A CONDITIONAL APPROACH TO EXTREME EVENT ATTRIBUTION

Richard L. Smith

University of North Carolina, Chapel Hill, USA
rls@email.unc.edu

IDAG 2023, Exeter, July 5, 2023

Slides, datasets etc.: http://rls.sites.oasis.unc.edu/ClimExt/intro.html

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Objectives

1. Develop an automated method of extreme event attribution analysis using only public data sources

2. Trying to extend existing approaches, not contradict them

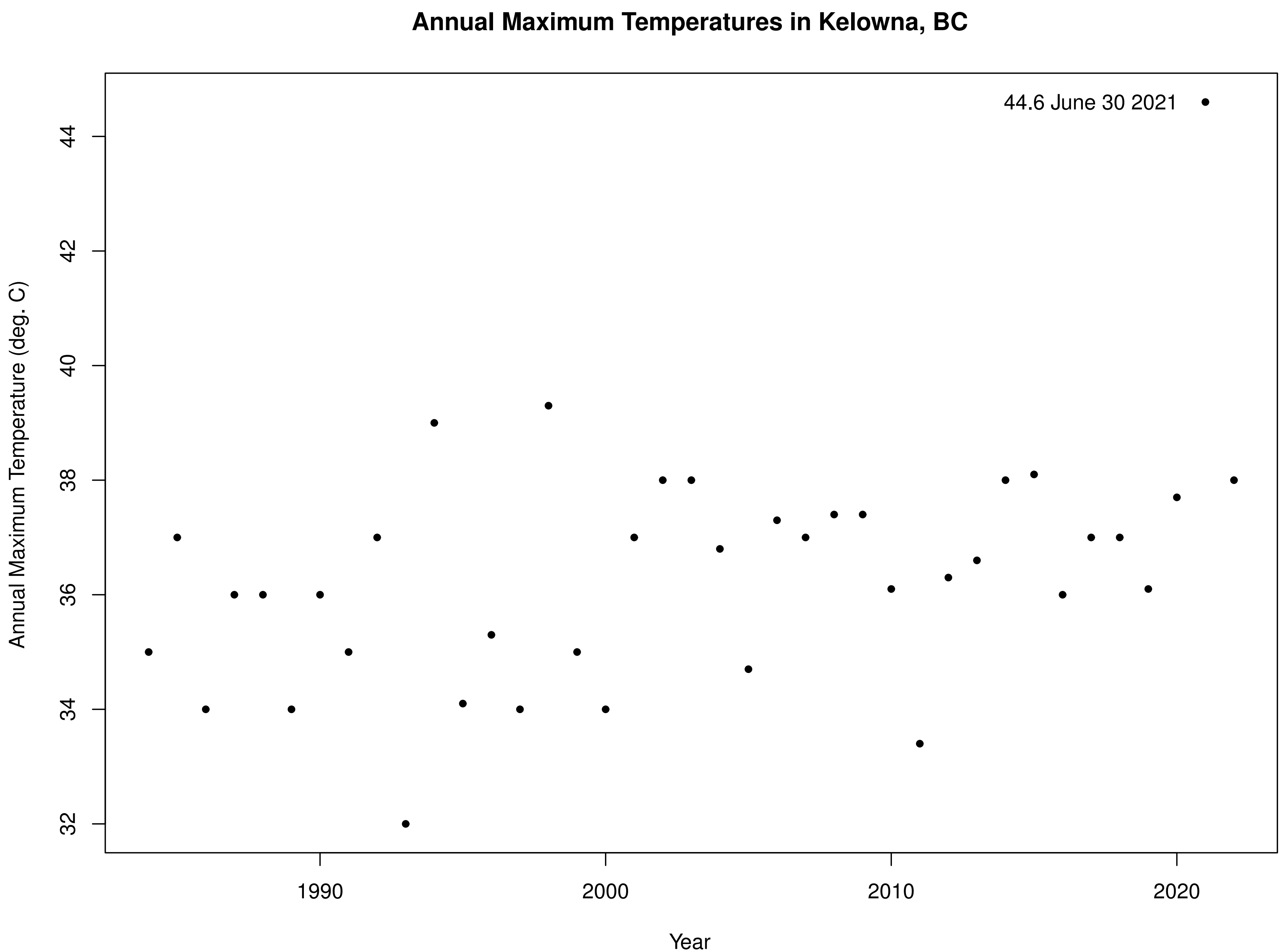
3. Acknowledging that dynamical methods will ultimately outperform statistical methods, but the latter are much quicker to calculate and provide an independent validation

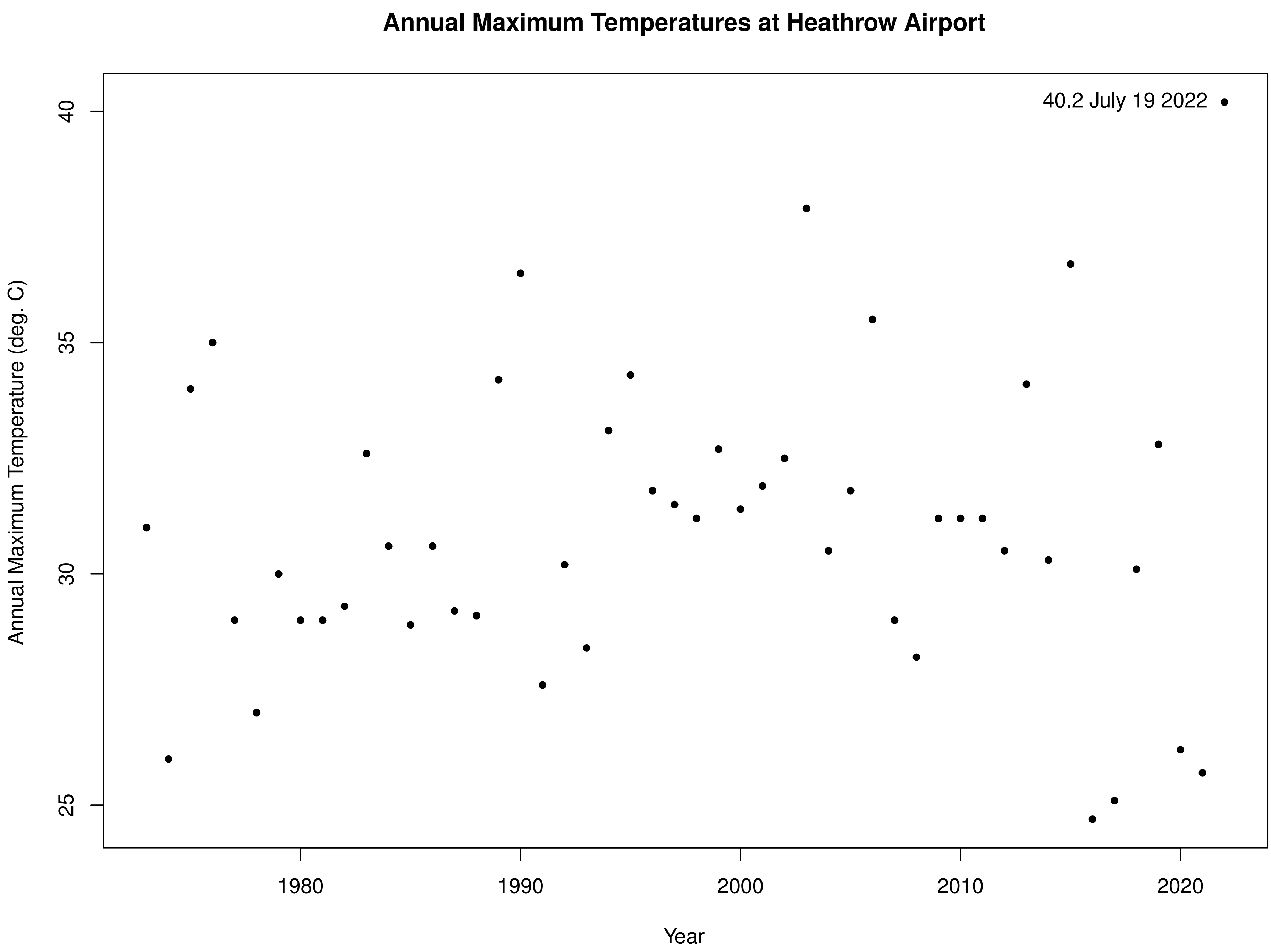
4. Key idea of this talk: include a *conditioning variable* — some regional climate indicator at a more localized scale than GMST

5. Second key idea: projections of future extreme event probabilities

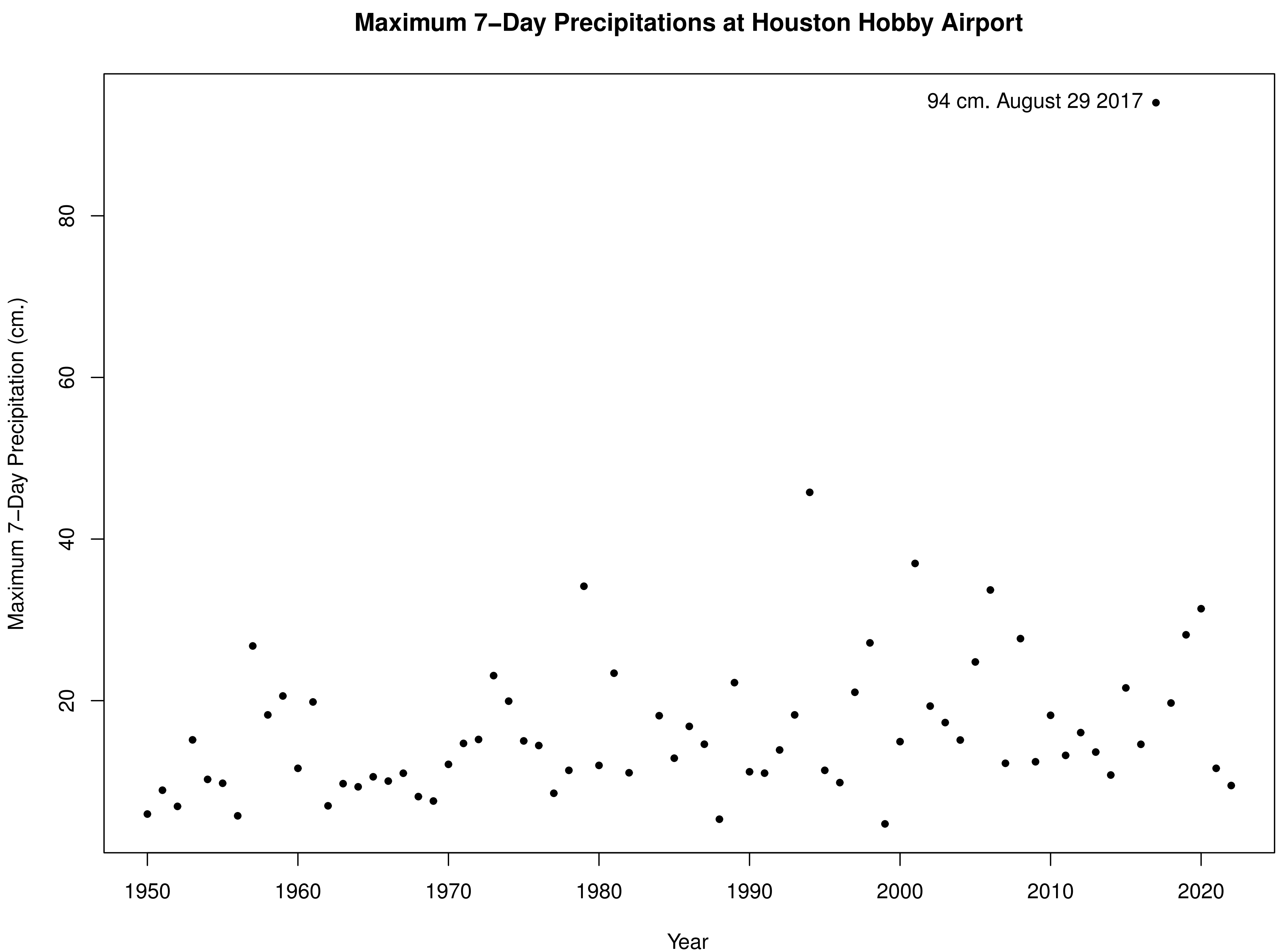
I. Introduction

I begin with three examples of datasets that contain extreme events


Annual Maximum Temperatures in Kelowna, BC
44.6 June 30 2021
Annual Maximum Temperature (deg. C)
32
34
36
38
40
42
44
1990
2000
2010
2020
Year

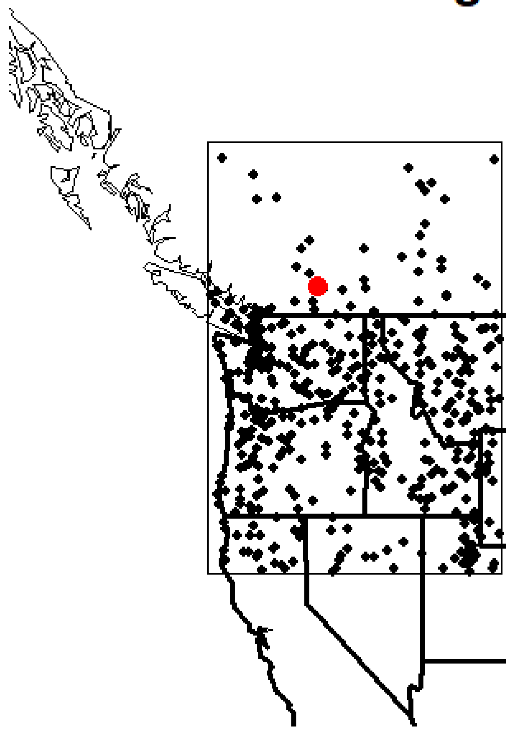


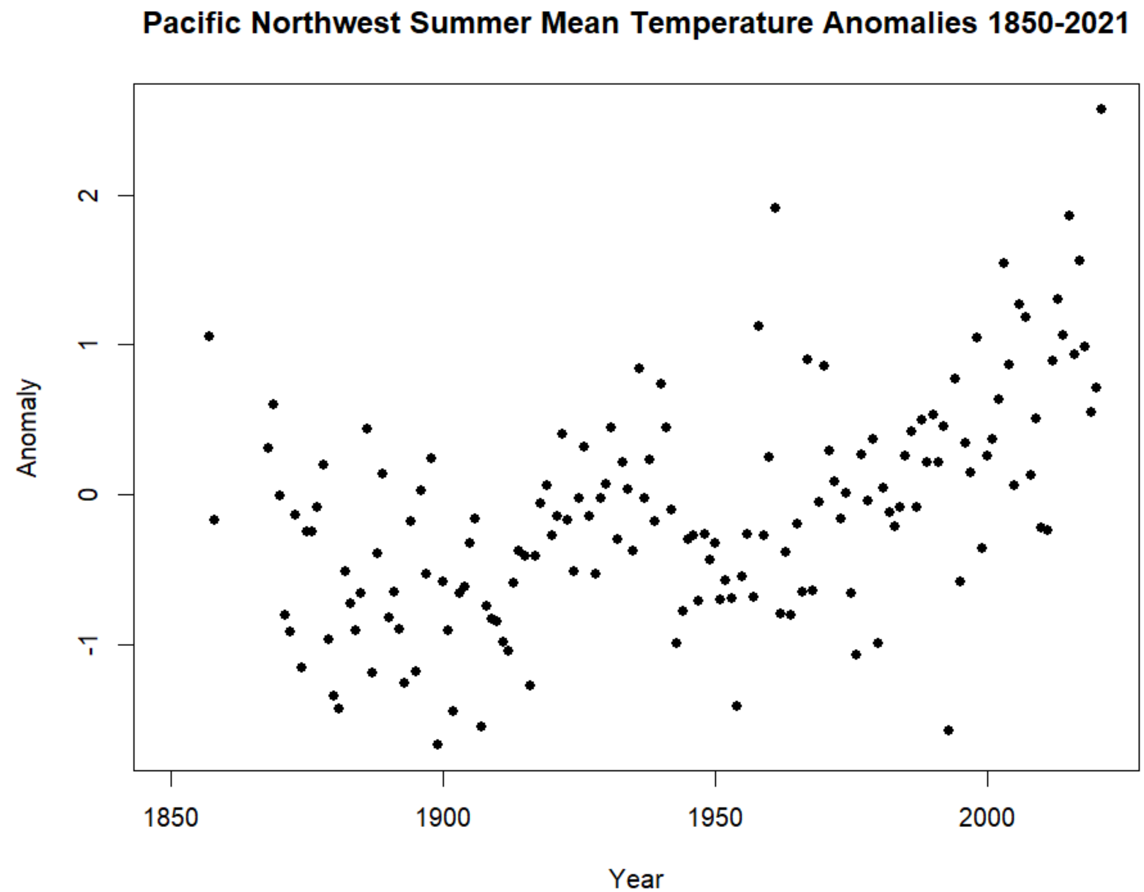
Annual Maximum Temperatures at Heathrow Airport
40.2 July 19 2022
Annual Maximum Temperature (deg. C)
40
35
30
25
1980
1990
2000
2010
2020
Year


Maximum 7−Day Precipitations at Houston Hobby Airport
94 cm. August 29 2017
Maximum 7−Day Precipitation (cm.)
20
40
60
80
1950
1960
1970
1980
1990
2000
2010
2020
Year

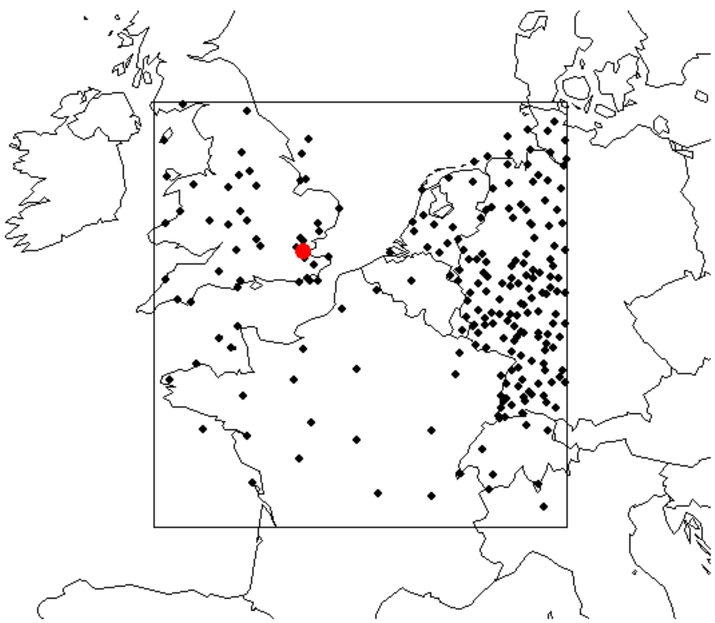
For each of these examples, I have collected weather data from multiple stations in the same region (from the Global Historical Climatological Network), and also calculated a *regional variable* that includes annual or seasonal maxima from spatially aggregated data (from the Climate Research Unit of the University of East Anglia)

Pacific Northwest Region

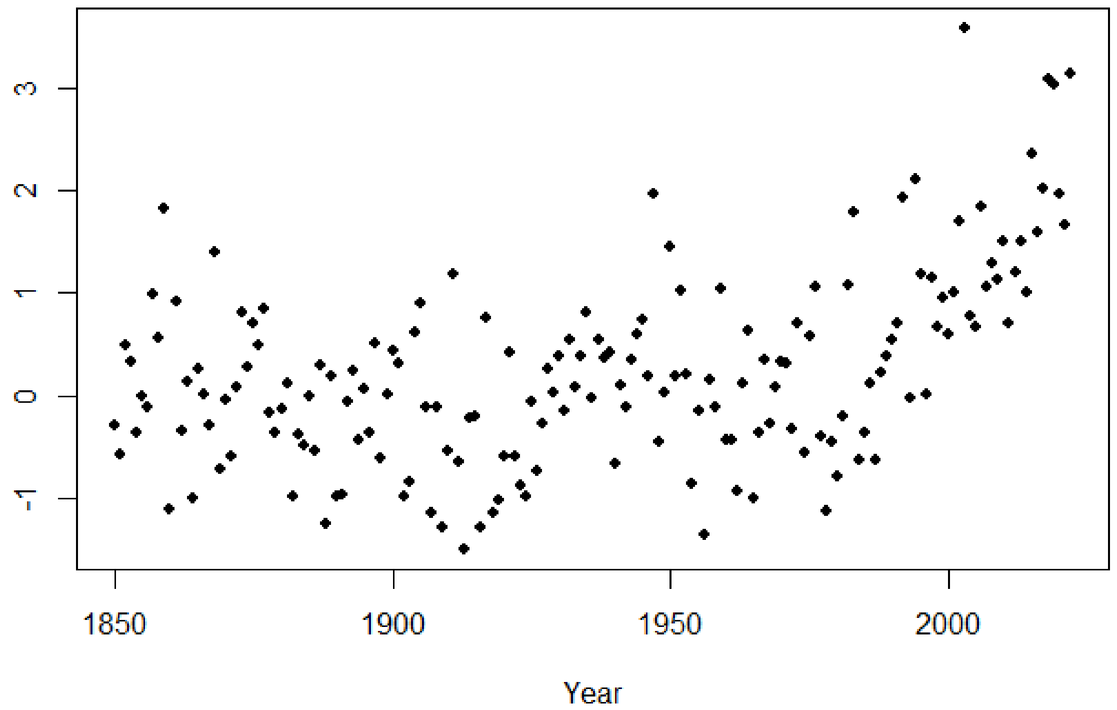


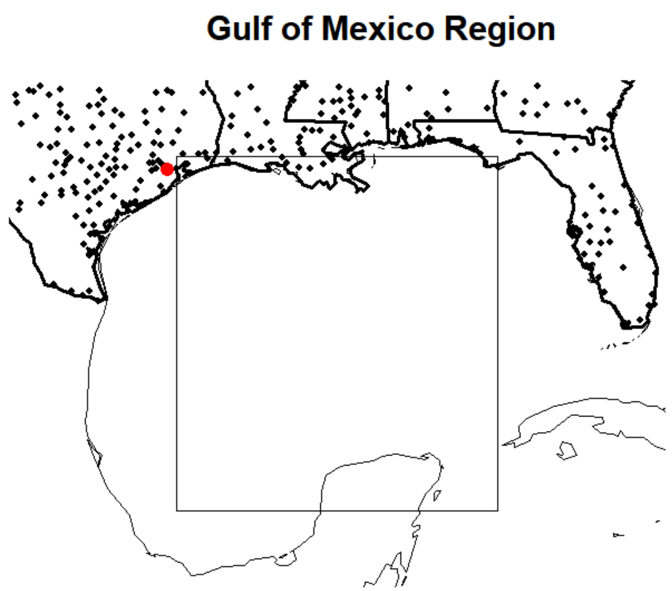


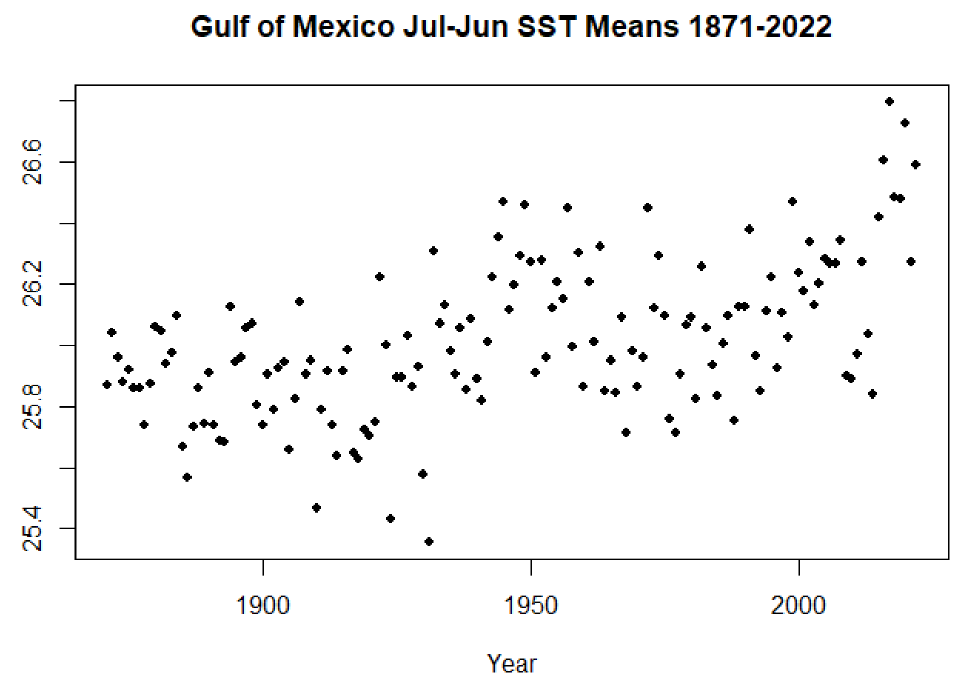
Northern Europe Region



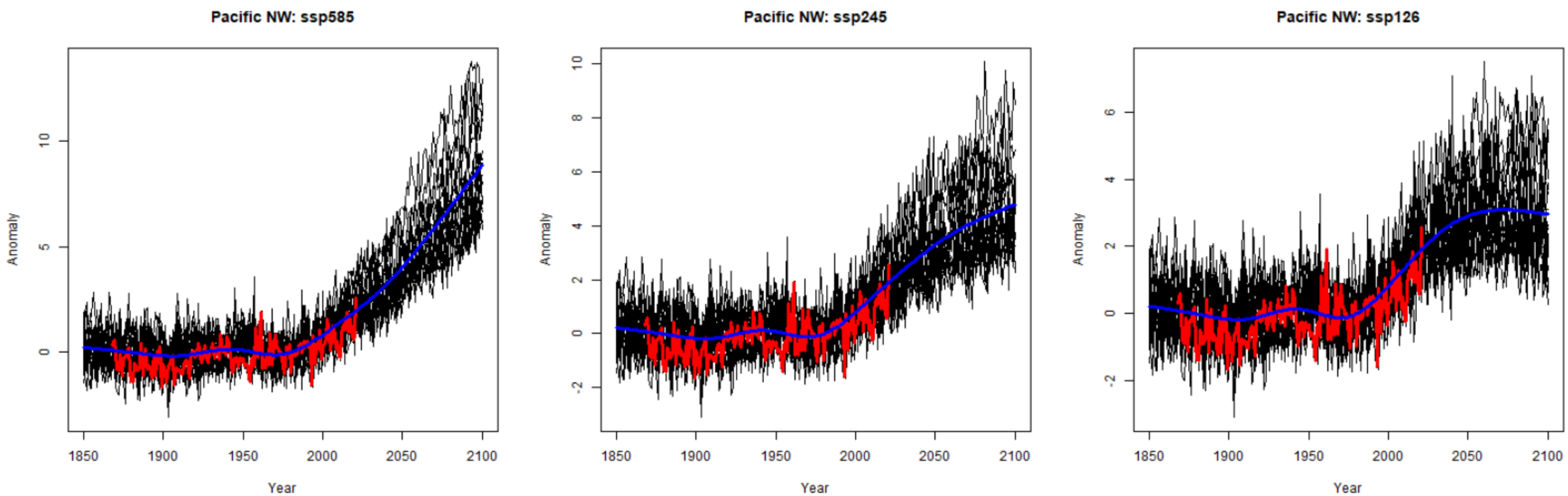
Northern Europe Summer Mean Temperature Anomalies 1850-2022

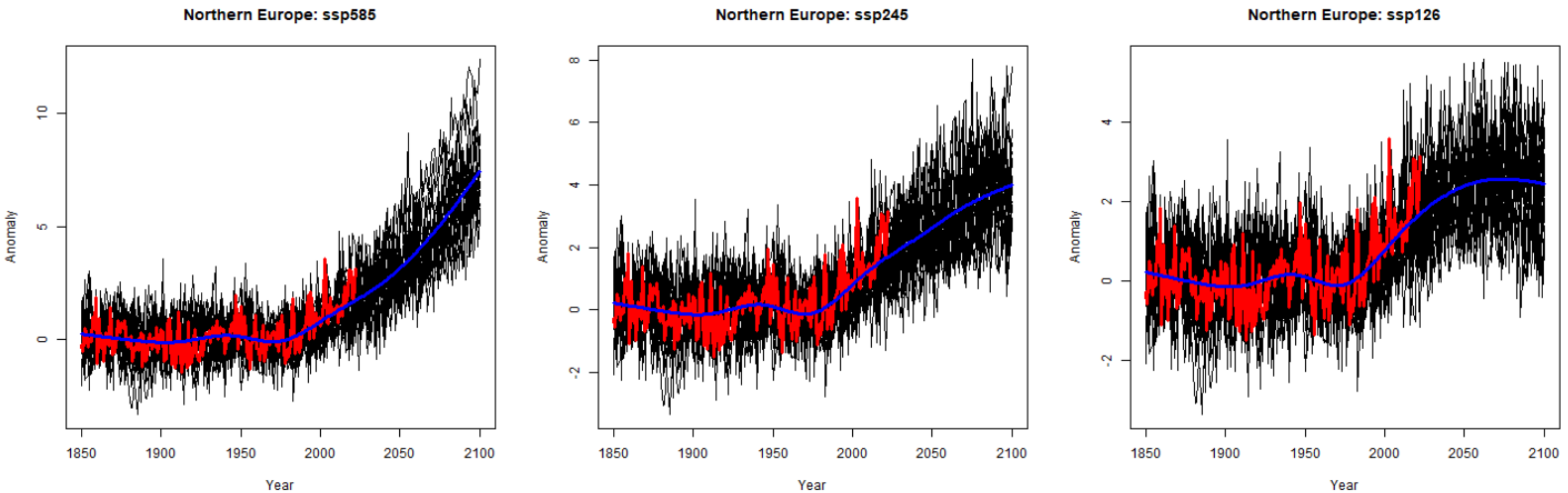


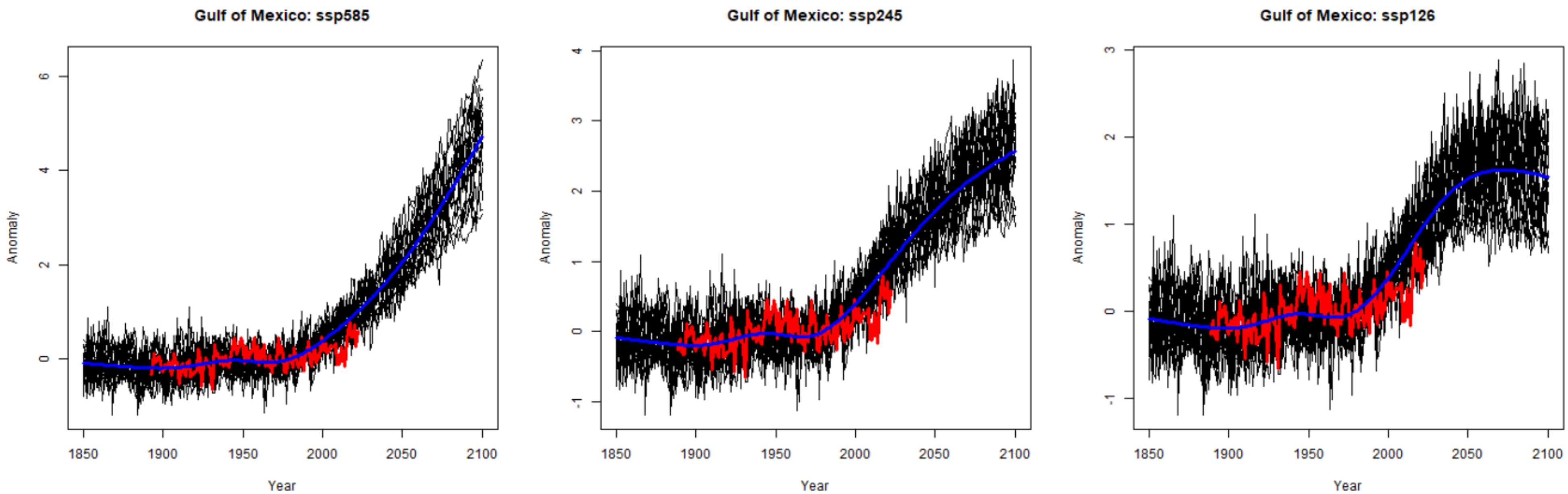

Gulf of Mexico Region


Gulf of Mexico Jul-Jun SST Means 1871-2022
26.6
26.2
25.8
25.4
1900
1950
2000
Year

I have also compiled 17 climate model datasets (from CMIP6) that correspond to the regional variables defined above


Pacific NW: ssp585
Pacific NW: ssp245
Pacific NW: ssp126
Anomaly
Year
1850
1900
1950
2000
2050
2100
0
5
10
-2
2
4
6
8


Northern Europe: ssp585
Northern Europe: ssp245
Northern Europe: ssp126
Anomaly
Year
1850
1900
1950
2000
2050
2100



Gulf of Mexico: ssp585
Anomaly
Year
Gulf of Mexico: ssp245
Anomaly
Year
Gulf of Mexico: ssp126
Anomaly
Year

II. Statistical Analysis

IIa. Used the Generalized Extreme Value (GEV) for each station with regional variable as a covariate

IIb. Combine stations using a spatial model

IIc. Climate models to project the regional variable forwards and backwards in time

IId. "End to end" analysis to show how the extreme event probability changes corresponding to climate variation (including uncertainty bounds)

IIa. GEV Analysis

$$G(y) \; = \; \mathrm{Pr}\{Y \le y\} \; = \; \exp\left\{-\left(1+\xi\frac{y-\mu}{\psi}\right)_{+}^{-1/\xi}\right\}$$

- Parameters $\mu,\ \psi,\ \xi$ depend on time and space
- Time dependence based on regional variable as a covariate
- Point of clarification: There is a debate in the literature about whether the analyzed data should include the extreme event of interest. The results I am showing here do *not* do this: the analyses for Kelowna, London and Houston are based on station data up to 2020, 2021 and 2016 respectively.

Covariate Models
(Risser and Wehner 2017, Russell et al. 2020)

$$\begin{aligned} \mu_{s,t} &= \theta_{s,1} + \theta_{s,4} X_t, \\ \log \psi_{s,t} &= \theta_{s,2} + \theta_{s,5} X_t, \\ \xi_{s,t} &= \theta_{s,3}, \end{aligned}$$

Define a parameter vector $\mathbf{\Theta}_s = \left(\theta_{s,1} \ \ \ldots \ \ \theta_{s,5} \right)$ at each site s; a 5-dimensional parameter vector for each site s.

Extension: $\log \left\{ \frac{1+\xi_{s,t}}{1-\xi_{s,t}} \right\} = \theta_{s,3} + \theta_{s,6} X_t$ (6-parameter model), also combined into $\mathbf{\Theta}_s$

IIb. Spatial Extremes Analysis

Objective: Come up with a model for interpolating the GEV distributions between stations, and also improving the analysis at individual stations by "borrowing strength" across stations.

- Latent process approach: Russell, Risser, Smith and Kunkel (2020)
- Idea is to "combine strength" across different stations
- Fit a spatial model to all the stations, then project backwards to specific locations (including the stations)
- Several other approaches, see in particular Zhang, Risser, Wehner and O'Brien (forthcoming)

Kelowna, B.C. (Single Station Approach)

5-Par Model:

Parameter	Estimate	SE	t-val	p-val
θ_1	34.8265	0.2511	138.7138	0.0000
θ_2	0.0703	0.1812	0.3882	0.6979
θ_3	-0.3709	0.3533	-1.0497	0.2939
θ_4	1.8317	0.2708	6.7642	0.0000
θ_5	-0.0958	0.3372	-0.2841	0.7763

MLE probability of exceeding 44.6°C in 2021, given X_{2021}: 0.

Bayesian posterior mean: 0.012 (1-in-83-year event, even *given* the high regional temperature)

6-Par Model:

Parameter	Estimate	SE	t-val	p-val
θ_1	34.8386	0.2809	124.0051	0.0000
θ_2	0.1397	0.1866	0.7486	0.4541
θ_3	-0.9475	0.4582	-2.0679	0.0386
θ_4	1.8494	0.2686	6.8861	0.0000
θ_5	-0.2301	0.2755	-0.8352	0.4036
θ_6	1.1113	0.7217	1.5399	0.1236

MLE probability for 2021 is 0.072, Bayesian 0.076 (1-in-13-year)

Results: Kelowna (6-Par Spatial Model)

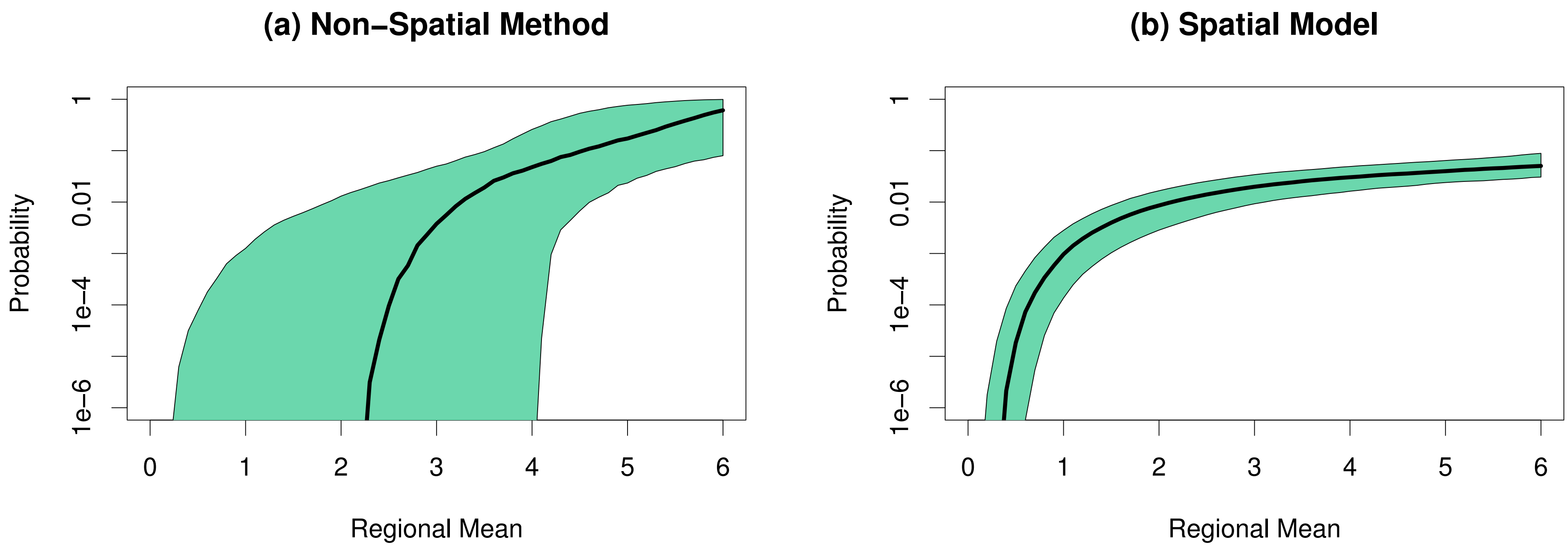
MLE Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	34.8386	0.2809	124.0051	0.0000
θ_2	0.1397	0.1866	0.7486	0.4541
θ_3	-0.9475	0.4582	-2.0679	0.0386
θ_4	1.8494	0.2686	6.8861	0.0000
θ_5	-0.2301	0.2755	-0.8352	0.4036
θ_6	1.1113	0.7217	1.5399	0.1236

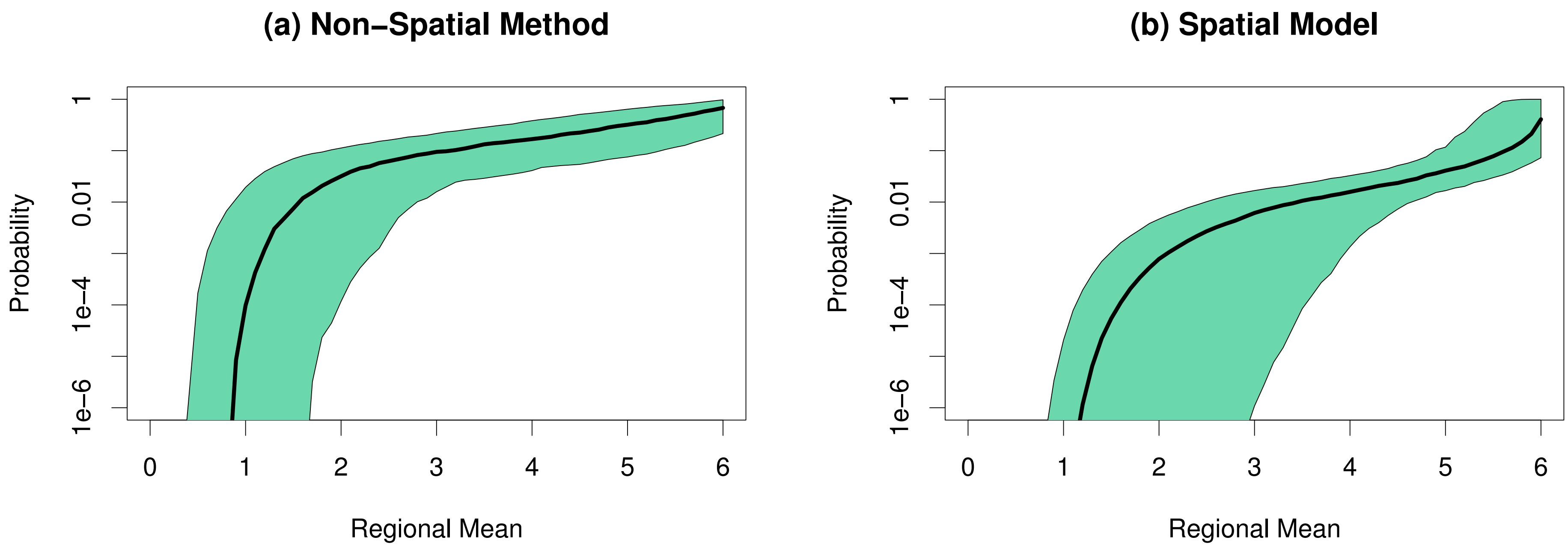
Spatial Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	34.8437	0.1767	197.2273	0.0000
θ_2	0.1099	0.0808	1.3597	0.1739
θ_3	-0.5908	0.1272	-4.6438	0.0000
θ_4	1.7402	0.1530	11.3750	0.0000
θ_5	-0.3748	0.1219	-3.0754	0.0021
θ_6	0.4290	0.2025	2.1185	0.0341

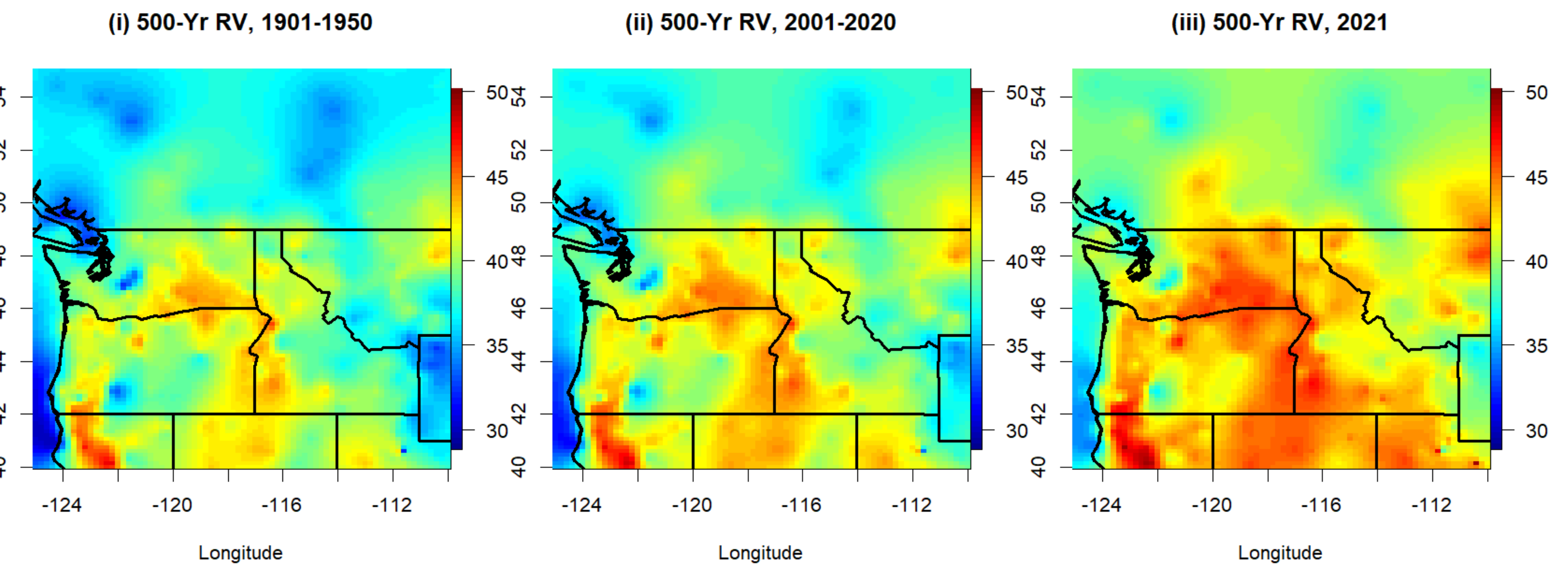
Estimates and 66% Credible Intervals for Mean Exceedance Probability: Comox, B.C.



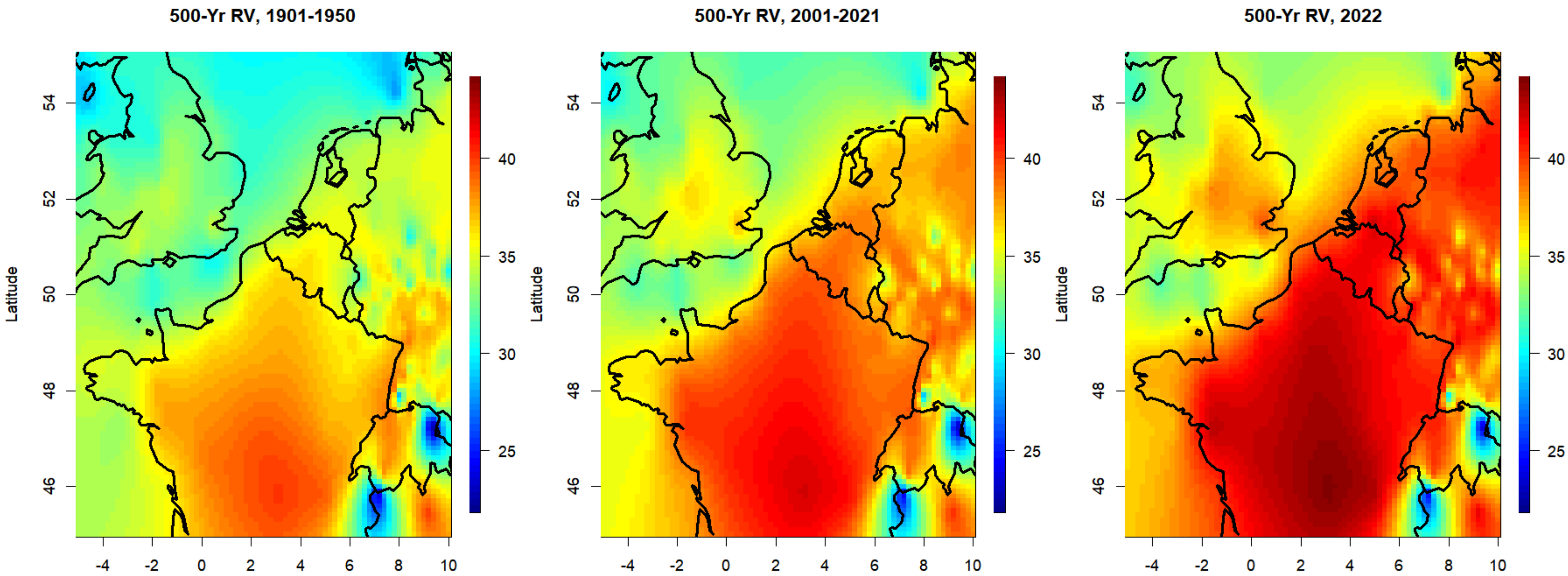
Estimates and 66% Credible Intervals for Mean Exceedance Probability: Kelowna (with monotonicity constraint)



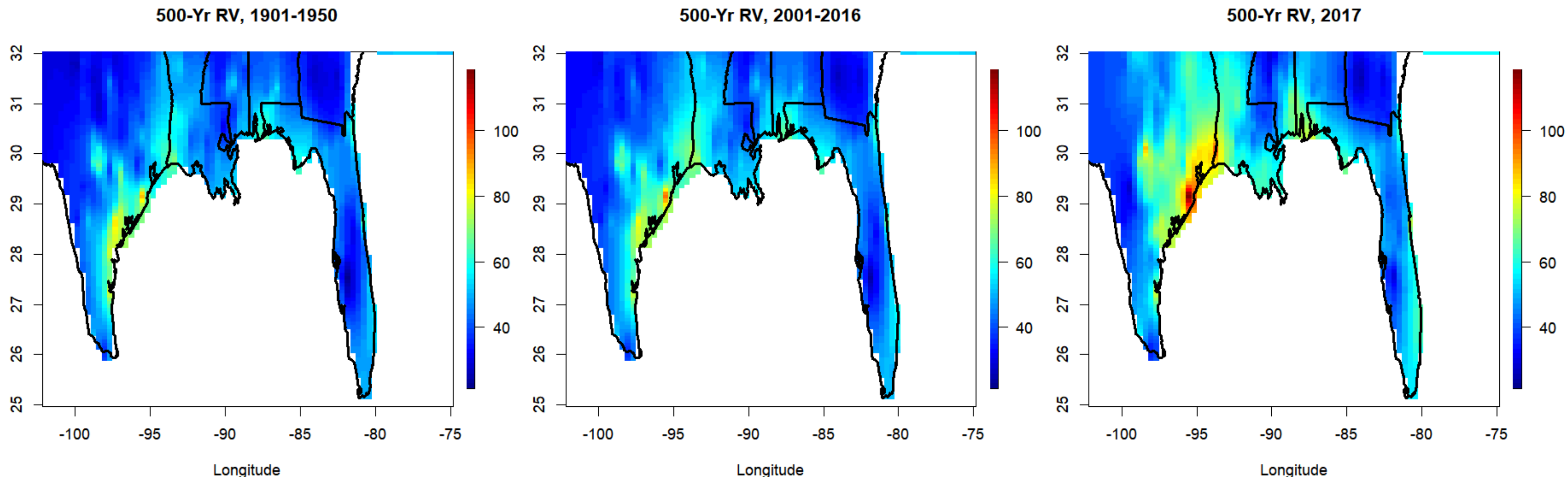
PNW: 500-year return values for (i), (ii), (iii)



NEU: 500-year return values for (i), (ii), (iii)



GOM: 500-year return values for (i), (ii), (iii)



Houston, we have a problem

Looking at the Probabilities of Individual Events

Conditional probabilities of exceeding 2021 temp in PNW:

	(i) 1901–1950	(ii) 2001–2020	(iii) 2021
Kelowna (44.6°C)	3×10^{-12}	8.6×10^{-6}	0.0061
All Canadian stations	0.0081	0.0185	0.067

Conditional probabilities of exceeding 2022 temp in UK:

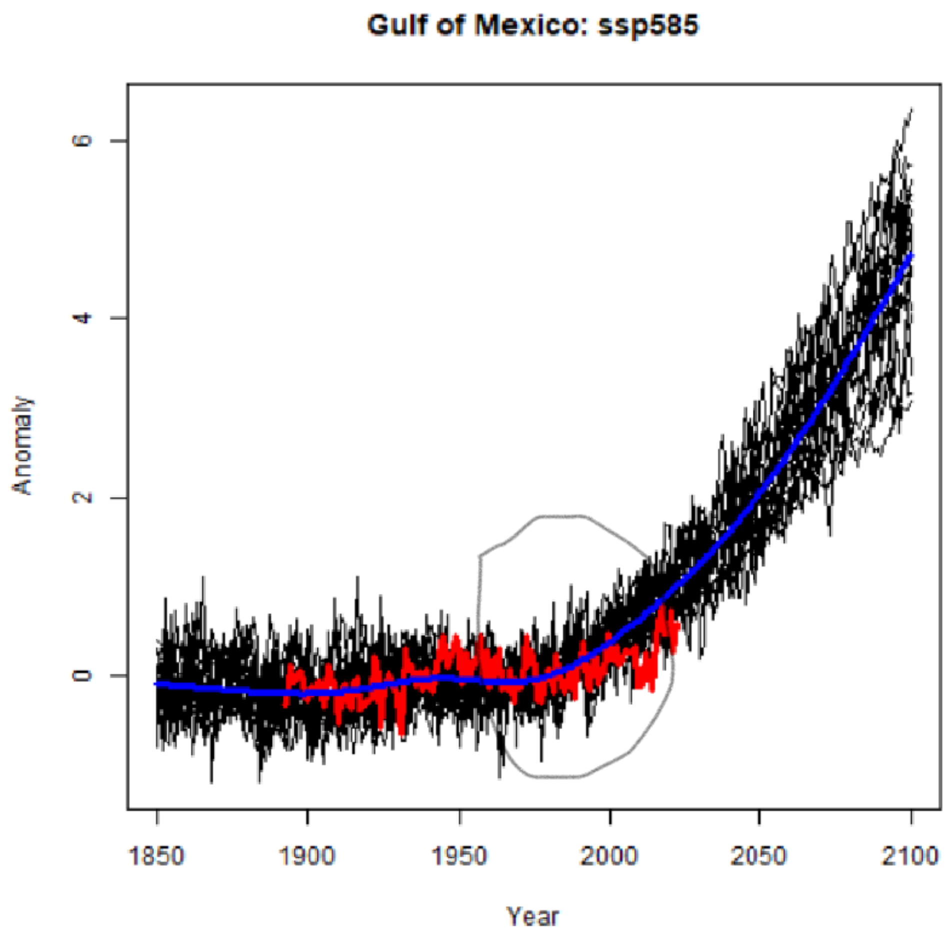
	(i) 1901–1950	(ii) 2001–2021	(iii) 2022
Heathrow	0	3.1×10^{-5}	0.017
All U.K. stations	0.0081	0.0319	0.095

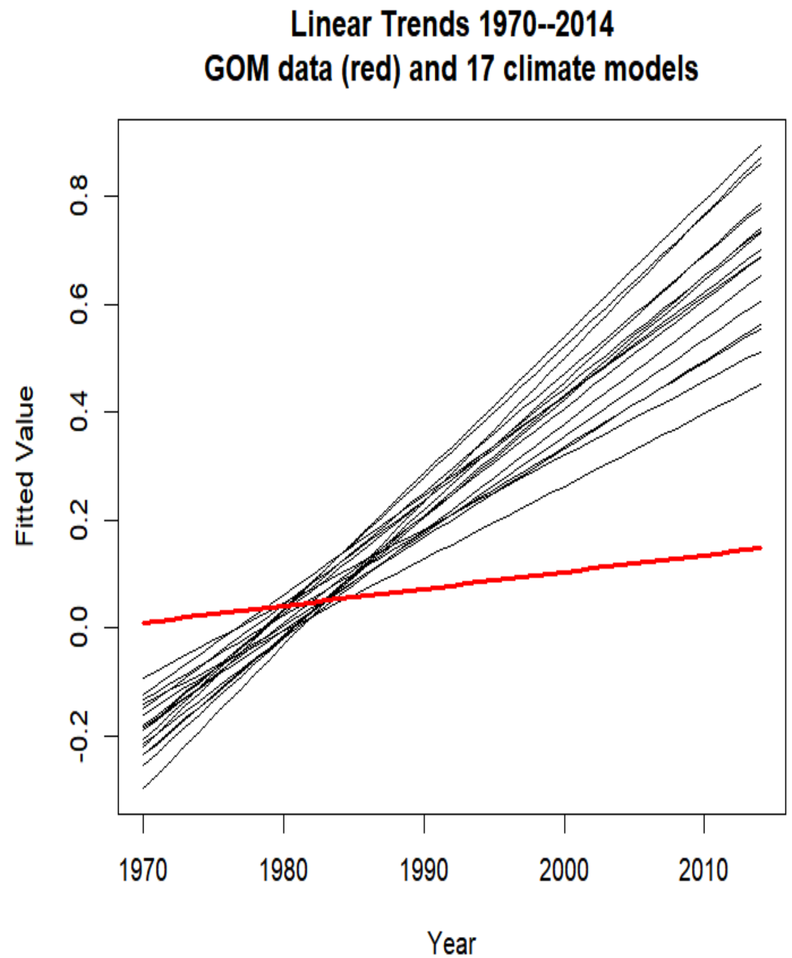
Conditional probabilities of exceeding 2017 precip in Houston:

	(i) 1901–1950	(ii) 2001–2016	(iii) 2017
Houston Hobby	4.7×10^{-5}	0.00014	0.0023
All stations $>$ 70 cm	0.00017	0.00030	0.0023

Still haven't introduced climate models into the discussion

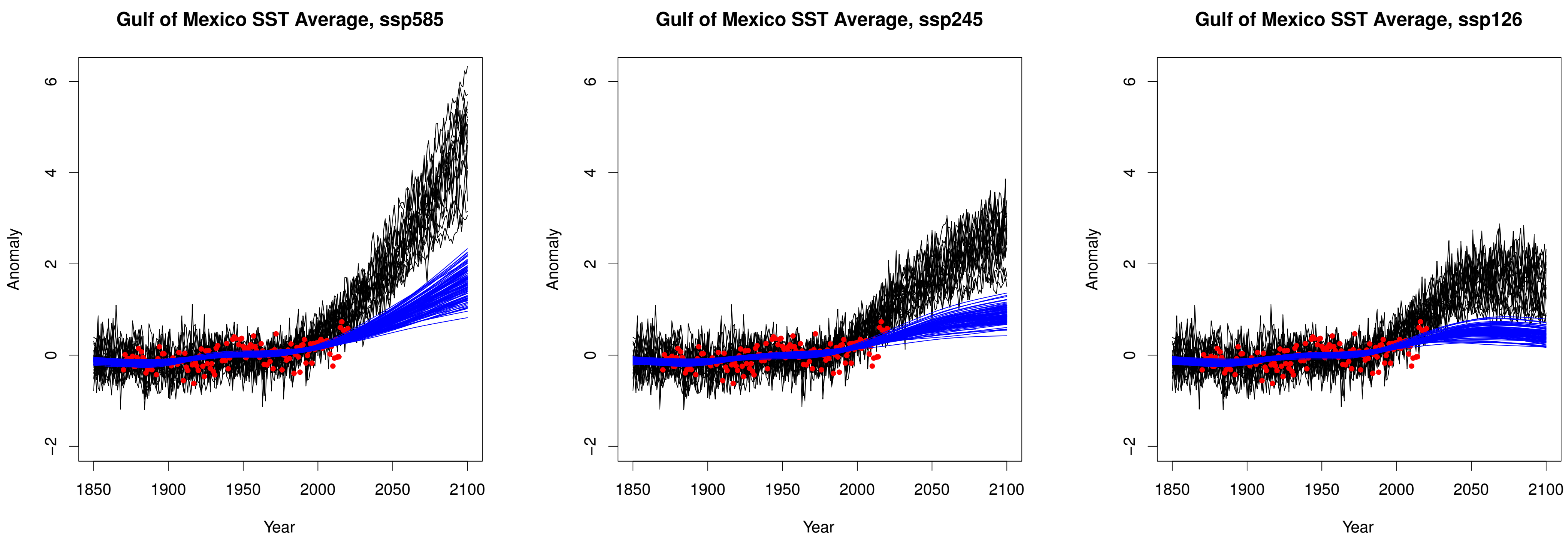
IIc: Projecting the Distribution of the Regional Variable Forwards and Backwards in Time



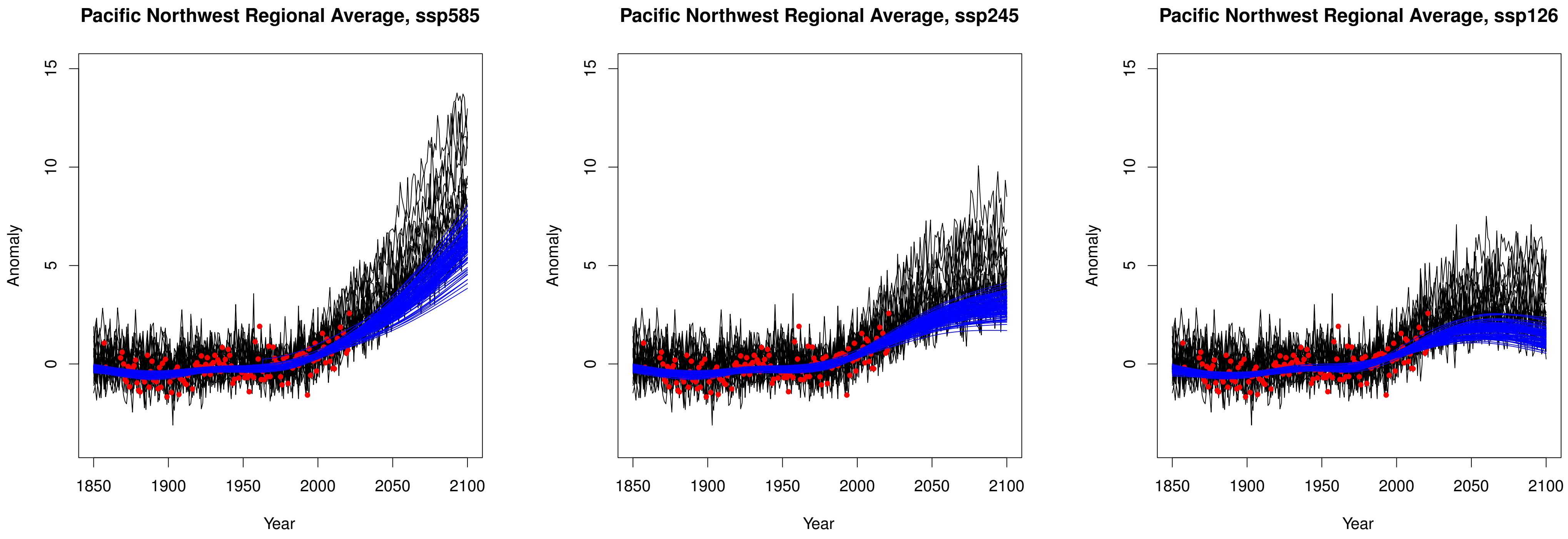


- Obvious method: regress observed regional value on 17 climate models, then use standard prediction theory
  - Objection: ignores variability in the covariates (climate model)
- To accommodate this feature, we need a model for the joint error distribution of 17 climate models. They are not independent!
- Typical solution: use principal components (empirical orthogonal functions), but it's not clear how to accommodate variability in the PCs (side note: Katzfuss, Hammerling and Smith (2017, GRL) proposed a Bayesian solution to detection and attribution, but did not resolve this question)
- Alternative: factor analysis (FA) instead of PCs
- FA models are based on unobserved latent components, easy to implement via Gibbs sampling (don't need Metropolis)
- But..... still susceptible to overfitting, possible lack of proprietary of posterior distribution
- I have avoided these issues by using a "shrinkage prior" formulation of Bhattacharya and Dunson (2011), allows arbitrarily many factors (I actually used 2)

Regional Variable Projections: Gulf of Mexico

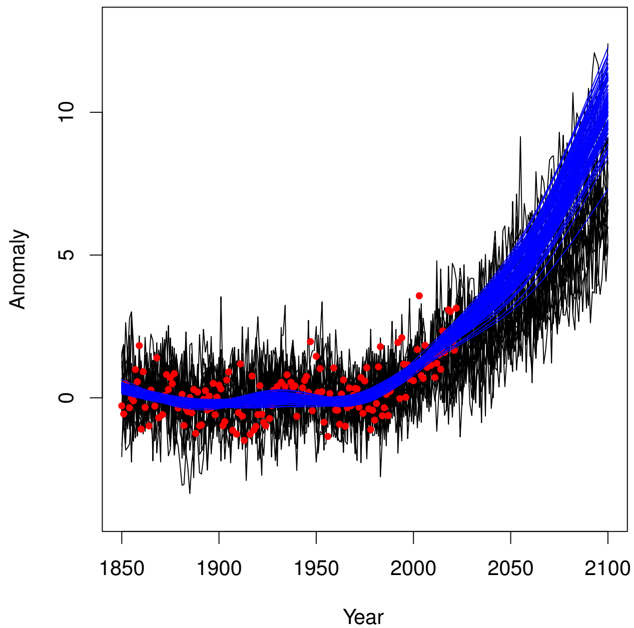


Regional Variable Projections: Pacific Northwest

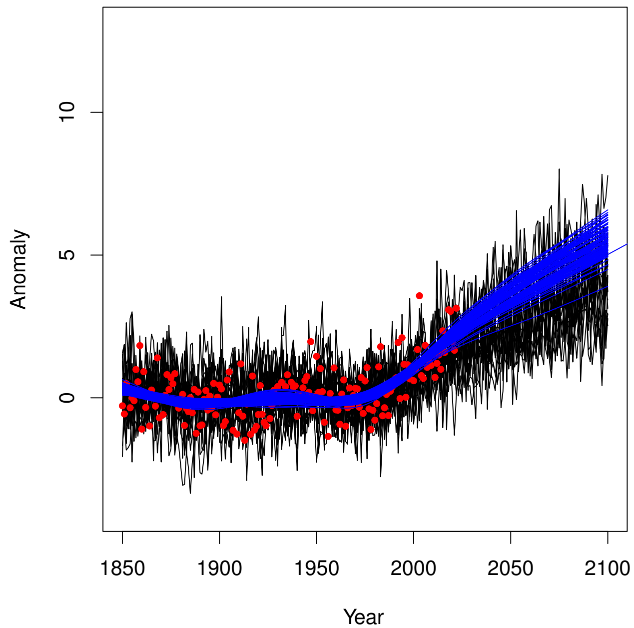


Regional Variable Projections: Northern Europe

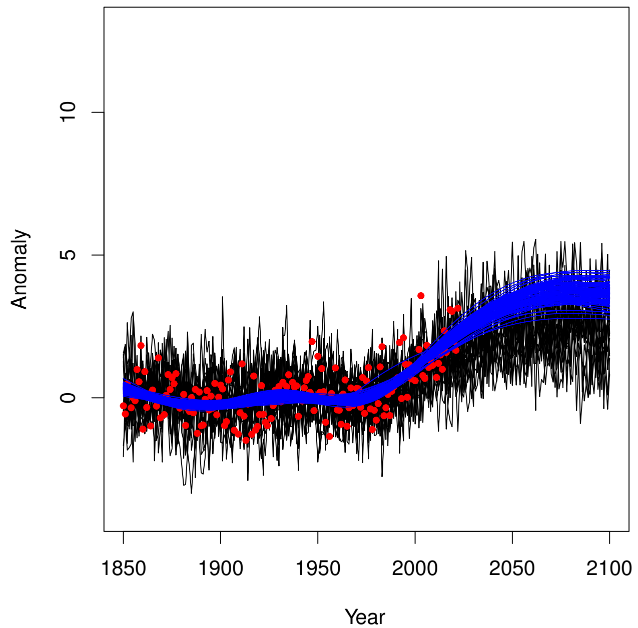
Northern Europe Regional Average, ssp585



Northern Europe Regional Average, ssp245



Northern Europe Regional Average, ssp126

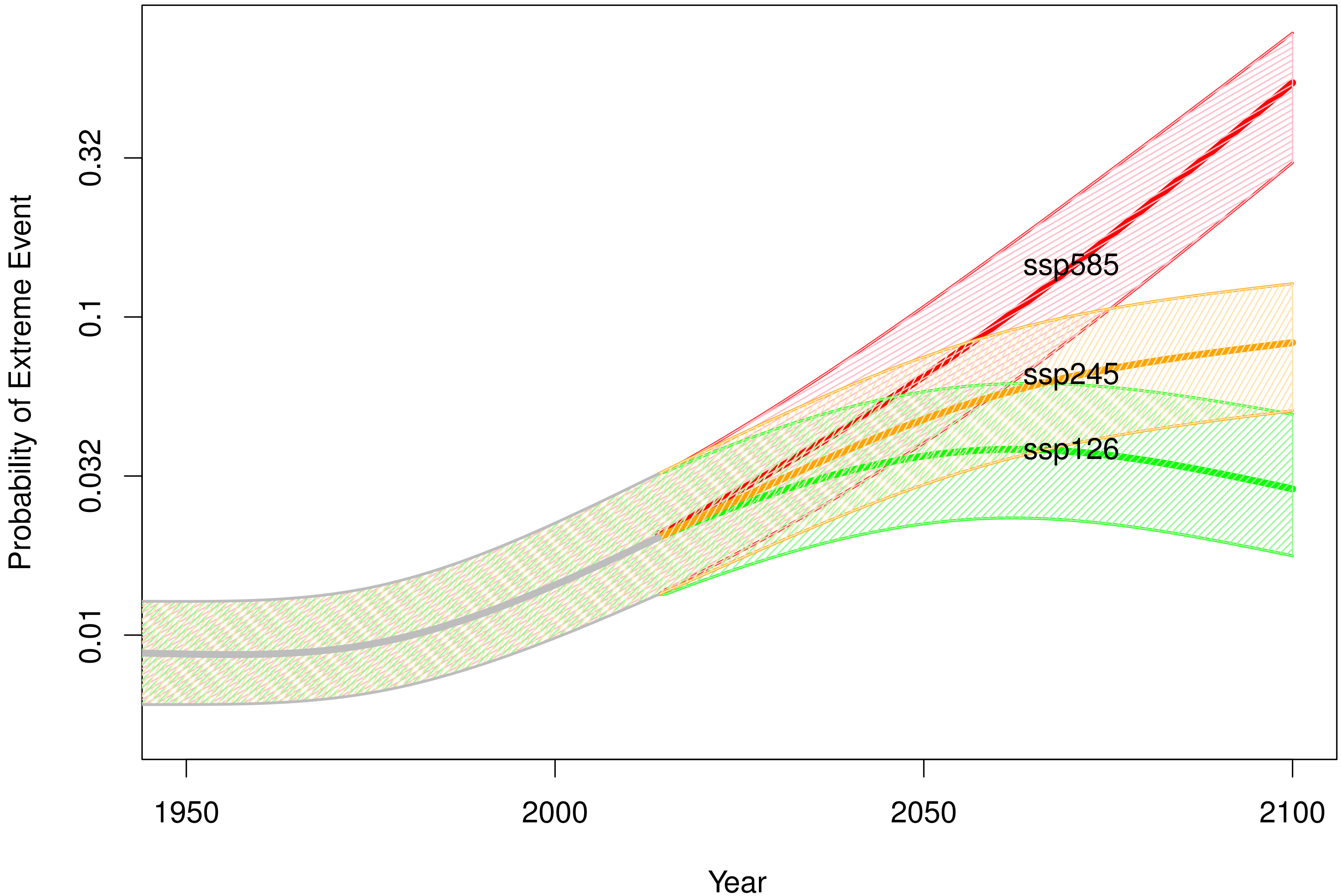


IId: End to End Analysis

- Generate Monte Carlo sample for regional variable condition on climate models
- Conditional on the regional variable, use the spatial GEV model to simulate values of the exceedance probabilities
- Compute 66% prediction intervals (“likely” in IPCC terminology)
- Plot the results

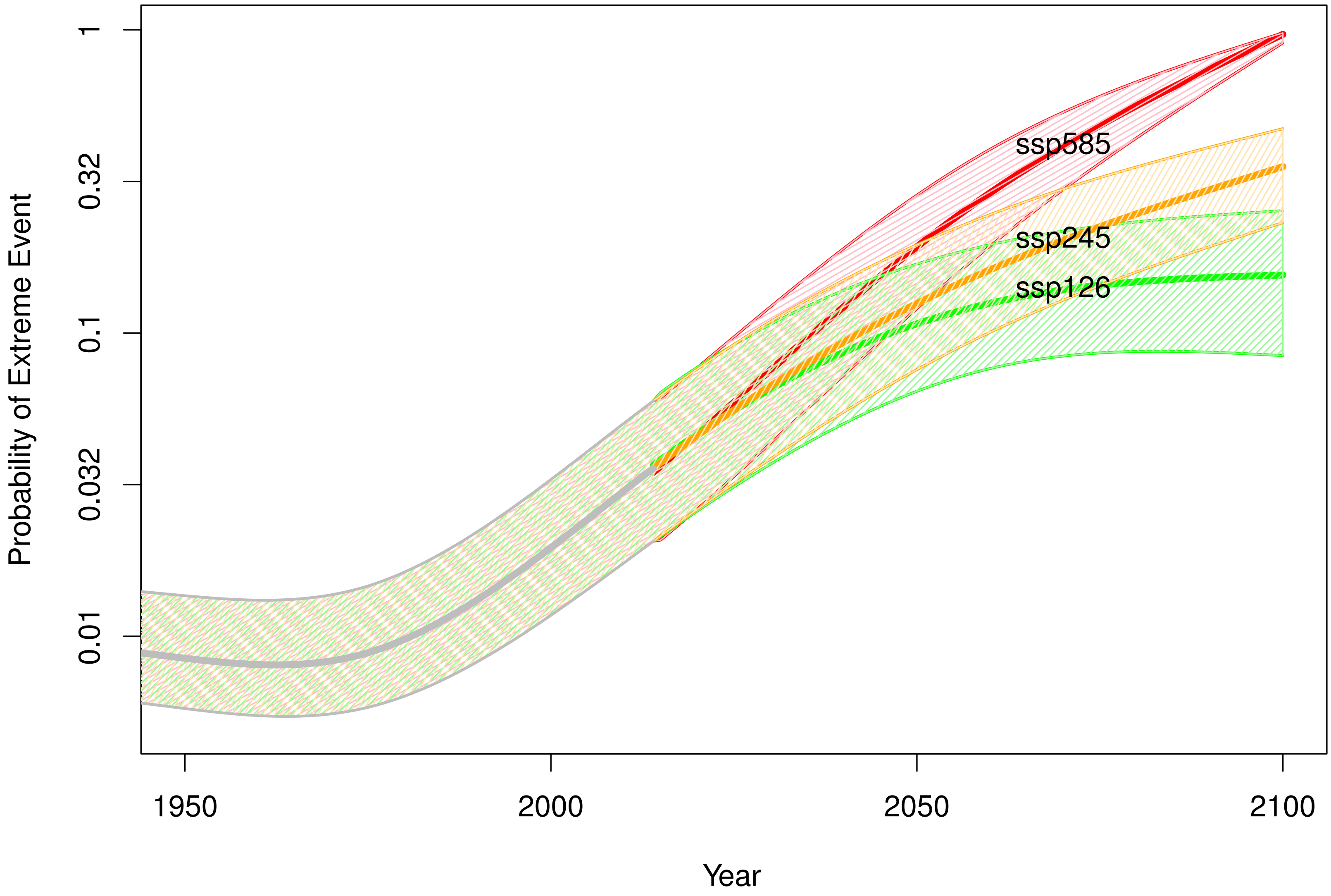
End To End Analysis: Mean Probability of Exceeding 2021 Value for All Stations in Canada

Mean probability over 1850–1949: 0.008; for 2023: 0.025;
for 2080: (0.035, 0.072, 0.22) under three scenarios; for 2100: (0.029, 0.083, 0.54)



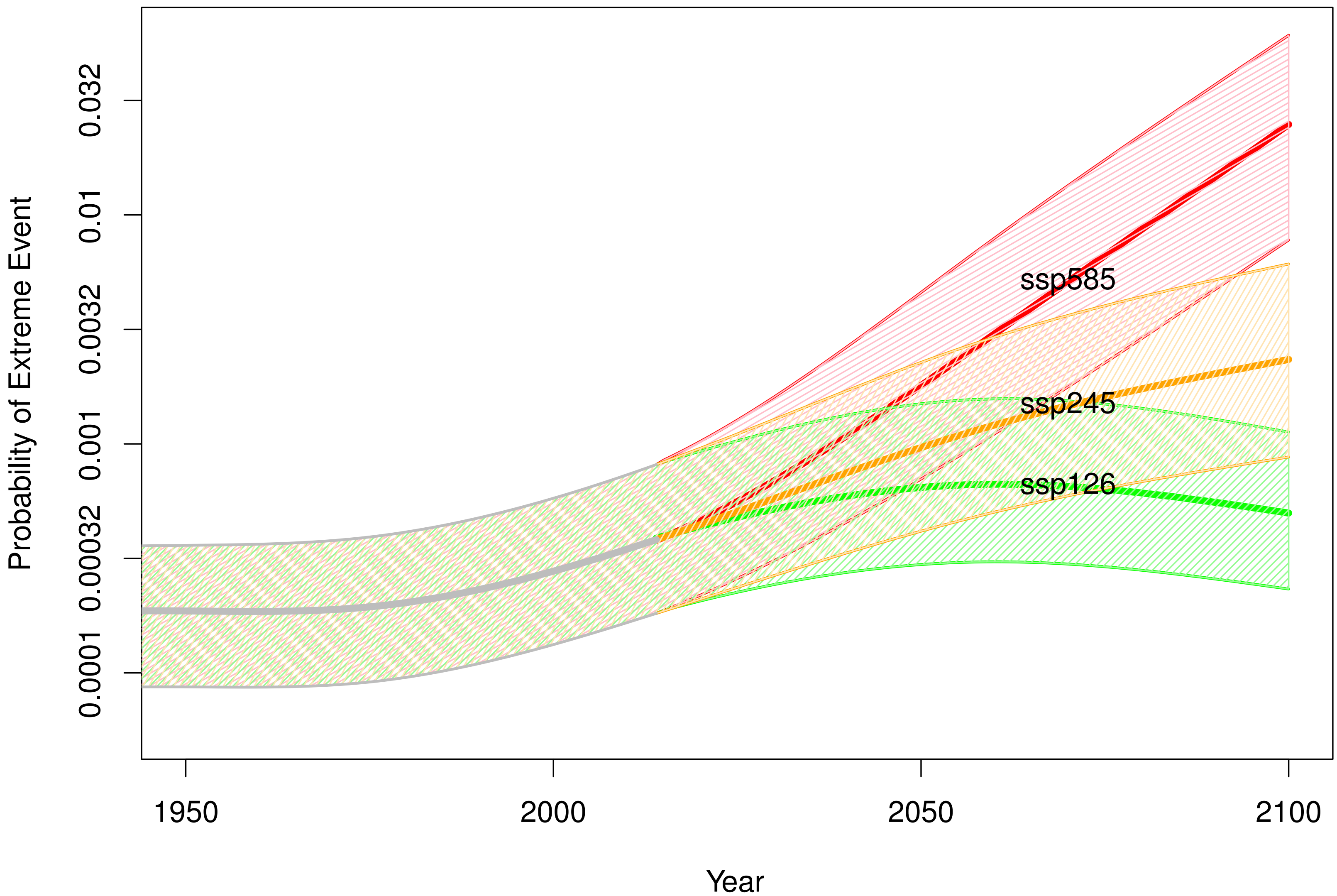
End To End Analysis: Mean Probability of Exceeding 2022 Value for All Stations in U.K.

Mean probability over 1850–1949: 0.008; for 2023: 0.052;
for 2080: (0.15, 0.25, 0.56) under three scenarios; for 2100: (0.16, 0.35, 0.97)



End To End Analysis: Mean Probability of Exceeding 2017 Value for 8 Stations near Houston

Mean probability over 1850–1949: 0.00015; for 2023: 0.00048;
for 2080: (0.00061, 0.0017, 0.0086) under three scenarios;
for 2100: (0.0005, 0.0023, 0.024)



III: Conclusions and Policy Implications

- We have only considered three scenarios for the future, and there are many others, but the analysis demonstrates that there is a *huge* difference among the scenarios for projected probabilities of future extreme events

- Calculation of confidence/prediction/credible intervals is a key point of this analysis. We need to *quantify uncertainty*

- The important caveat: this analysis still relies on statistical assumptions that are not directly verifiable. We need a range of alternative approaches in order to demonstrate that the qualitative conclusions are not dependent on one particular method of analysis.

Slides and datasets: http://rls.sites.oasis.unc.edu/ClimExt/intro.html