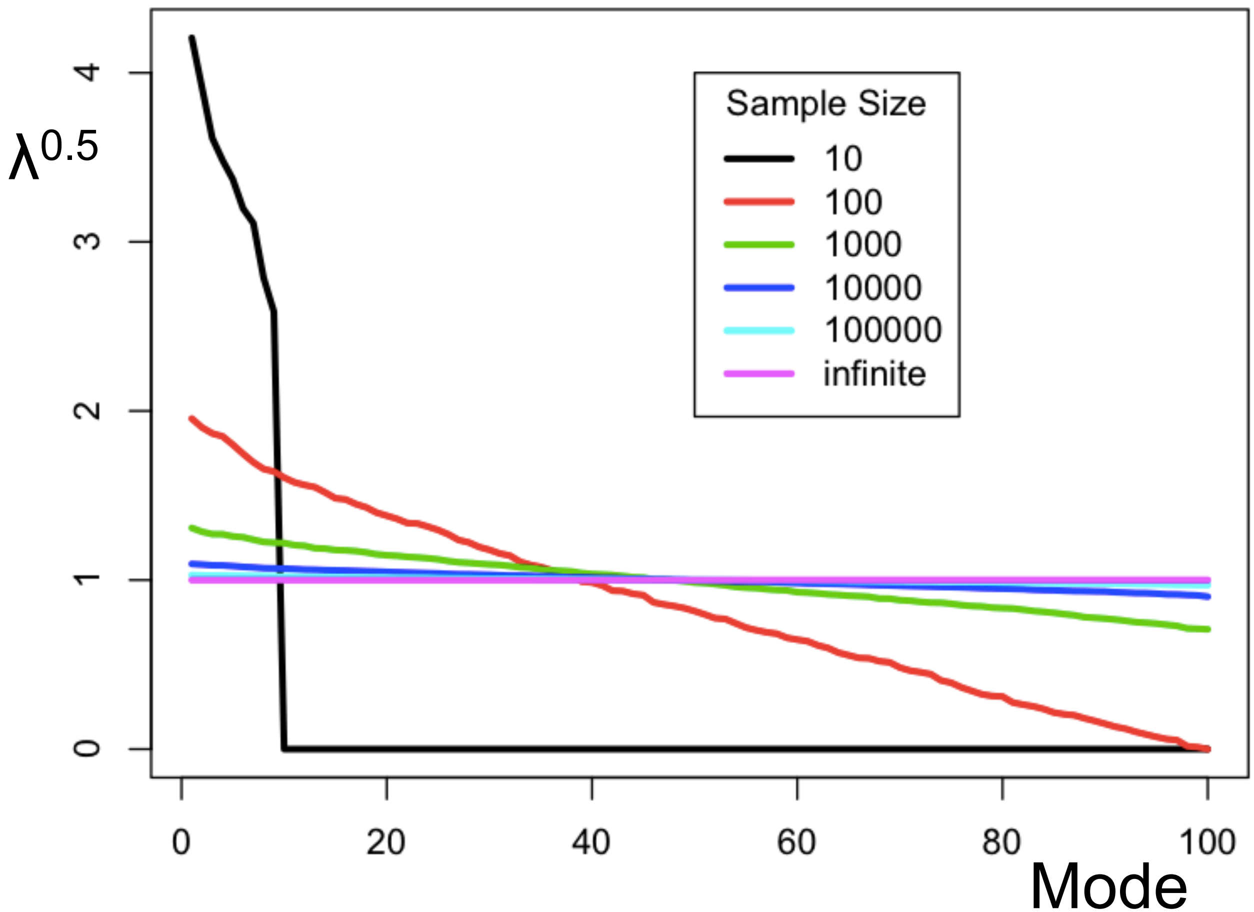
Photo: F. Zwiers

Methods

- Often identified using EOF analysis
- Typical configuration still generally has temporal dimension that is smaller than spatial dimension
- A frequent question is whether modes are *mixed*
  – Modes that represent similar amounts of variance are mixed (any set of orthonormal vectors that span the same subspace are equally plausible)
- North’s 1982 “rule of thumb” provides some guidance through an approximation $\delta\lambda \sim \lambda(2/n)^{1/2}$ of the sampling error of eigenvalue estimates

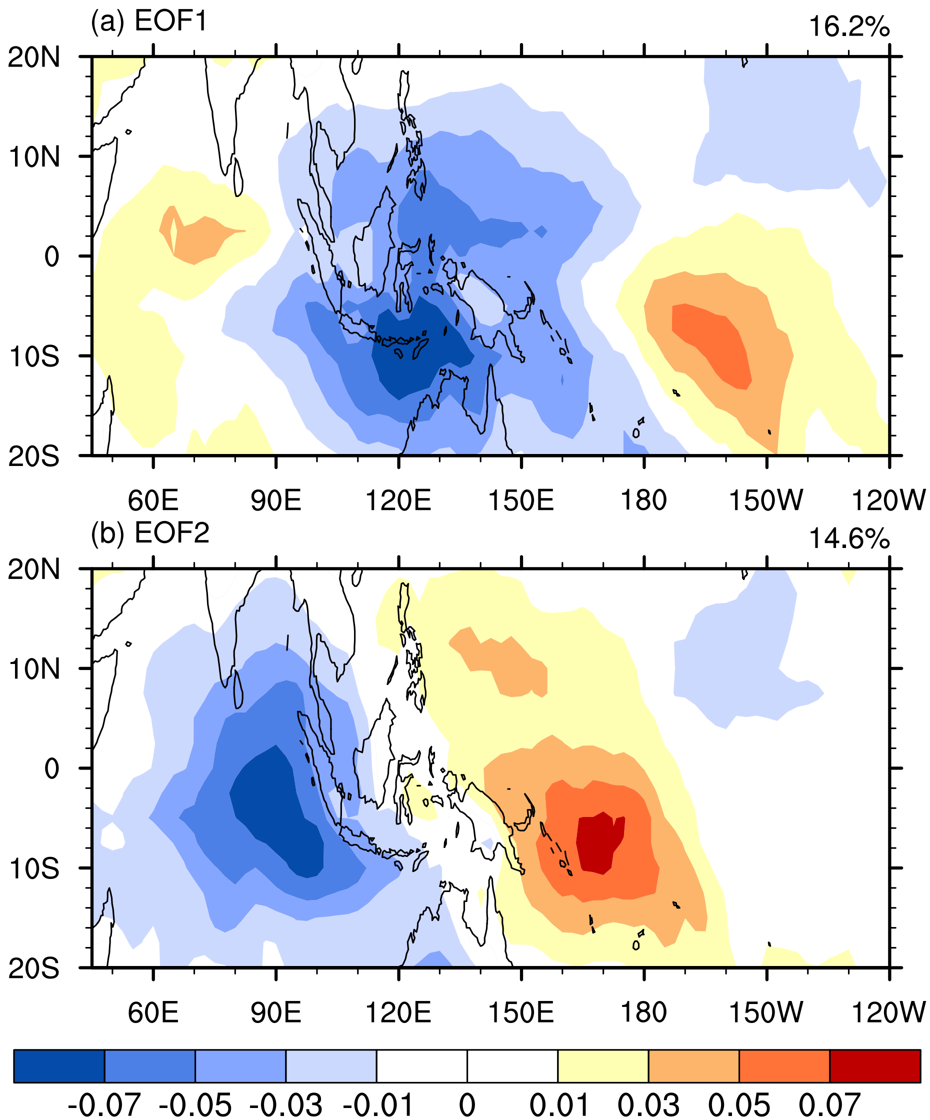
Eigenspectrum behavior

Eigen values of sample covariance matrix estimated from samples of 100-dimensional N(0,I) random vectors



Total variance is conserved, and thus for small n, eigenvalues are biased and EOFs are necessarily distorted

An example - MJO



Sperber, 2003; IPCC AR5 WG1 Fig 9.31

- Based on 7-years of data (selected based on tropical 200-hPa zonal mean wind variability), 101 days of 20-100 day band-pass-filtered AVHRR OLR data per year
- Analyzed fields have ~1250 grid cells (2.5°x2.5° lat-long)
- Modes in quadrature
- Explained variance likely over-estimated
- Lots of spatial noise
- Used to assess climate models (Sperber and Kim, 2012, IPCC WG1 Ch. 9)

Some concerns …

- Underestimation of eigenvalue uncertainty
  - North's "rule of thumb" but does not consider temporal dependence
- Bias of eigenvalue estimates
- Corresponding aliasing of spatial variability within derived EOFs
- North et al (1982) warn that convergence of the sample covariance matrix is slow – but this is not well appreciated.
- Regularized estimators such as the Ledoit-Wolf (2004) estimator $\Sigma^* = \rho_1 I + \rho_2 S$ are demonstrably "more accurate", but impact on modes not studied

3. Extremes

Photo: F. Zwiers

See WCRP summer school on extremes, ICTP, July, 2014

Methods

- Determined by sampling approach
  - Block maximum (e.g., the annual maximum)
  - Peaks over threshold

Usual assumptions (block maximum)

- Block maxima are iid (or iid after accounting for dependence on covariates)
- Blocks are long enough to ensure that the GEV approximates the distribution of extremes well

Spatial dependence

- Several methods available, but they are not in general use in climatology

Precipitation extremes

Photo: F. Zwiers (Longji)

Precipitation extremes

- Observational studies suggest intensification is occurring, although local detection is very hard (eg., Westra et al, 2013)
- Expectation of intensification is supported by
  - attribution of warming (eg, Bindoff et al, 2013),
  - attribution of observed increase in atmospheric water vapour content (eg, Santer et al, 2007), and
  - D&A studies of change in mean precipitation (eg., Zhang et al., 2007; Noake et al., 2012; Polson et al, 2013; Marvel and Bonfils, 2013; Wu et al, 2013) and surface salinity (eg., Pierce et al., 2012).
- D&A studies on extreme precipitation are very limited (eg, Min et al 2011, Zhang et al, 2013)

D&A on transformed extremes

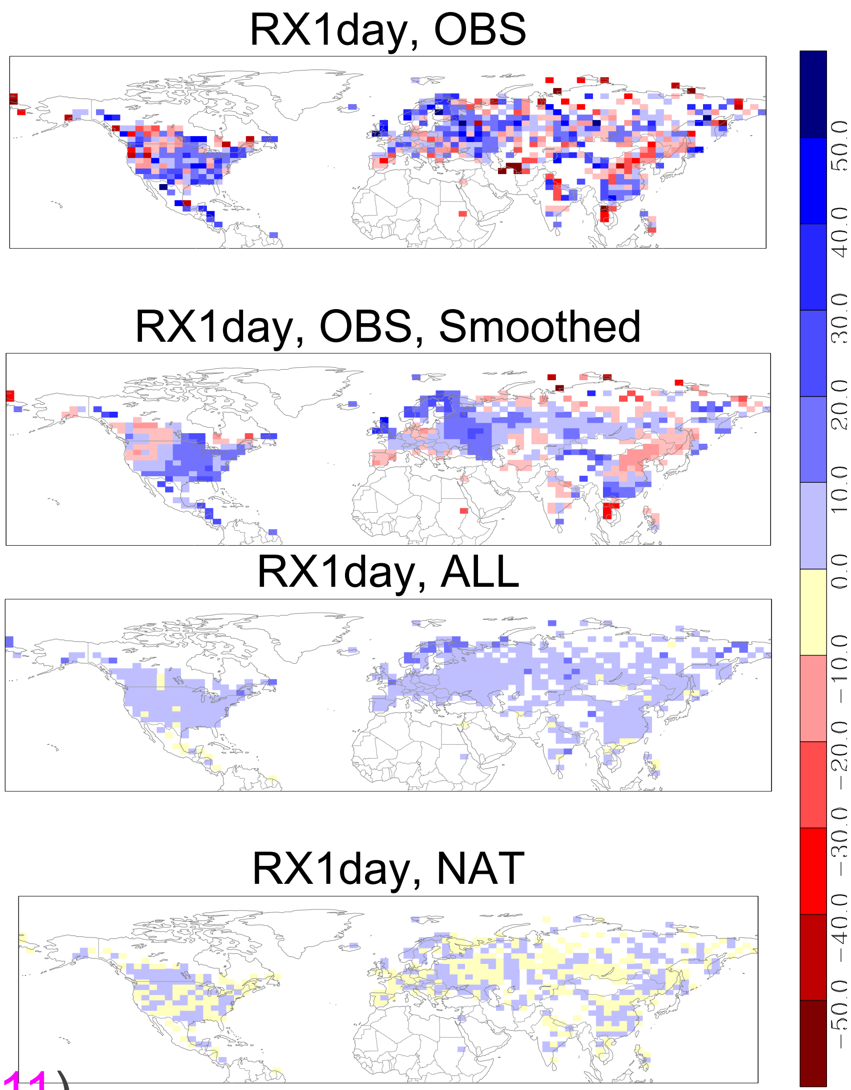
- Transform to a probability index
  - Fit an extreme value distribution locally
  - Apply probability integral transform
  - Transformed values have approximately the uniform distribution
  - Time and area averaging produces Gaussian values
- Apply standard D&A paradigm
- Examples include
  - Min et al 2011, 2013, Zhang et al, 2013, Kim et al 2015

Precipitation Extremes

Observed and simulated changes in a probability index of annual maximum 1-day precipitation (1951-2005)

Attributed intensification:
- 3.3% [1.1 – 5.8]%
- 5.2%/˚C [1.3 – 9.3]%/˚C

Estimated waiting time for 1950's 20-year event:
- ~15-yr in the early 2000's



Zhang et al., 2013 (see also Min et al 2011)

Some concerns (block max approach)

- Annual cycle
  - How long is the block really?
  - Does not lying within the domain of convergence weaken interpretation?
  - Does using a generic goodness of fit test (standard practice) increase confidence materially?
- Event frequency
  - How big is the block if the block size is itself stochastic (eg, non-zero precipitation events in a block)?
  - What is the impact on interpretation if the expected event frequency (ie, average block size) is also affected by forcing (Schar et al, 2016)?

Some concerns …

- Spatial scaling
  - Can we use model output at one scale to interpret changes in observations at a different scale?
- Temporal scaling
  - Do models exhibit the same temporal scaling that is seen in observations (e.g., power law behavior in precipitation extremes across different accumulation periods)?
  - Will temporal scaling change in the future?

4. Climate change detection and attribution

Photo: F. Zwiers

Methods

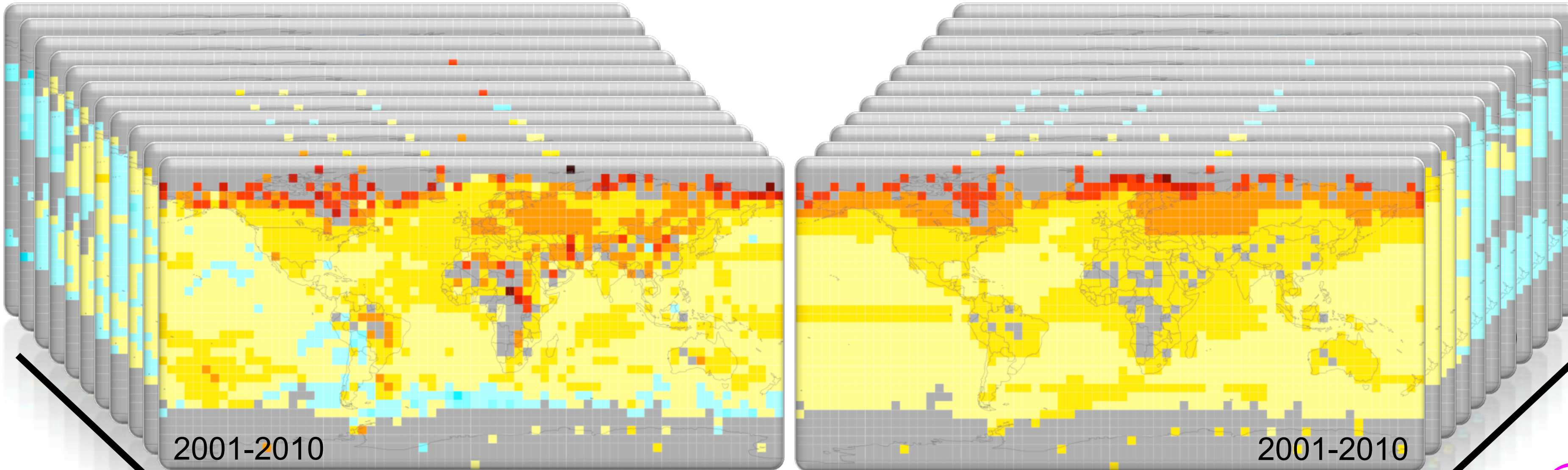
- Involve simple statistical models
- Complex implementation due to data volumes (which are both small and large)

Usual assumptions

- Key forcings have been identified
- Signals and noise are additive
- Model simulation of large-scale forcing response patterns ok, but signal amplitude is uncertain

→ leads to a regression formulation

Observations (HadCRUT4) Multi-model mean (ALL forcings)



11 decades (1901-1911 to 2001-2011)

$\mathbf{Y}$ $\mathbf{X}$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Evaluate scaling factors $\leftarrow \widehat{\boldsymbol{\beta}}$

$\hat{\boldsymbol{\varepsilon}} \rightarrow$ Evaluate residuals

After Weaver and Zwiers (2000)

That formulation has been evolving

$$\mathbf{Y} = \sum_{i=1}^{s} \boxed{\beta_i} \mathbf{X}_i + \boldsymbol{\varepsilon}$$

$$\mathbf{Y} = \mathbf{Y}^* + \boldsymbol{\varepsilon}_{\boldsymbol{y}}$$
$$\mathbf{X}_{\boldsymbol{i}} = \mathbf{X}_{\boldsymbol{i}}^* + \boldsymbol{\varepsilon}_{\boldsymbol{x}_{\boldsymbol{i}}}$$
$$\mathbf{Y}^* = \sum_{i=1}^{s} \boxed{\beta_i} \mathbf{X}_i^*$$

$$\mathbf{Y} = \mathbf{Y}^* + \boldsymbol{\varepsilon}_{\boldsymbol{y}}$$
$$\mathbf{X}_{\boldsymbol{i}} = \boxed{\mathbf{X}_{\boldsymbol{i}}^*} + \boldsymbol{\varepsilon}_{\boldsymbol{x}_{\boldsymbol{i}}}$$
$$\mathbf{Y}^* = \sum_{i=1}^{s} \mathbf{X}_i^*$$

- Hasselmann (1979, 1993)
- Hegerl et al (1996, 1997)
- Tett et al (1999)
- Allan and Stott (2003)
- Huntingford et al (2006)
- Hegerl and Zwiers (2011)

- Ribes et al (2013a, 2013b)
- Hannart et al (2014)
- Hannart (2016)

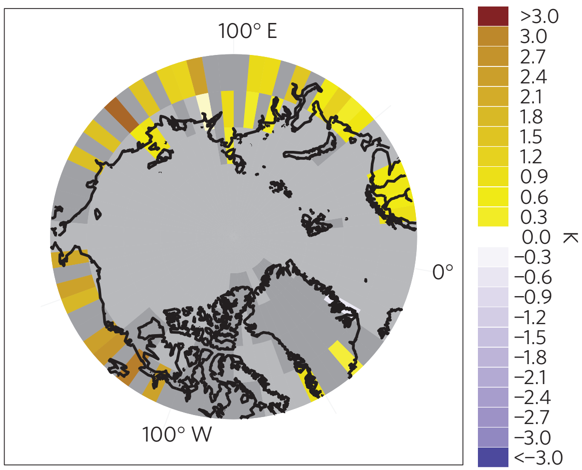
- Ribes et al (2016)

Arctic temperature

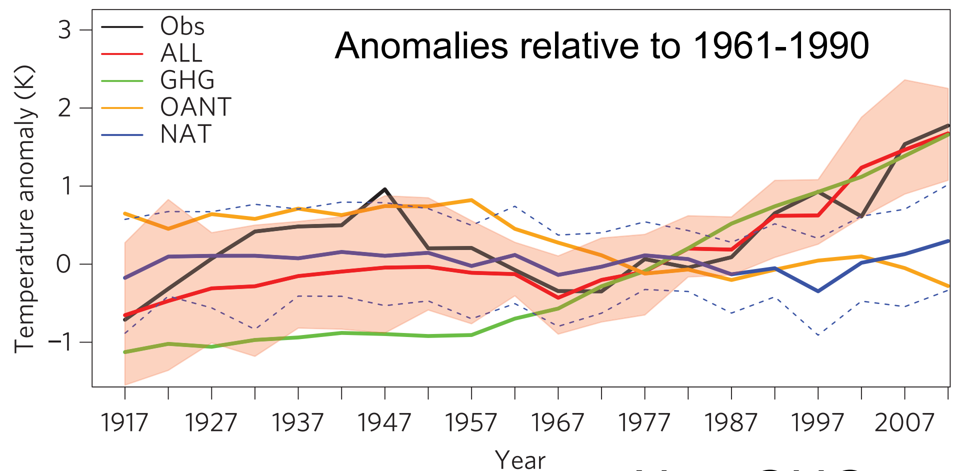
Arctic temperature change – 1913-2012

Najafi et al (2015)

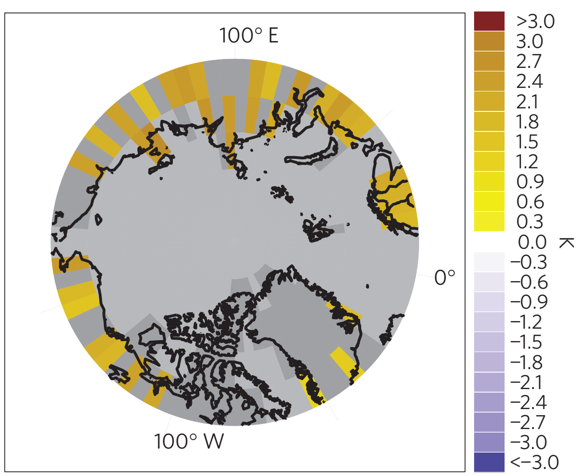
Observed trends



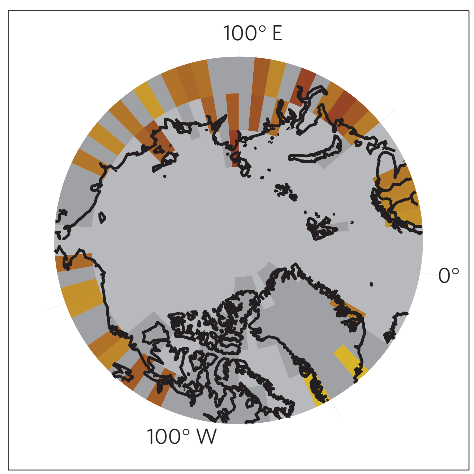
5-year mean Arctic mean temperature



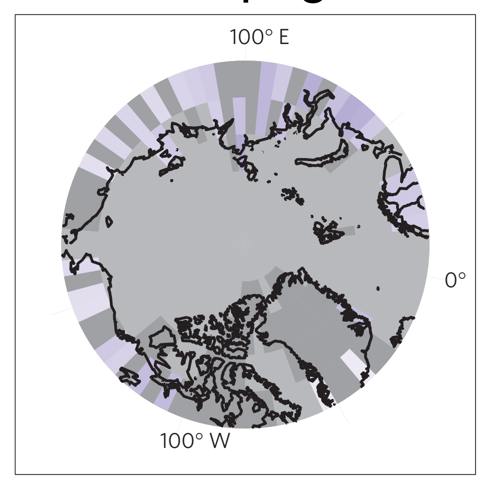
Model trends - ALL



GHG-only



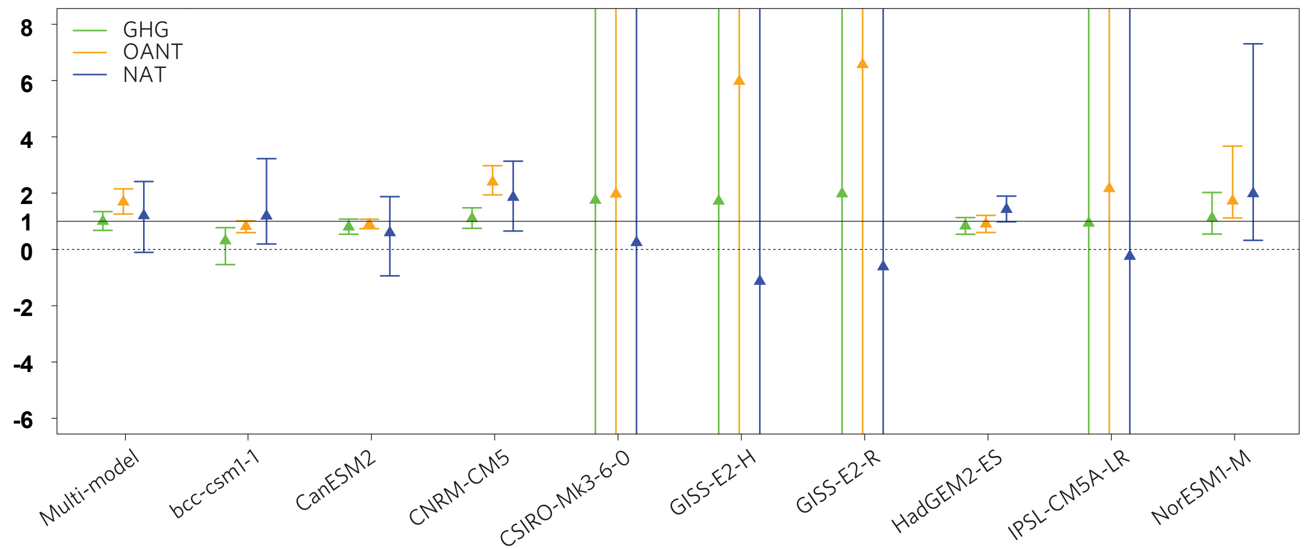
Non-GHG Anthropogenic



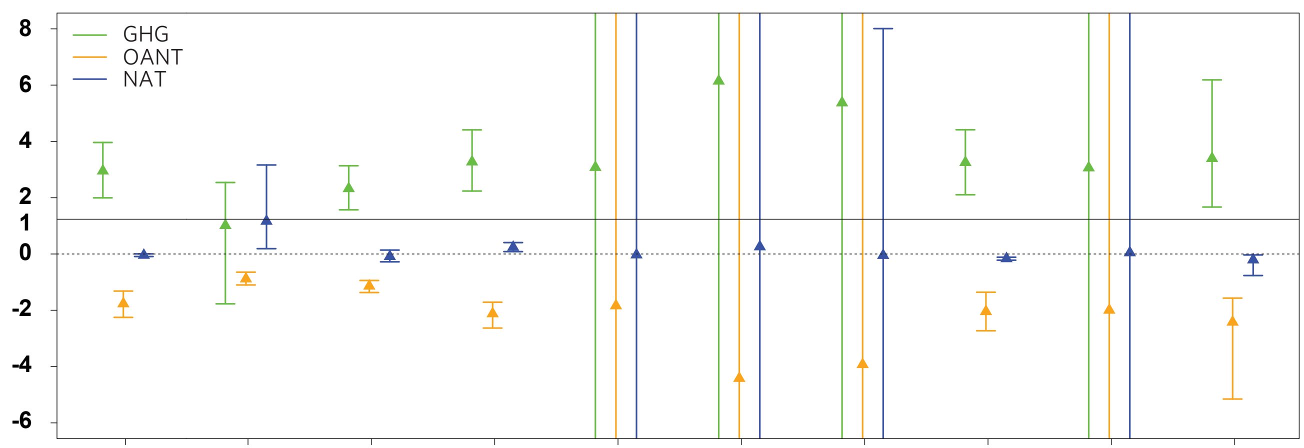
5-yr mean Arctic temperature 1913-2012

D&A analysis based on HadCRUT4 observations for Arctic land area north of 65°N latitude, ALL, GHG and NAT forced runs (9 models, 40 runs total for each forcing) and preindustrial control runs (42 models, 24,800 years)

Scaling factors β



Attributed warming (°C/100-yr)



Some concerns

- Most studies implicitly assume Gaussian noise (generally not a large concern)
- Sampling variability in the estimation of the noise covariance matrix is not accounted for well
  - Hannart (2016) proposes a solution
- Most studies treat inter-model differences as sampling variability equivalent to internal variability
  - Hannart et al (2014) proposes a partial solution
  - Ribes et al (2016) propose an alternative approach
  - In reality, we do not have a comprehensive statistical framework that allows us to describe how the available ensembles of opportunity have been obtained

Some concerns …

- Many studies still use ad-hoc methods for covariance matrix regularization (e.g., EOF-truncation)
  - Some now use better approaches (e.g., the Ledoit-Wolf (2004) estimator) following Ribes et al (2013a, 2013b)
- Many studies do not discuss basic assumptions
  - Key forcings have been identified (and thus there are no other confounding influences)
  - Additivity of signals and noise, or dependence of noise on mean state
- Tendency to attribute based only on statistical evidence (see discussion in Mitchell et al., 2001)

Conclusions

Conclusions

- Evident that some areas have been thought about much more deeply than others
- We should report more completely on methods and assumptions
  - The statistical framework does not need to be perfect
  - But the context for statistical inference should be well understood
  - We sometimes use methods without being clear about the statistical model we are using (eg, event attribution)
- Users bear responsibility as good consumers of the results of our analyses
  - Important, because the costs of application are often borne broadly