Some considerations and methods for uncertainty quantification for event attribution

Christopher Paciorek UC Berkeley Statistics

Various contributions from:

Mark Risser	Lawrence Berkeley National Lab
Daithi Stone	Lawrence Berkeley National Lab
Soyoung Jeon	UTEP
Michael Wehner	Lawrence Berkeley National Lab

NCAR IMAGE Workshop
April 2016

Work funded under the CASCADE project by DOE DE-AC0205CH11231

Model-based event attribution

- Interest lies in understanding whether climate change (or human influence more specifically) is responsible for the occurrence or magnitude of an event or for changes in the magnitudes or probability of a type of events.
- Climate scientists can run climate model(s) multiple times with differing initial conditions, providing an ensemble of replicates
- Experimental design is straightforward compared to observation-based methods for which temporal dependence and confounding are big issues.
- But the model is not the real world.

- Ensemble of n_A model runs under ALL forcings (anthropogenic + natural)
  - Can estimate p_A
- Ensemble of n_N model runs under NAT forcings (natural only)
  - Can estimate p_N

- Estimate either:
  - risk ratio

$$\mathrm{RR} = \frac{p_A}{p_N}$$

  - fraction attributable risk, FAR (probability of necessary causation)

$$\mathrm{FAR} = 1 - \frac{p_N}{p_A}$$

NAS Report on Event Attribution: key points

- Importance of consistent results based on physical principles, observations, numerical models that can reproduce the event, as well as across multiple models and multiple studies
- More confidence in events related to temperature
- Confidence in attribution results is highest for hot and cold events and lowest for severe convective storms and extratropical cyclones
- Extreme events have multiple causes and natural variability is always a component
- Attribution results are sensitive to how the attribution question is posed
- Stating assumptions and quantifying uncertainty is important
- Community standards for attribution analyses would be useful
- Selection bias is an issue and attribution studies are not a representative sample of events
- Systematic attribution is important and would allow evaluation of attribution performance, including uncertainty

NAS Report on Event Attribution: statistical issues

- Statistically-savvy committee members: Gabi Hegerle, Chris Paciorek, Ted Shepherd, Francis Zwiers
- Discussions of causation, uncertainty, and statistics were at the forefront of the committee discussions.
- Some key issues:
  - Multiple causation
  - FAR vs. RR and causation vs. estimation of changes in probability
    - FAR interpreted in terms of causation but communication of multiple causes hard and data not usually available to estimate for THE event
    - Interpretation of constrained model runs in terms of conditional probability

$$\frac{P_A(\mathrm{Event}|\mathrm{Nino})}{P_N(\mathrm{Event}|\mathrm{Nino})} \times \frac{P_A(\mathrm{Nino})}{P_N(\mathrm{Nino})} = \frac{P_A(\mathrm{Event}, \mathrm{Nino})}{P_N(\mathrm{Event}, \mathrm{Nino})}$$

  - Vacuousness of null hypothesis significance testing
  - Multiple sources of selection bias mean that one can't assess influence of climate change on extreme events from collection of existing studies
  - Structural uncertainty (in particular bias) from use of models is hard to quantify
  - Existing statistics in event attribution often mix Bayesian and frequentist concepts
  - Uncertainty in event attribution is largely uncalibrated/unvalidated

Selection bias: thought experiment

Gregory (Scotland Yard detective): Is there any other point to which you would wish to draw attention?
Sherlock Holmes: To the curious incident of the dog in the night-time.
Gregory: The dog did nothing in the night-time
Holmes: That was the curious incident.

Setting: Consider a scientific literature comprised of event attribution analyses of extreme events that occur in a given year.

Question: Will this be biased in terms of collective understanding of the influence of climate change on extreme events?

Thought experiment:

- World with two types of extreme events (say hot and cold events) of equal societal importance.
- Suppose hot events are becoming much more likely ($p_A = .01$, $p_N = 0$) and cold events much less likely ($p_A = 0$, $p_N = .01$).
- Only hot events actually occur, suppose 10 occur.
- Suppose we have unlimited data.
- 10 events are analyzed, and all will be judged to be attributed to climate change.
- Headline: climate change causing extreme events to become more likely.
- We might call this "occurrence bias".
- Some similarity to the file drawer problem in meta-analysis

Possible solution:

- Predefine extreme events, including all possible event types based on historical record, and analyze systematically

UQ: sources of uncertainty

- Sampling uncertainty (limited ensemble sizes for model analyses, limited observational record for observation-based analyses)
- Uncertainty from internal variability (long-term dependence)
- Generalizability uncertainty from conditioned analyses (fixed SSTs)
- Generalizability uncertainty / selection bias from analysis of individual events
- Structural uncertainty
  - Parametric uncertainty
  - Model structure uncertainty (variation across models)
  - Model inadequacy
  - Boundary condition uncertainty
  - Forcing uncertainty
- Counterfactual state uncertainty
  - With fixed SSTs, what is the counterfactual SST state that should be used
  - What would analogous world look like, but conditioned on measure-zero state?

UQ: Sampling uncertainty - current

Goal: quantify uncertainty in risk ratio estimation given limited ensemble sizes.

Data: Initial condition ensembles (induces a distribution over event magnitude/occurrence)

Estimation:
- Count events (binomial uncertainty)
- Extreme value methods

Popular method: bootstrap from ensemble members and report bootstrap distribution of estimated FAR or RR graphically or as 95% interval
  - Bayesian interpretation
  - Interval corresponds to percentile bootstrap

Stott et al. (2004; Nature)

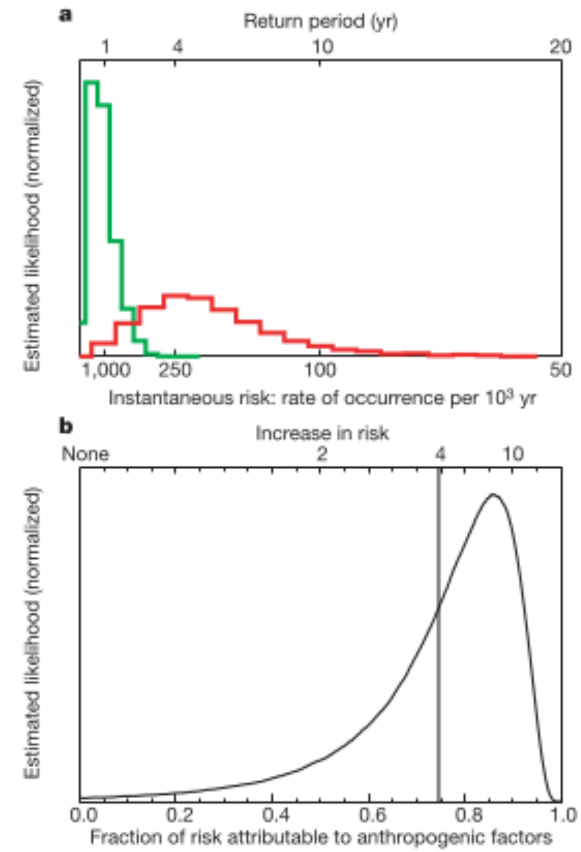


Figure 4 Change in risk of mean European summer temperatures exceeding the 1.6 K threshold. **a**, Histograms of instantaneous return periods under late-twentieth-century conditions in the absence of anthropogenic climate change (green line) and with anthropogenic climate change (red line). **b**, Fraction attributable risk (FAR). Also shown, as the vertical line, is the 'best estimate' FAR, the mean risk attributable to anthropogenic factors averaged over the distribution.

'ublishing Group NATURE | VOL 432 | 2 DECEMBER 2004 | www.nature.com/nature

UQ: Sampling uncertainty – potential methods

Goal: quantify uncertainty in risk ratio estimation given limited ensemble sizes.

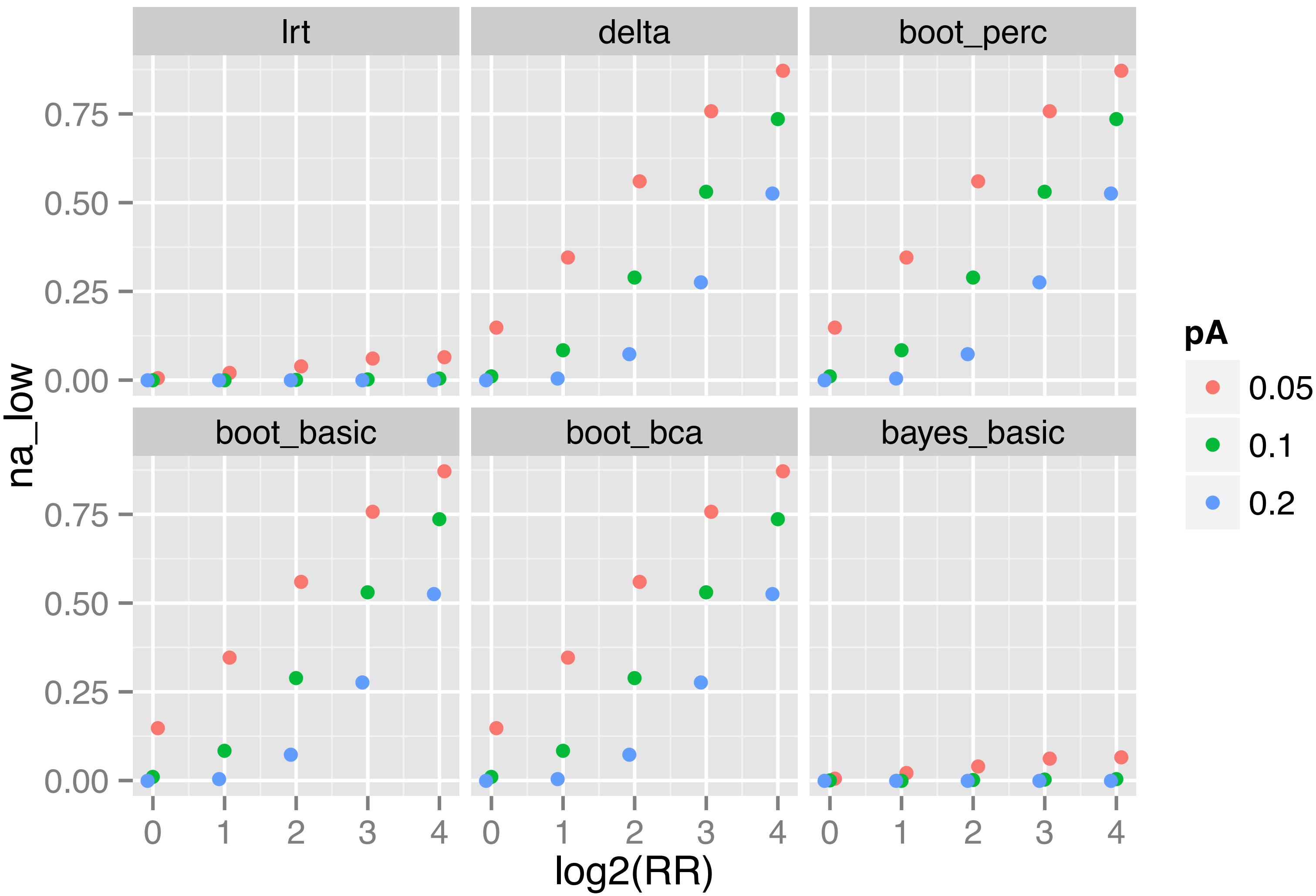
Data: Initial condition ensembles (induces a distribution over event magnitude/occurrence)

Potential methods for uncertainty in $RR=p_A/p_N$ based on binomial or EVA:
- Delta method
- Bootstrap
  - Standard error of log(RR)
  - Percentile, basic bootstrap, bootstrap-t, BCA confidence intervals
- Likelihood ratio-based interval: invert test for $RR=RR_0$
- Exact confidence interval under binomial sampling (Wang & Shan (2015); Biometrics)
  - Enumerates probabilities of more extreme data under a given rank function
  - Only under binomial setting

Considerations:
- Assumptions for delta method and bootstrap not well-satisfied for small sample sizes and proportions near zero (and some bootstrap samples do not provide an estimate)
- No estimate of uncertainty when $\widehat{p_N} = 0$ using delta method or bootstrap
- Likelihood ratio method still relies on asymptotics
- Wang & Shan (2015) approach intriguing but computationally intensive

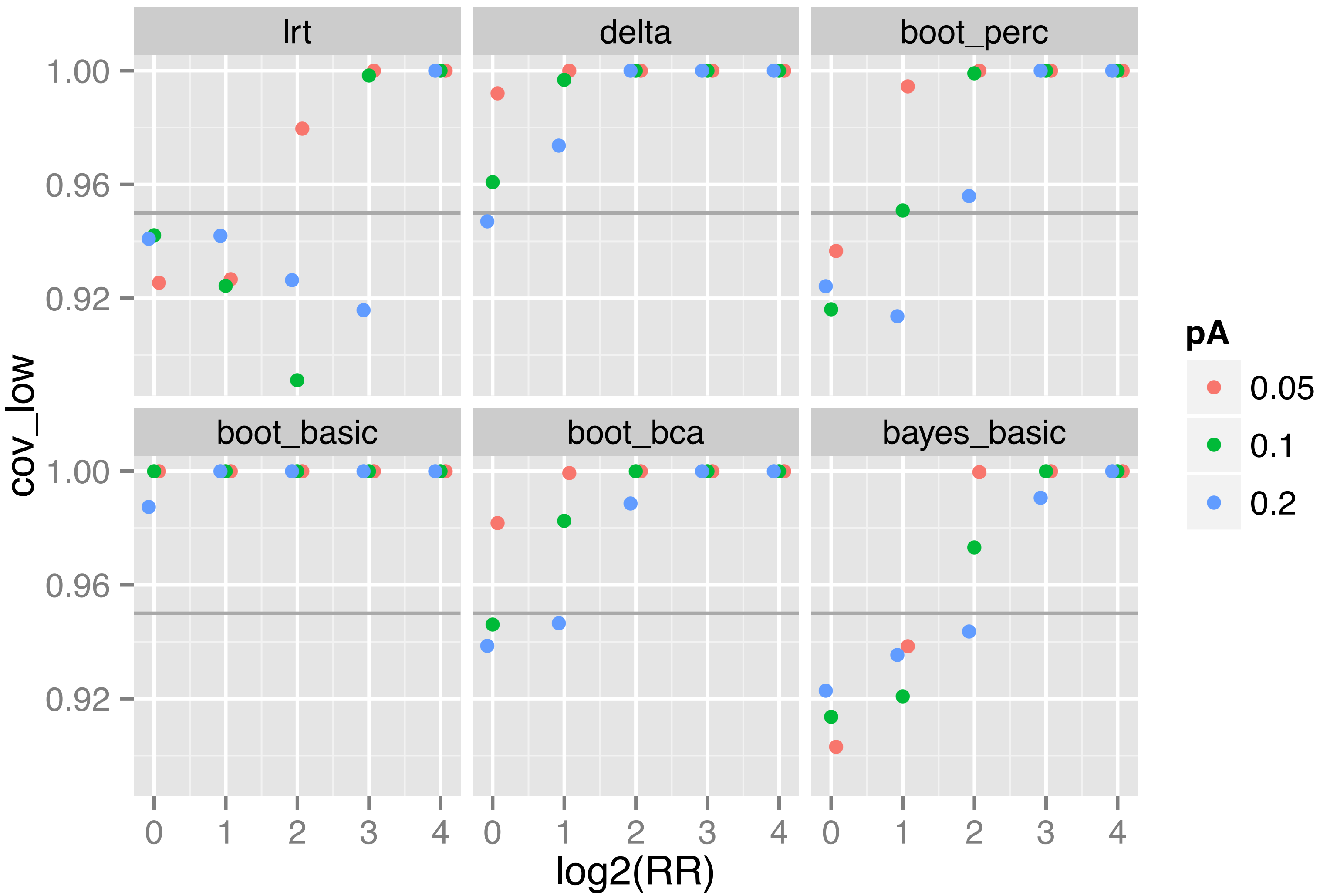
UQ: Sampling uncertainty – assessment

Proportion of simulations for which uncertainty estimate could not be calculated



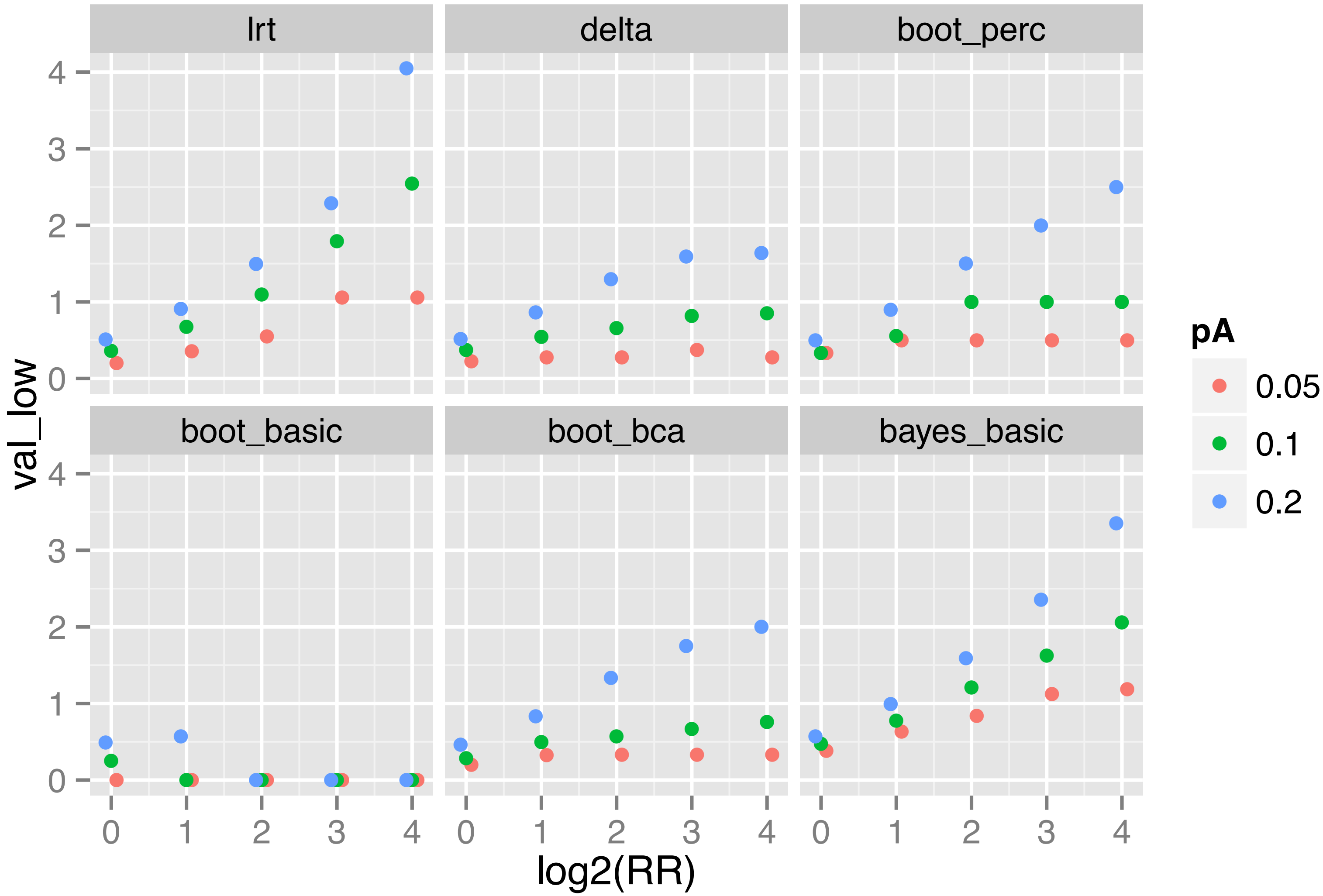
UQ: Sampling uncertainty – assessment

Coverage results for one-sided (L, ∞) intervals for RR



UQ: Sampling uncertainty – assessment

Average lower endpoint, L, for one-sided (L, ∞) intervals for RR



UQ: Conditioning on boundary conditions (SSTs)

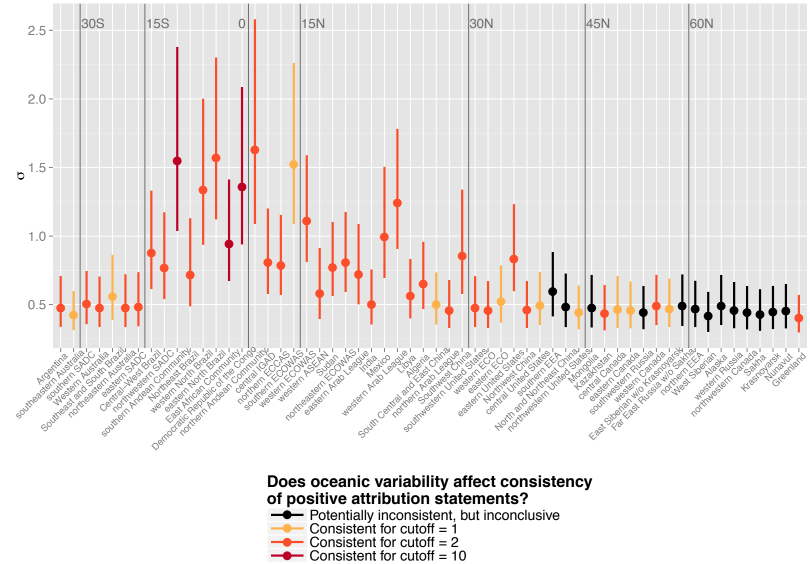
Question: How much do event attribution statements change for different years (different SST conditions)?

Data:

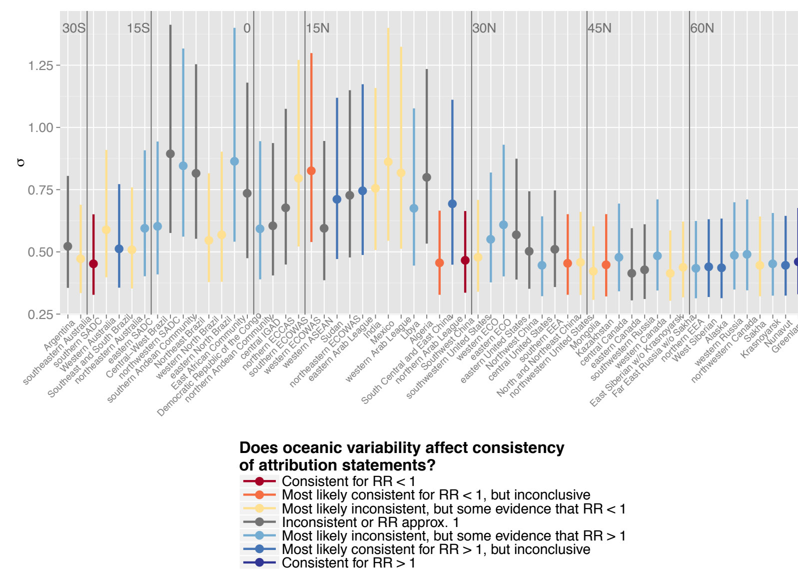
- All and natural forcings CAM5 model ensembles for 1960s-2013
- Count events bigger than some threshold

$$Z_{t,A} \sim \text{Bin}(n_t, p_{t,A})$$
$$Z_{t,N} \sim \text{Bin}(n_t, p_{t,N})$$
$$\text{logit}(p_{t,A}) = \beta_{0,A} + \alpha_t + \delta_t$$
$$\text{logit}(p_{t,N}) = \beta_{0,N} + \alpha_t$$
$$\delta_t \sim \mathcal{N}(0, \sigma^2)$$
$$\text{RR}_t \approx \exp(\beta_{0,A} - \beta_{0,N}) \exp(\delta_t)$$
$$\sigma^2 \approx \text{Var}(\log \text{RR}_t)$$

Hot events

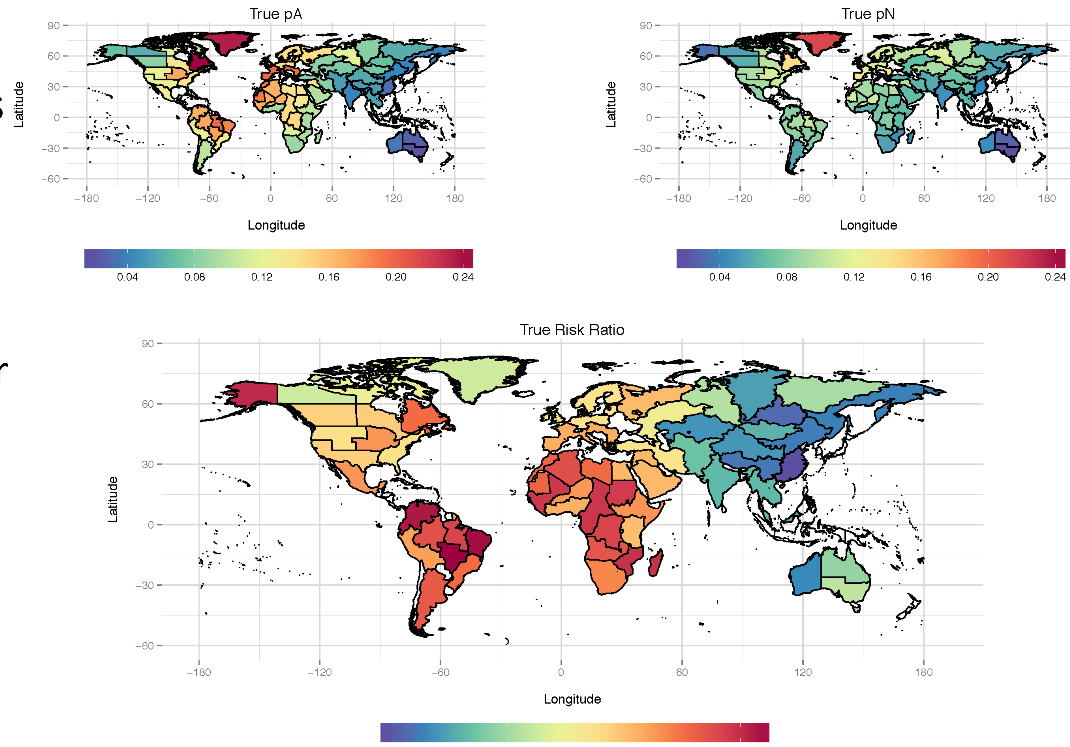


Wet events



UQ: Multiplicity in systematic attribution

Goal: develop procedure for classifying regions in terms of magnitude and uncertainty of risk ratio and use for ongoing monthly attribution forecasts



Simulations to compare:

- Unadjusted estimates
- Classical FDR
  - Thought to be conservative under positive spatial dependence
- FDR accounting for dependence
  - Sun et al. (2015; JRSSB)
  - Possible sensitivity to model misspecification
  - Uses Bayesian estimators
- Unadjusted Bayesian estimators
  - Rely on shrinkage to deal with multiplicity

UQ: structural uncertainty

Strategies for structural uncertainty: model, parameter, boundary conditions

- Frequentist or Bayesian?
- Sample from distribution of plausible parameters/models as prior distribution?
    - Does ad hoc assemblage of models constitute anything we can use as a such a prior? Simple sensitivity analysis instead?
    - Condition on observations to filter models or parameters (what summary stats to use?)
- Parametric uncertainty
    - Approaches to designing statistical emulators with stochastic outputs (distribution over event magnitudes or counts of events)
- Are any of the UQ methods from statistics or physics helpful for model inadequacy?

```
Z_{t,A} & \sim \mbox{Bin}(n_t, p_{t,A}) \\
Z_{t,N}  & \sim  \mbox{Bin}(n_t, p_{t,N}) \\
\mbox{logit}(p_{t,A}) & =  \beta_{0,A} + \alpha_t + \delta_t \\
\mbox{logit}(p_{t,N}) & =  \beta_{0,N} + \alpha_t \\
\delta_t & \sim  \mathcal{N}(0,\sigma^2) \\
\mbox{RR}_t  & \approx  \exp(\beta_{0,A}-\beta_{0,N}) \exp(\delta_t)
\\
\sigma^2 & \approx  \mbox{Var}(\log \mbox{RR}_t)
```