A CONDITIONAL APPROACH TO EXTREME EVENT ATTRIBUTION

Richard L. Smith

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

EVA 2023, Milan, June 29, 2023

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

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

The Guardian, June 27, 2023

Current heatwave across US south made five times more likely by climate crisis

Latest 'heat dome' event over Texas and Louisiana, plus much of Mexico, driven by human-cause climate change, scientists find

A temperature display reading 99F (about 37.2C) in late afternoon in Houston, Texas, at the weekend. Photograph: Xinhua/Shutterstock

The record heatwave roiling parts of Texas, Louisiana and Mexico was made at least five times more likely due to human-caused climate change, scientists have found, marking the latest in a series of recent extreme "heat dome" events that have scorched various parts of the world.

Outline

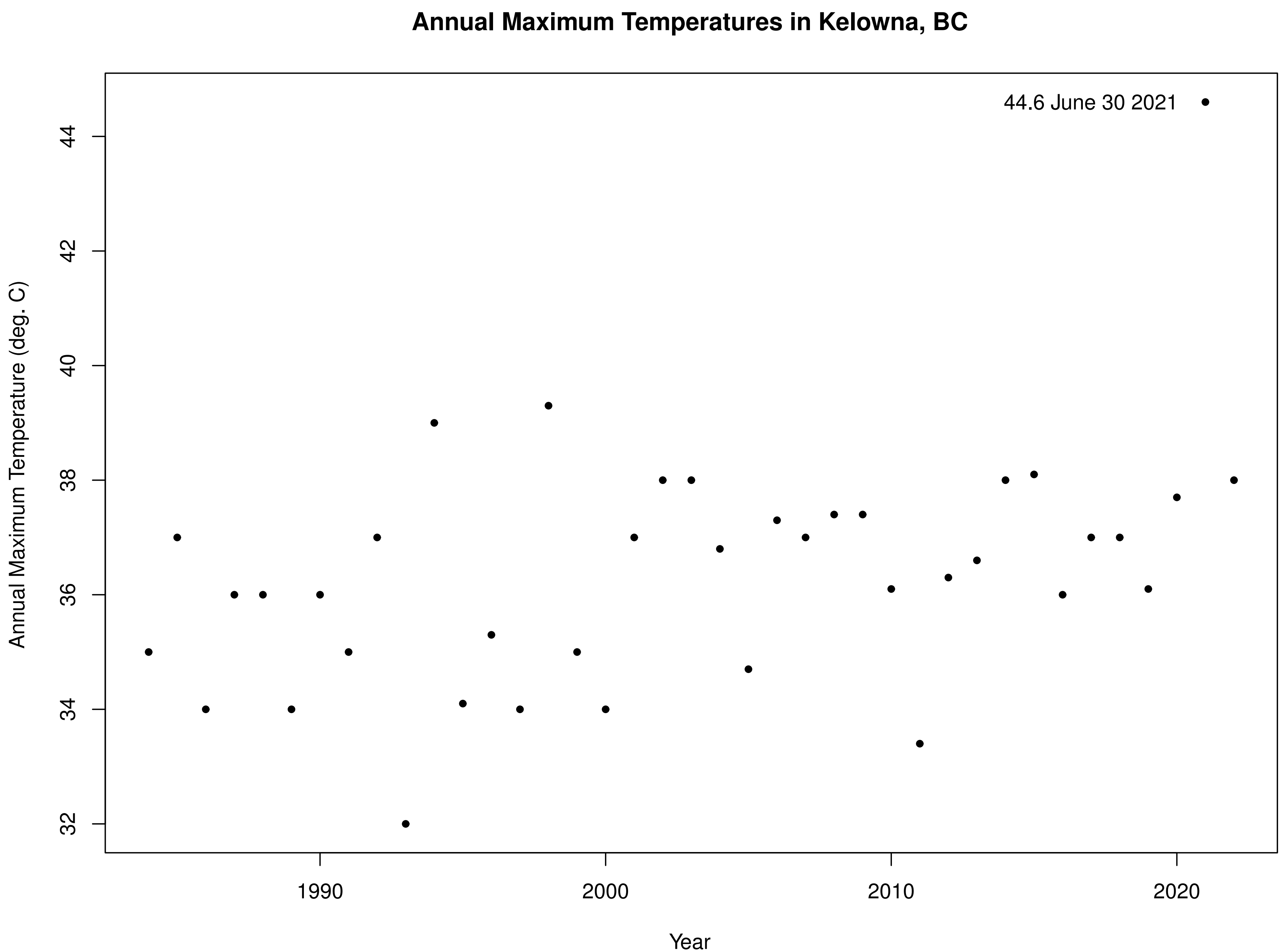
I. Introduction and Data

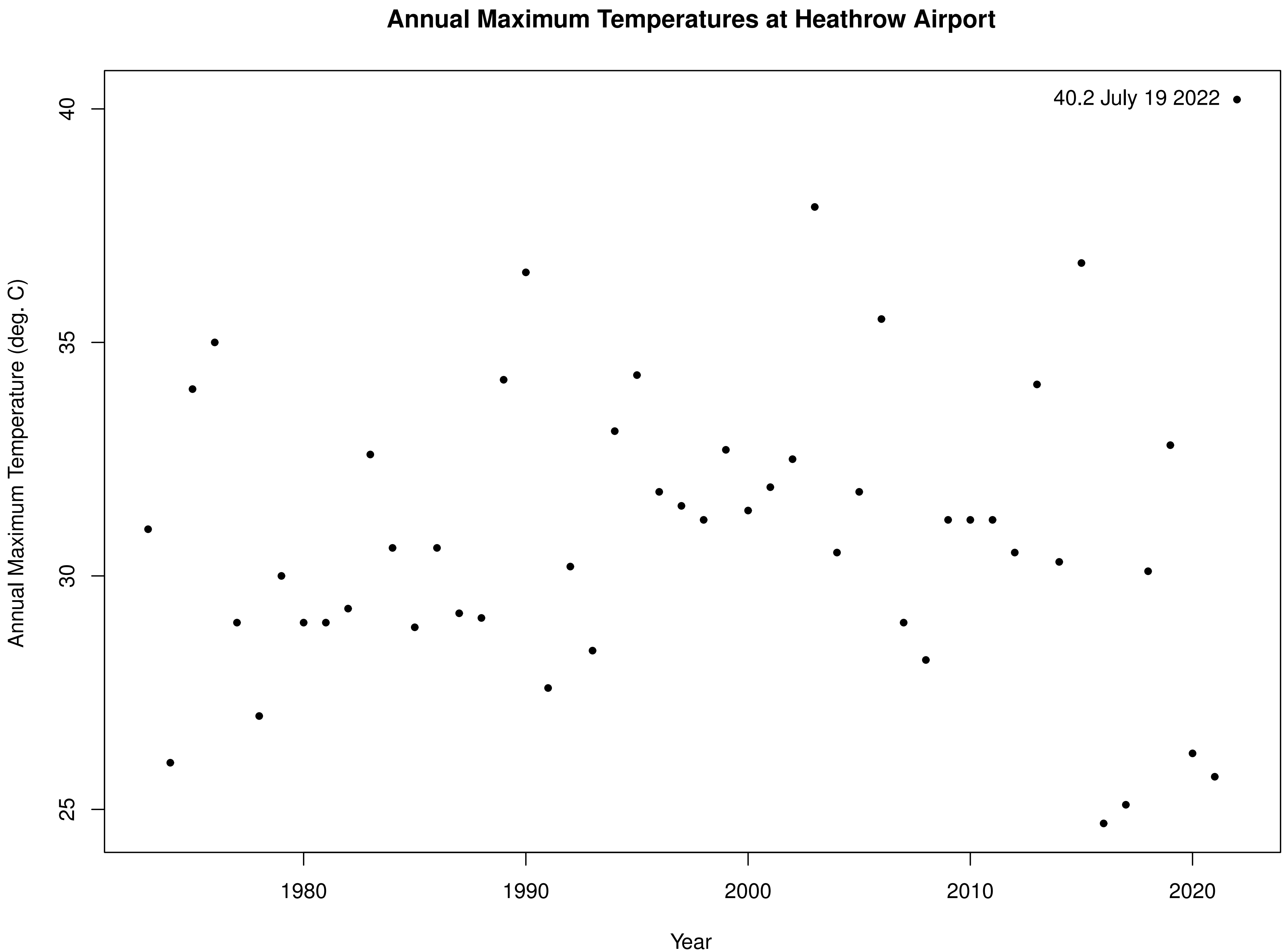
II. Statistical Analysis

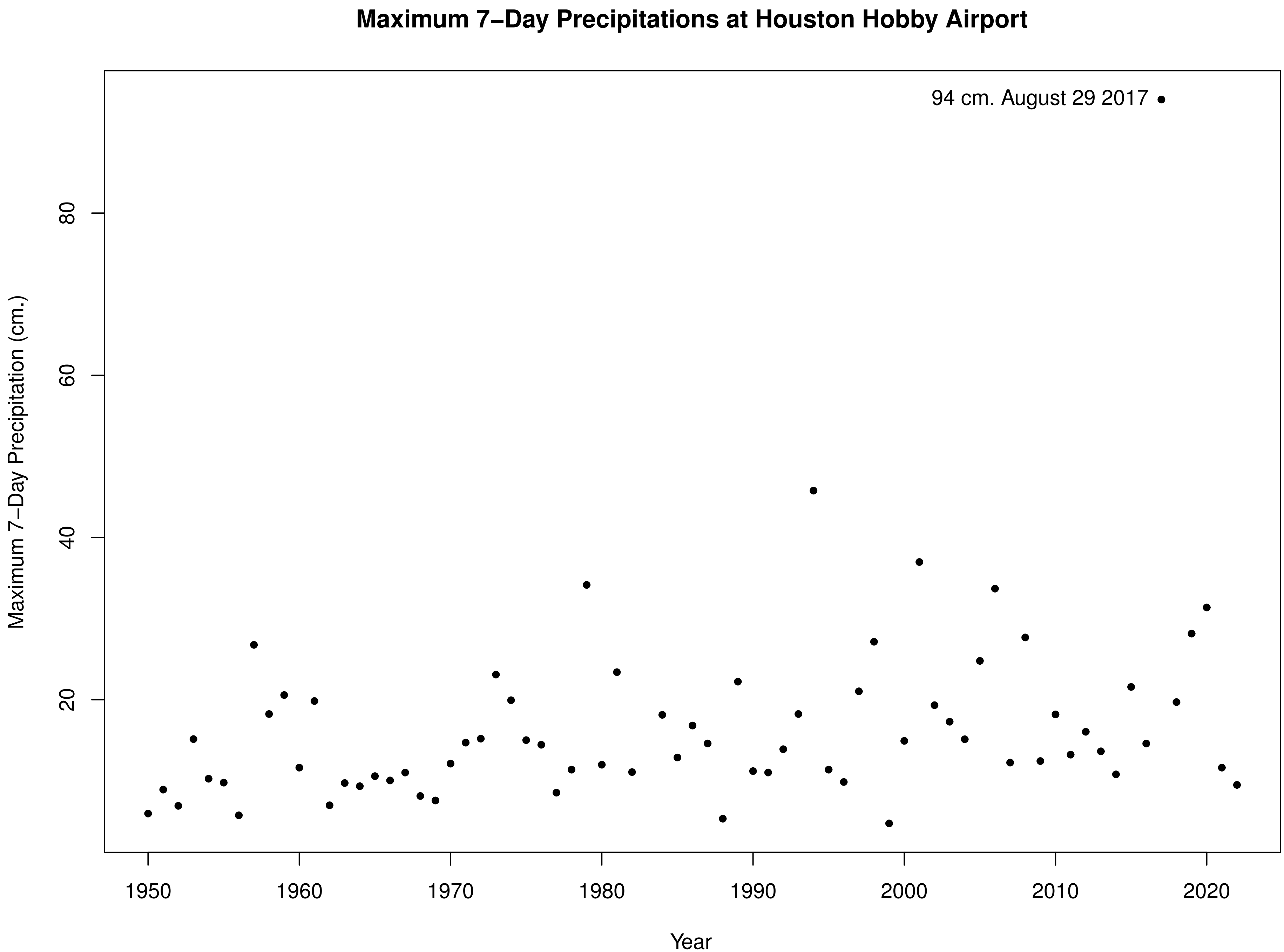
III. Conclusions and Policy Implications

I. Introduction

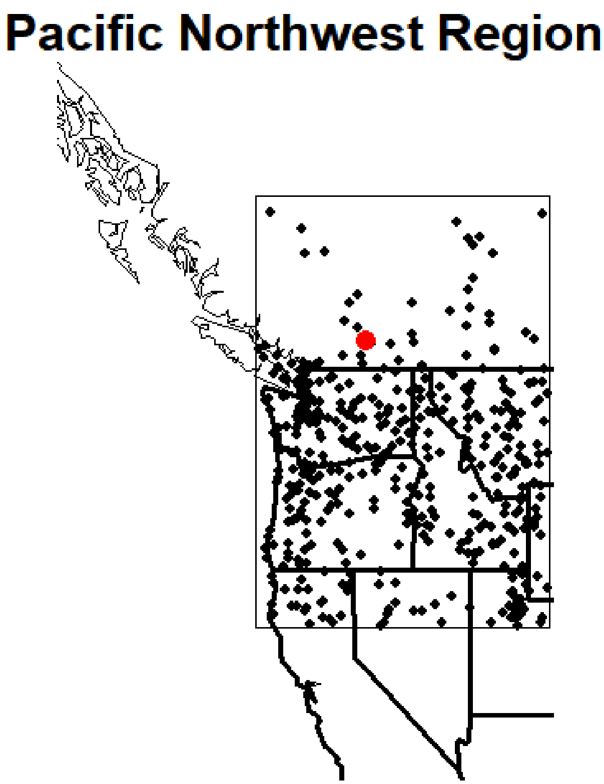
I begin with three examples of datasets that contain extreme events

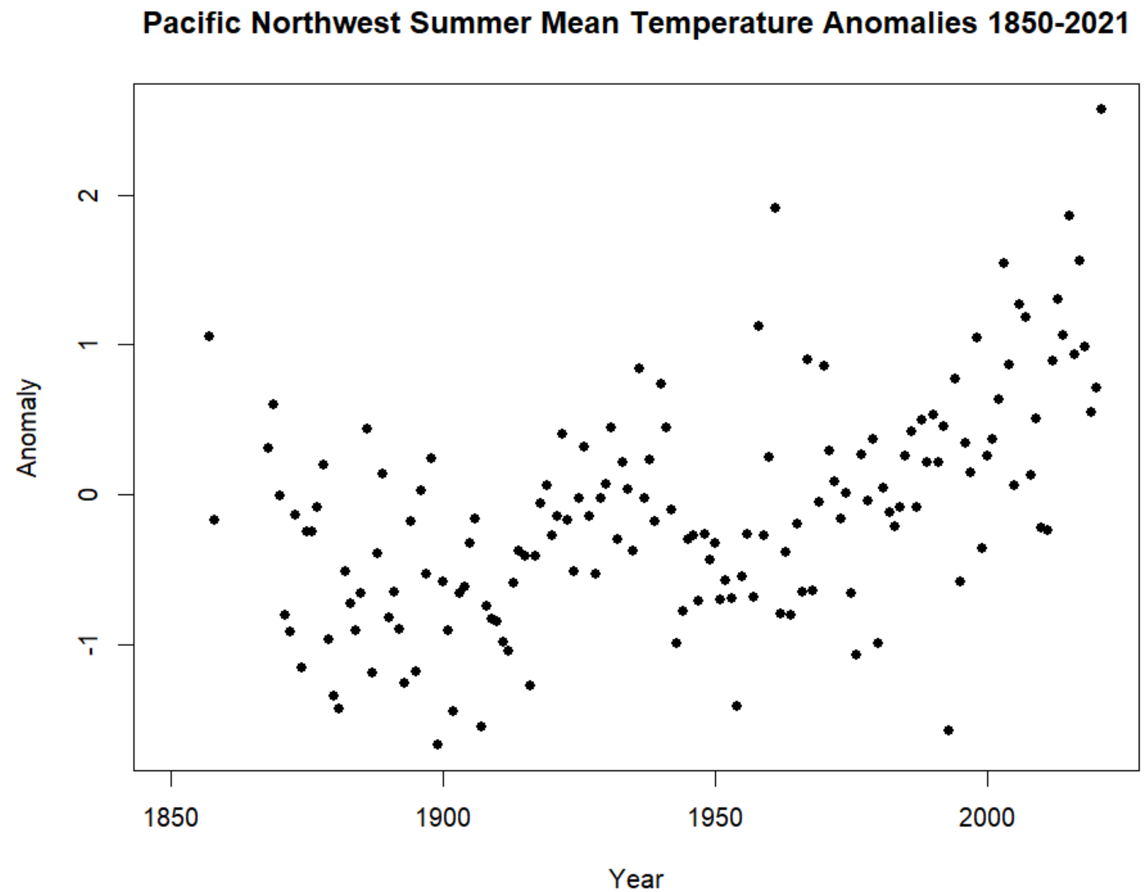

Annual Maximum Temperatures in Kelowna, BC
44.6 June 30 2021
Annual Maximum Temperature (deg. C)
32
34
36
38
40
42
44
1990
2000
2010
2020
Year

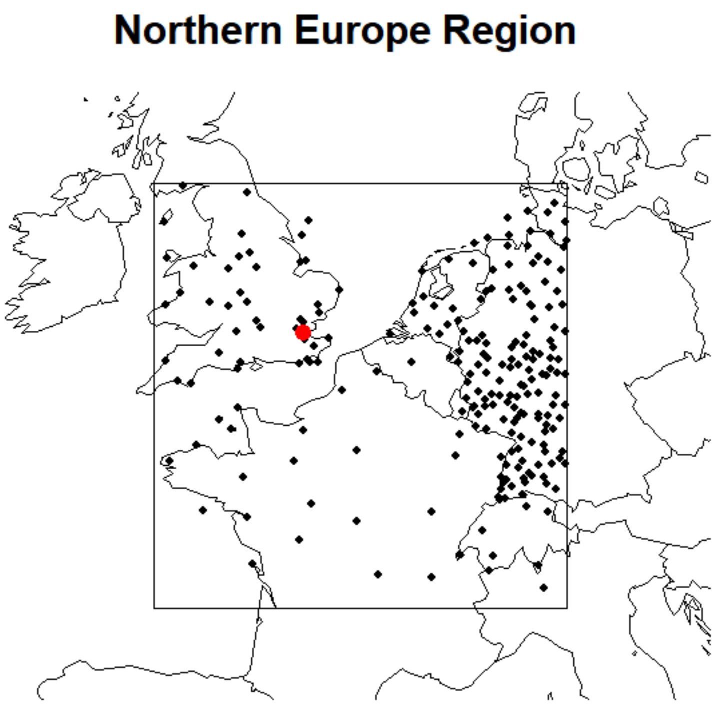

Annual Maximum Temperatures at Heathrow Airport
40.2 July 19 2022
Annual Maximum Temperature (deg. C)
40
35
30
25
1980
1990
2000
2010
2020
Year

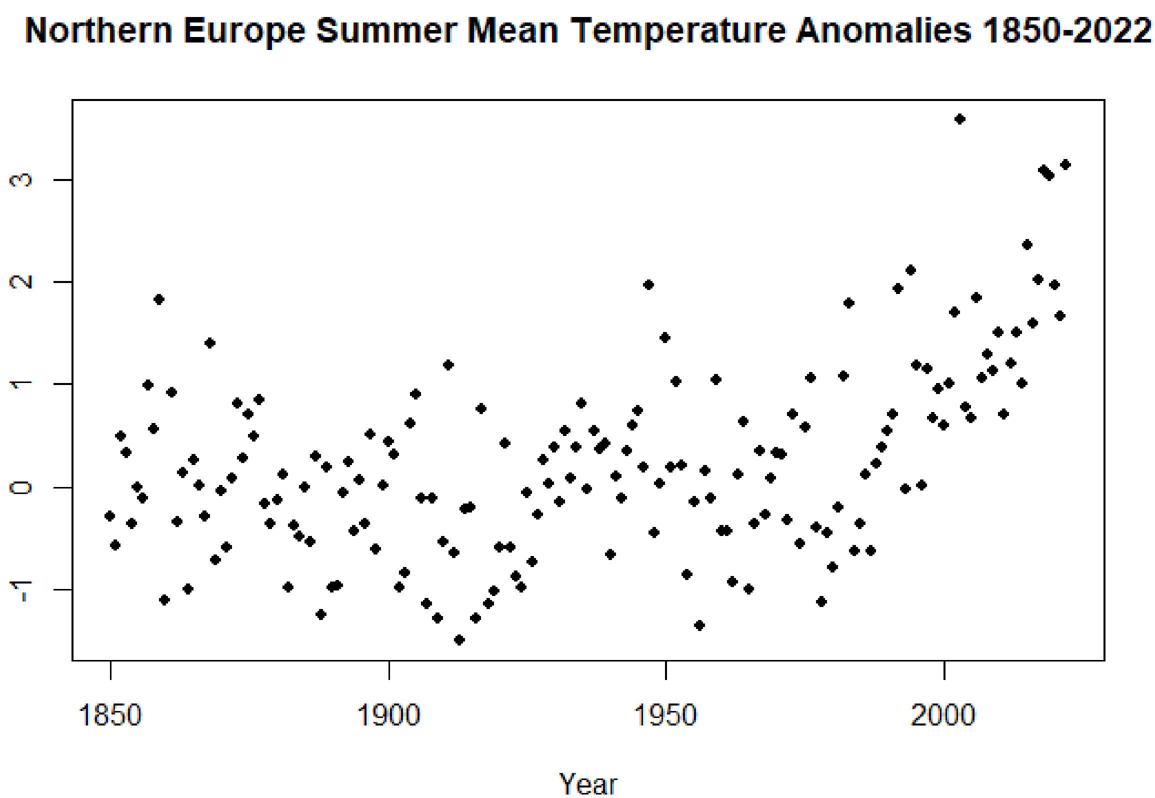

Maximum 7−Day Precipitations at Houston Hobby Airport
94 cm. August 29 2017
Maximum 7−Day Precipitation (cm.)
20
40
60
80
1950
1960
1970
1980
1990
2000
2010
2020
Year

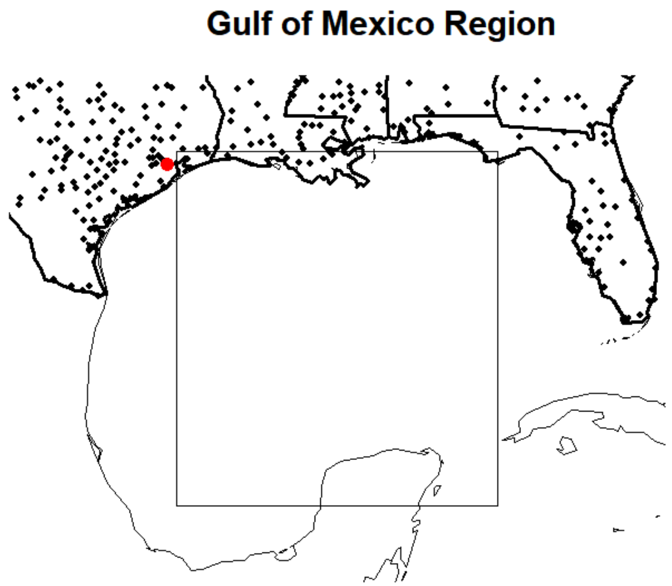
For each of these examples, I have collected weather data from multiple stations in the same region (from the Global Historical Climatological Network), and also calculated a *regional variable* that includes annual or seasonal maxima from spatially aggregated data (from the Climate Research Unit of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

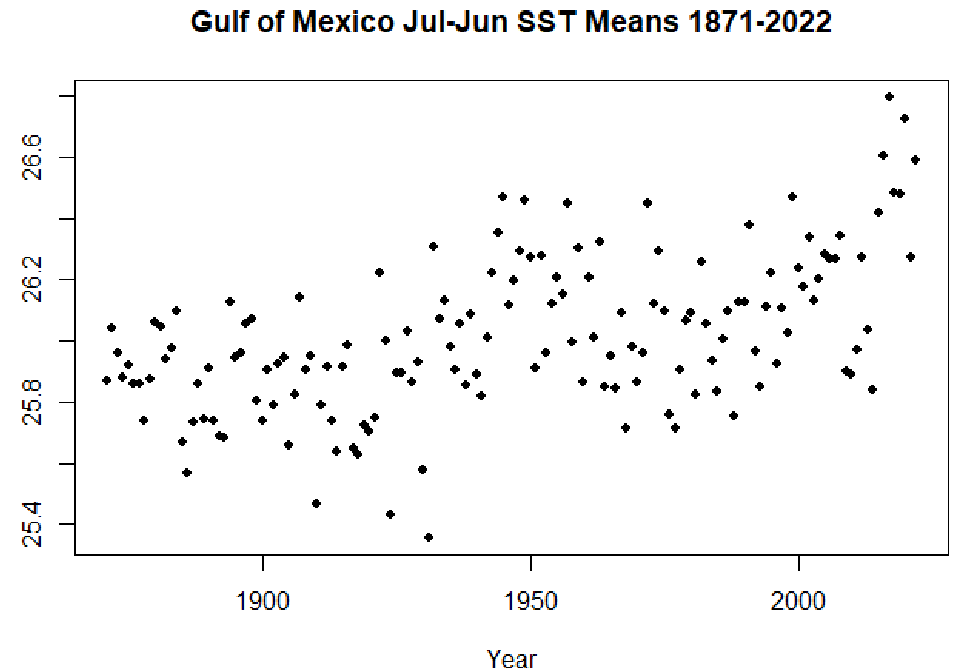

Pacific Northwest Region


Pacific Northwest Summer Mean Temperature Anomalies 1850-2021
Anomaly
Year
-1
0
1
2
1850
1900
1950
2000

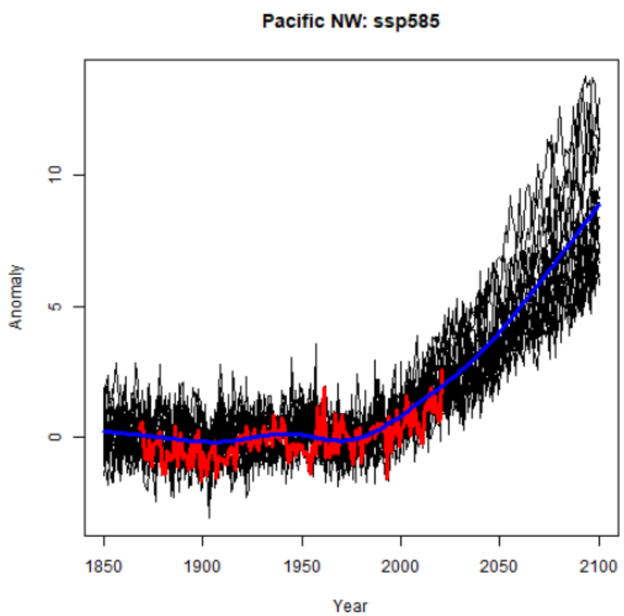

Northern Europe Region

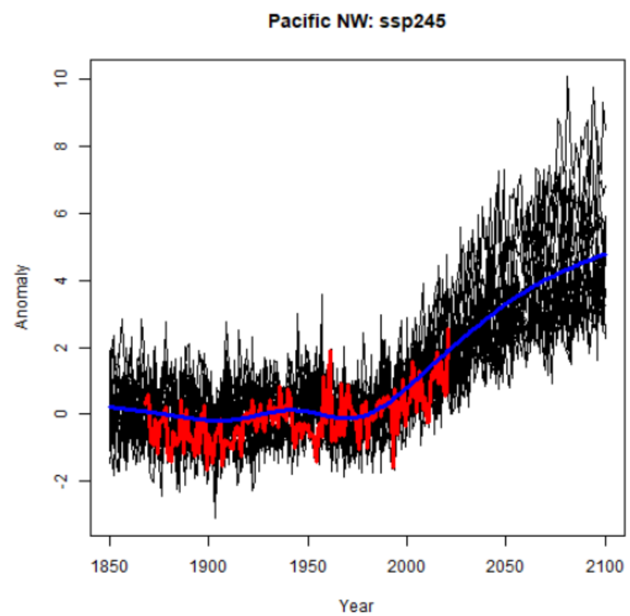

Northern Europe Summer Mean Temperature Anomalies 1850-2022
-1
0
1
2
3
1850
1900
1950
2000
Year

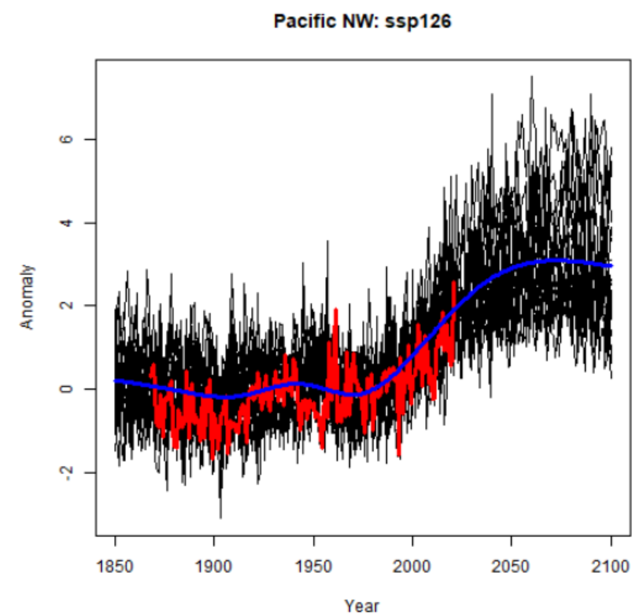

Gulf of Mexico Region

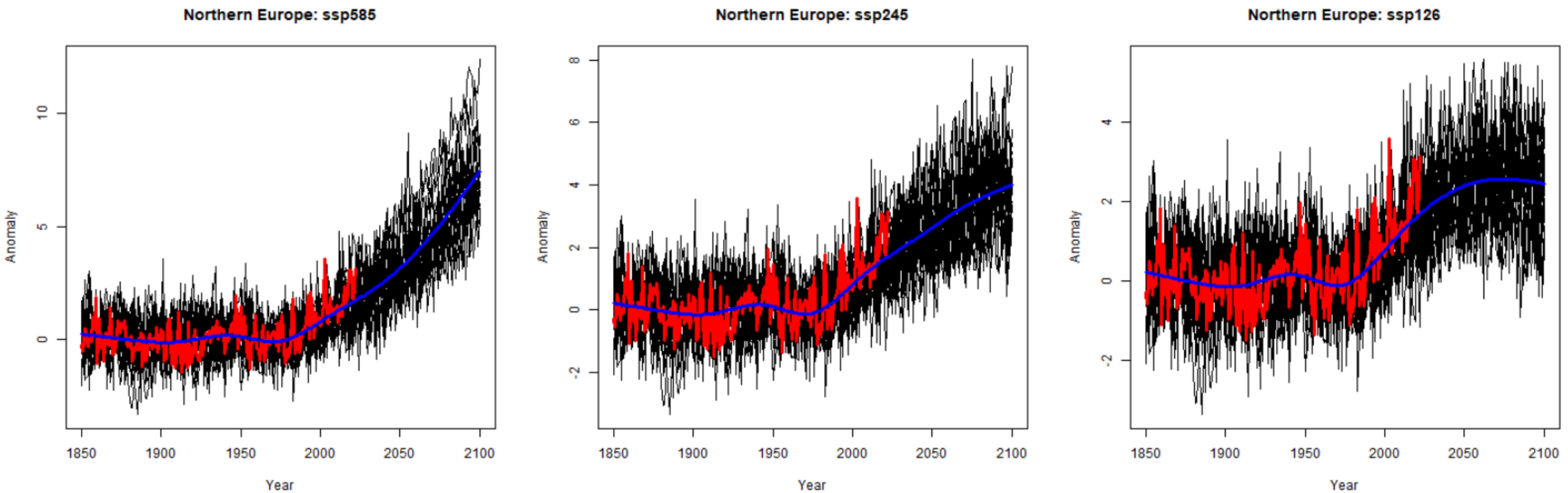

Gulf of Mexico Jul-Jun SST Means 1871-2022
25.4
25.8
26.2
26.6
1900
1950
2000
Year

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

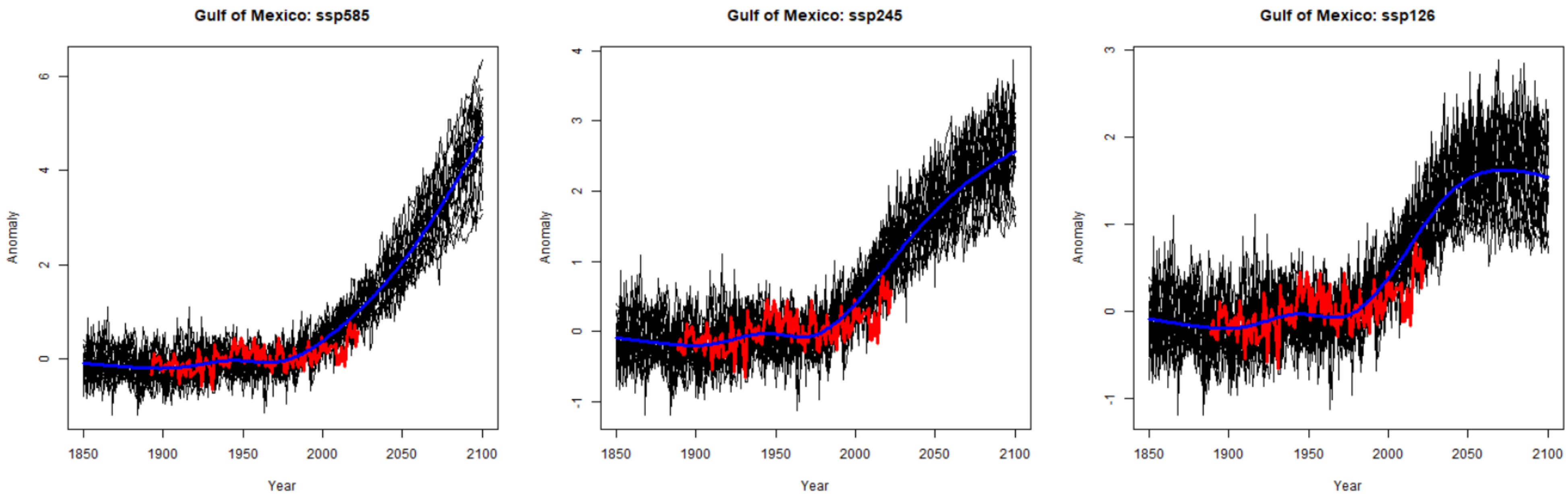

Pacific NW: ssp585
Anomaly
Year
0
5
10
1850
1900
1950
2000
2050
2100


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


Pacific NW: ssp126
Anomaly
Year
-2
0
2
4
6
1850
1900
1950
2000
2050
2100



Northern Europe: ssp585
Anomaly
Year
Northern Europe: ssp245
Anomaly
Year
Northern Europe: ssp126
Anomaly
Year


Gulf of Mexico: ssp585
Gulf of Mexico: ssp245
Gulf of Mexico: ssp126
Anomaly
Year
1850
1900
1950
2000
2050
2100

II. Statistical Analysis Outline of the Approach

IIa. 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

- Fit GEV with covariate to each station

- For each of our three examples, we use the regional temperature averages defined earlier as the covariate of interest. Of course, many other choices are possible.

- 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.

GEV Distribution with a Covariate

Data: $Y_{s,t}$ is annual maximum temperature or precipitation in year t at location s

$$\Pr\left\{Y_{s,t} \le y\right\} = \exp\left\{-\left(1+\xi_{s,t}\frac{y-\mu_{s,t}}{\psi_{s,t}}\right)_+^{-1/\xi_{s,t}}\right\}$$

where $x_+ = \max(x,0)$ and $\mu_{s,t}, \psi_{s,t}, \xi_{s,t}$ are location, scale and shape parameters. Notation: $\mathrm{GEV}(\mu_{s,t}, \psi_{s,t}, \xi_{s,t})$.

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

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

where $\boldsymbol{\Theta}_s = \begin{pmatrix} \theta_{s,1} & \ldots & \theta_{s,5} \end{pmatrix}$ is 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)

Application to Kelowna, B.C.

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)

Application to Heathrow, London

5-Par Model:

Parameter	Estimate	SE	t-val	p-val
θ_1	29.0626	0.2947	98.6323	0.0000
θ_2	0.5502	0.1312	4.1922	0.0000
θ_3	-0.1041	0.0988	-1.0538	0.2920
θ_4	2.0747	0.2224	9.3281	0.0000
θ_5	-0.0038	0.0948	-0.0399	0.9682

MLE probability of exceeding 40.2°C in 2022, given X_{2022}: 0.041

Bayes posterior mean probability is 0.068.

6-Par Model:

Parameter	Estimate	SE	t-val	p-val
θ_1	28.9863	0.2870	101.0043	0.0000
θ_2	0.5131	0.1289	3.9795	0.0001
θ_3	-0.0130	0.2757	-0.0472	0.9624
θ_4	2.1770	0.2192	9.9295	0.0000
θ_5	0.0283	0.0926	0.3061	0.7595
θ_6	-0.2591	0.2101	-1.2331	0.2175

MLE probability of exceeding 40.2°C in 2022, given X_{2022}: 0.0008

Bayes posterior mean probability is 0.043.

Application to Houston Hobby Airport

5-Par Model:

Parameter	Estimate	SE	t-val	p-val
θ_1	11.6179	0.6408	18.1313	0.0000
θ_2	1.4864	0.1142	13.0167	0.0000
θ_3	0.0963	0.0943	1.0208	0.3074
θ_4	8.0246	2.6788	2.9956	0.0027
θ_5	1.4649	0.4975	2.9444	0.0032

MLE probability of exceeding 94 cm. in 2017, given X_{2017}: 0.013

Bayesian probability: 0.029

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
  - Originated by Coles and Casson (1999)
  - Further developed by Cooley and many others
  - A key issue: spatial correlations in residuals
  - Extension to multivariate spatial processes: analysis here follows a paper by Russell, Risser, Smith and Kunkel (2020)
- Direct modeling of spatial dependence using a max-stable process or alternatives (Zhang, Risser, Wehner and O'Brien; Hector and Reich 2023; Zhang, Shaby and Wadsworth 2020; Huser and Wadsworth 2019; ...)

Concept of Approach

- True (latent) process $\mathbf{\Theta}$ (KN-dimensional, K=5 or 6)
- Estimated process $\hat{\mathbf{\Theta}}$ (GEV estimates at each site)
- Assume $(\hat{\mathbf{\Theta}} \mid \mathbf{\Theta}) \ \sim \ \mathcal{N}_{KN}(\mathbf{\Theta}, W)$
- Spatial model $(\mathbf{\Theta} \ \sim \ \mathcal{N}_{KN}(\boldsymbol{\mu} \otimes I_N, V(\phi)))$
- $\hat{\mathbf{\Theta}} \ \sim \ \mathcal{N}_{KN}(\boldsymbol{\mu} \otimes I_N, V(\phi) + W)$
- Estimate W empirically, $\boldsymbol{\mu}$ and ϕ by MLE
- Hence generate $\mathbf{\Theta} \mid \hat{\mathbf{\Theta}}$
- Model for $V(\phi)$: *co-regionalization* (Wackernagel 2003, Finley et al. 2008, etc.)

Results: Kelowna (6-Par)

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

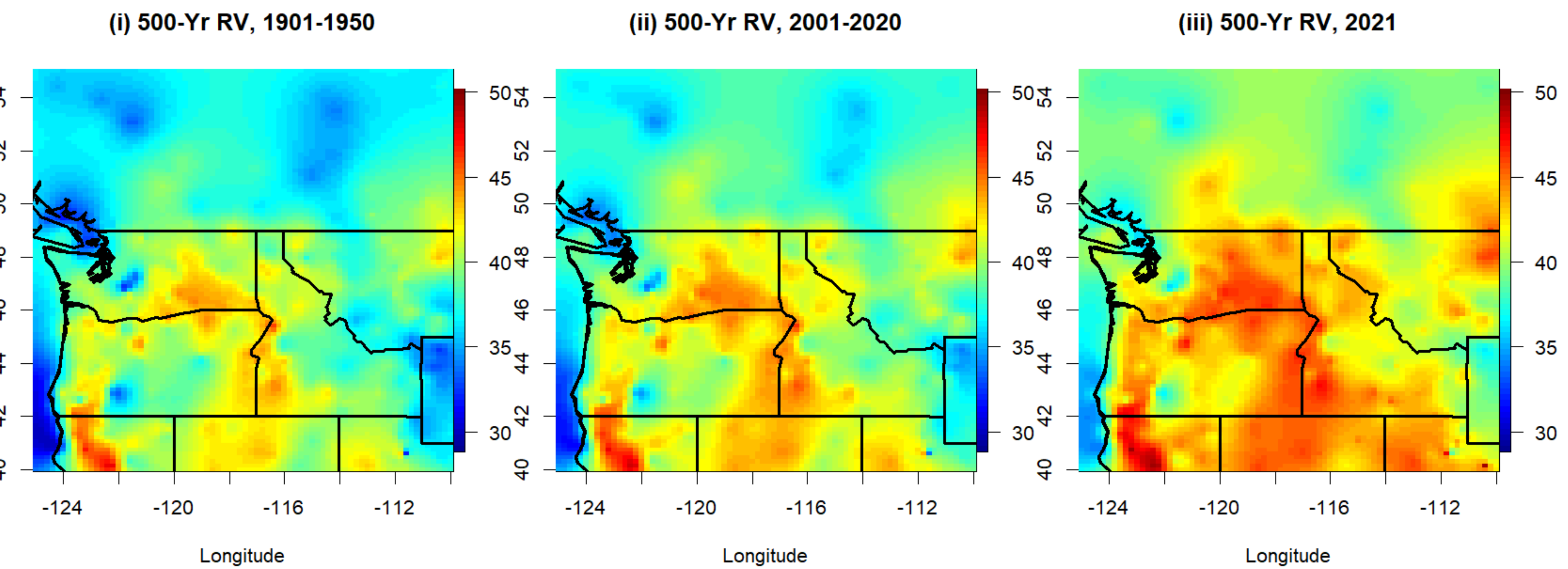
Spatial Analysis

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

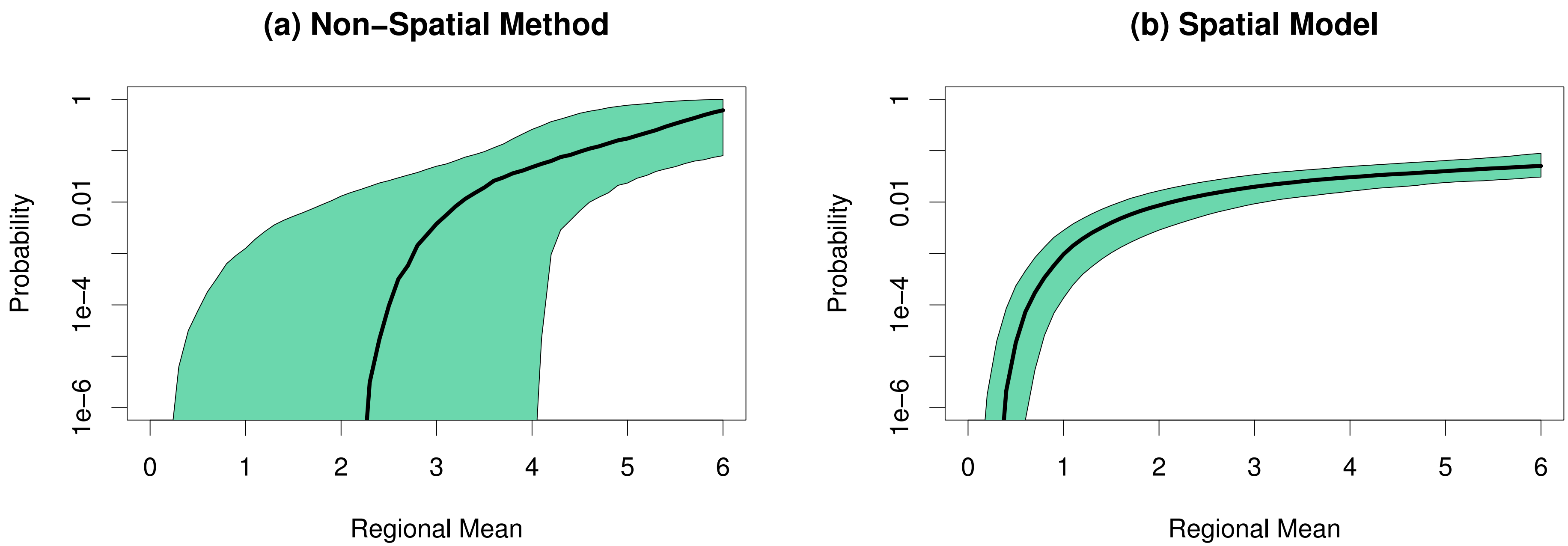
Conditional probabilities of exceeding the 2021 value:

	(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

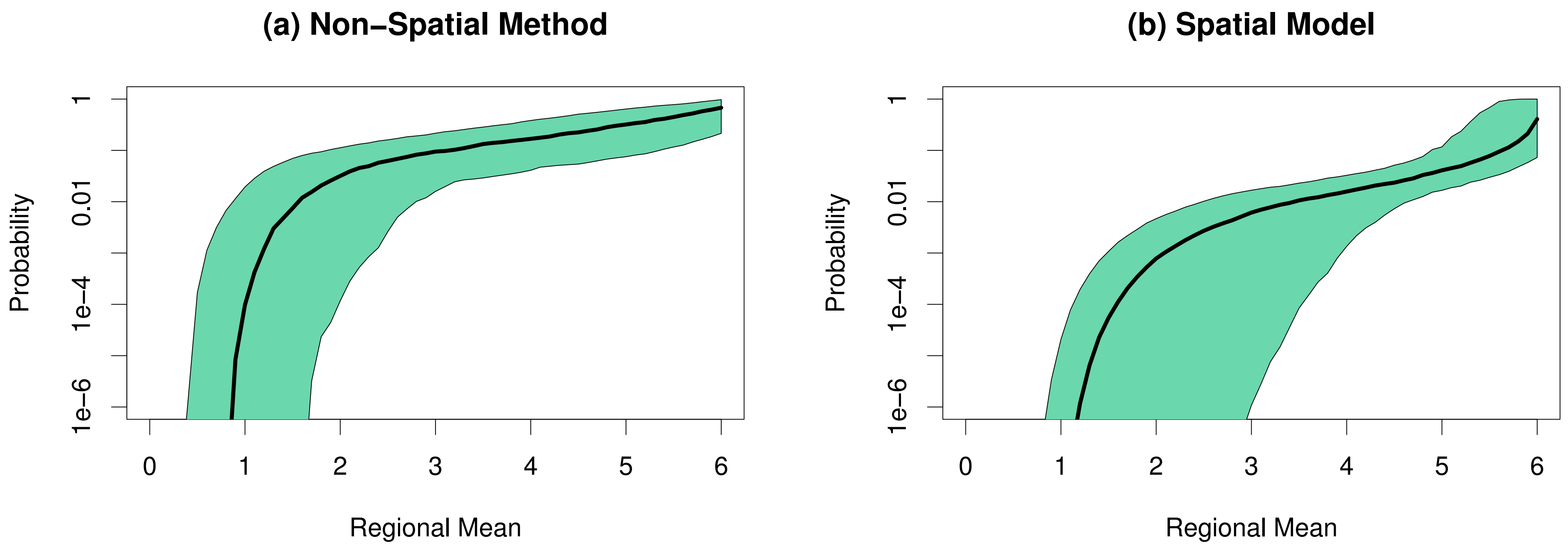
Map of 500-year return values for (i), (ii), (iii)



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)



Results: Heathrow (6-Par)

MLE Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	28.9863	0.2870	101.0043	0.0000
θ_2	0.5131	0.1289	3.9795	0.0001
θ_3	-0.0130	0.2757	-0.0472	0.9624
θ_4	2.1770	0.2192	9.9295	0.0000
θ_5	0.0283	0.0926	0.3061	0.7595
θ_6	-0.2591	0.2101	-1.2331	0.2175

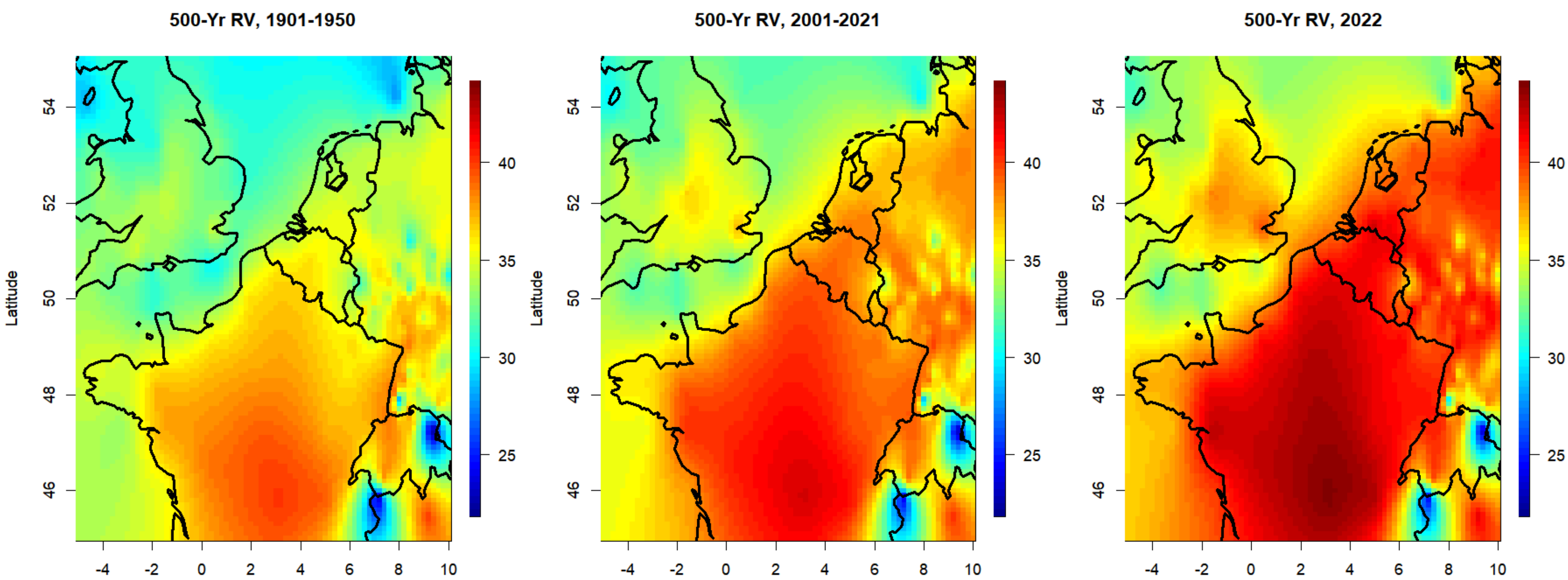
Spatial Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	29.5414	0.1691	174.6576	0.0000
θ_2	0.5271	0.0587	8.9810	0.0000
θ_3	-0.5628	0.0579	-9.7281	0.0000
θ_4	1.7354	0.0964	18.0000	0.0000
θ_5	0.0604	0.0479	1.2610	0.2073
θ_6	-0.0046	0.0710	-0.0644	0.9486

Conditional probabilities of exceeding the 2022 value:

	(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

Map of 500-year return values for (i), (ii), (iii)



Results: Houston (5-Par)

MLE Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	11.6179	0.6408	18.1313	0.0000
θ_2	1.4864	0.1142	13.0167	0.0000
θ_3	0.0963	0.0943	1.0208	0.3074
θ_4	8.0246	2.6788	2.9956	0.0027
θ_5	1.4649	0.4975	2.9444	0.0032

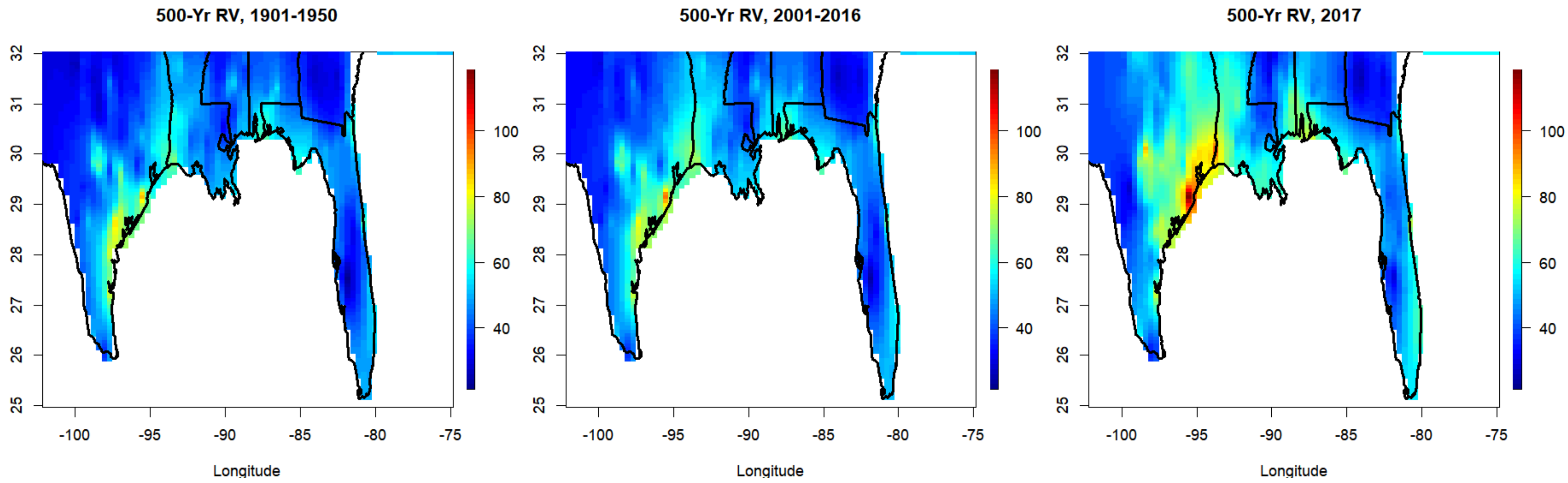
Spatial Analysis

Parameter	Estimate	SE	t-val	p-val
θ_1	11.1723	0.3810	29.3261	0.0000
θ_2	1.5000	0.0662	22.6702	0.0000
θ_3	0.0911	0.0470	1.9373	0.0527
θ_4	4.1167	1.2635	3.2581	0.0011
θ_5	0.8484	0.2410	3.5200	0.0004

Conditional probabilities of exceeding the 2022 value:

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

Map of 500-year return values for (i), (ii), (iii)

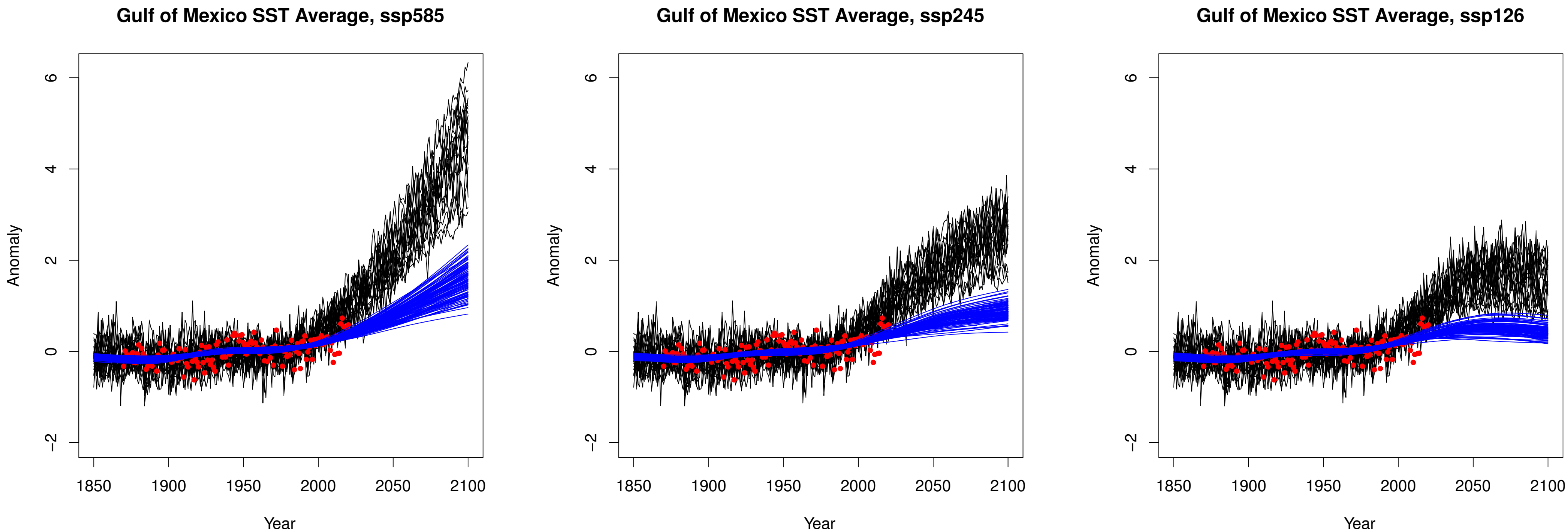


Houston, we have a problem

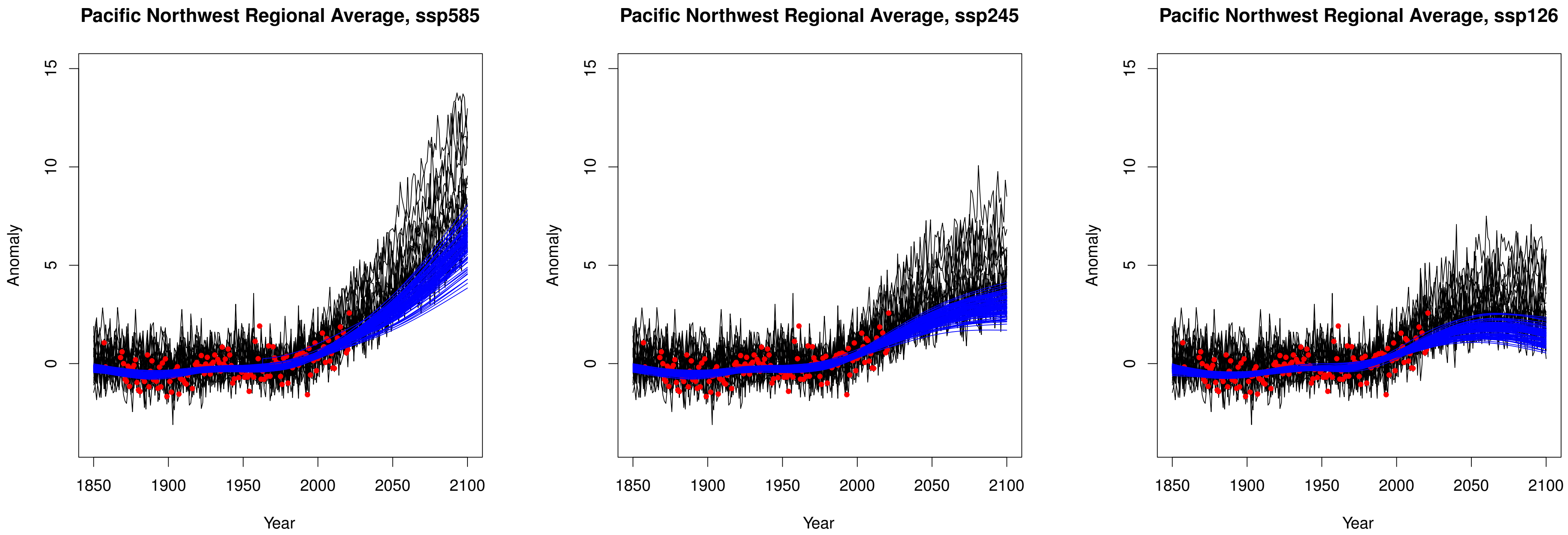
IIc: Projecting the Distribution of the Regional Variable Forwards and Backwards in Time

- 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)

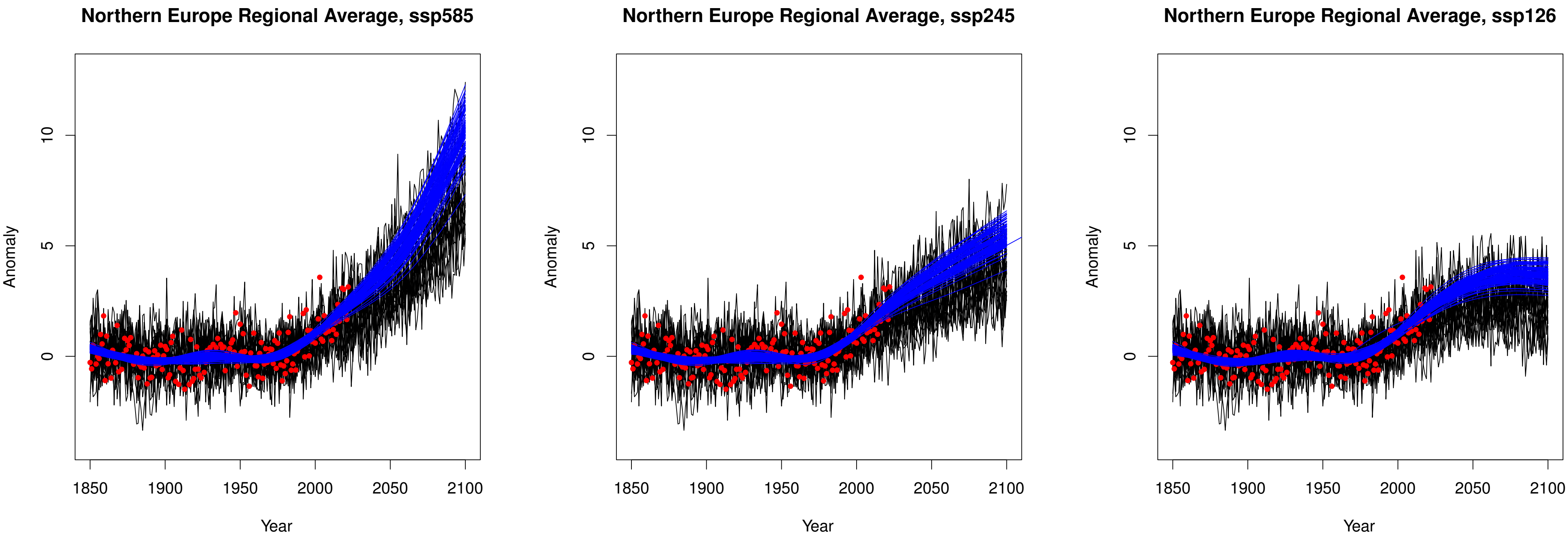
Regional Variable Projections: Gulf of Mexico



Regional Variable Projections: Pacific Northwest



Regional Variable Projections: Northern Europe

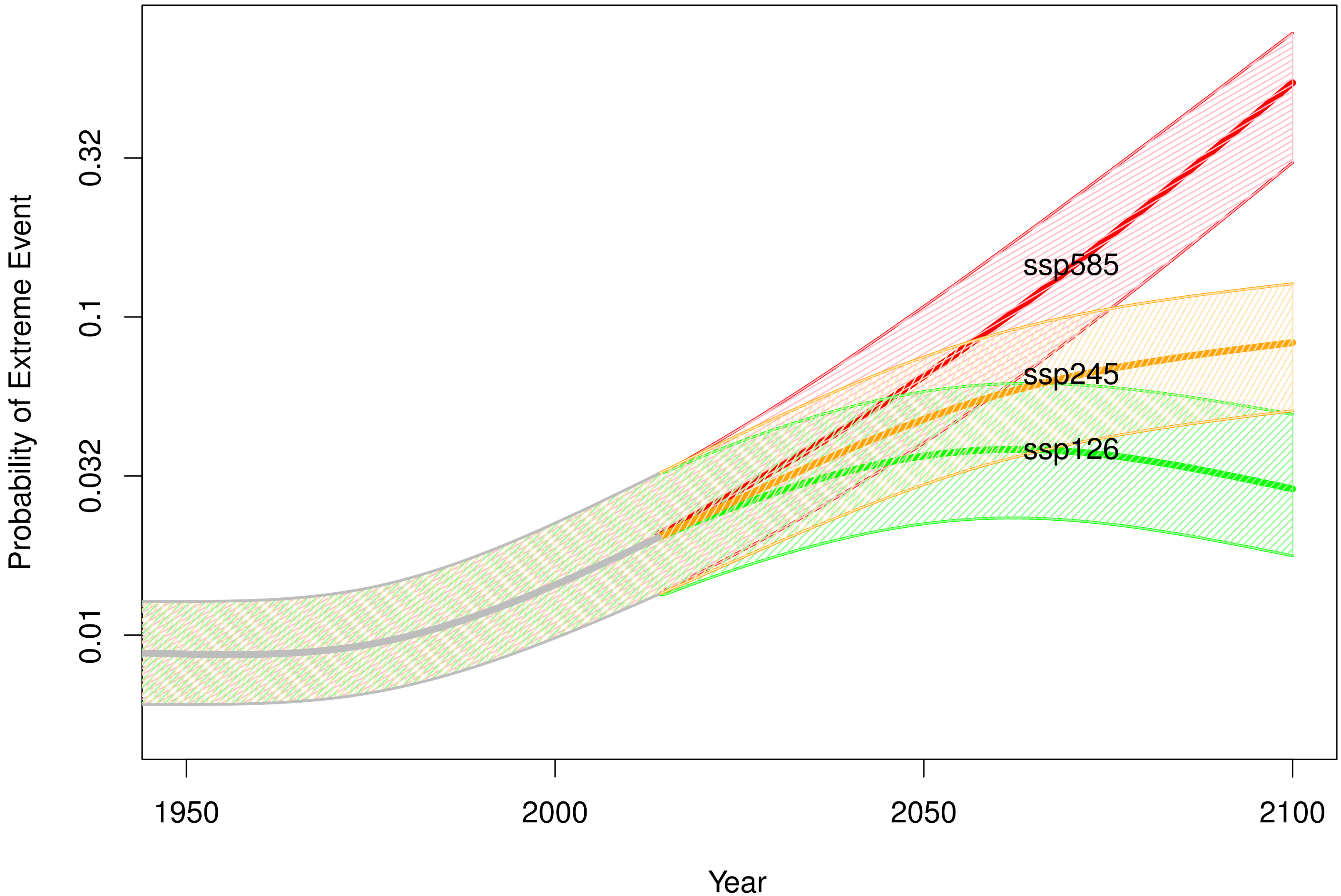


IId: End to End Analysis

- 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

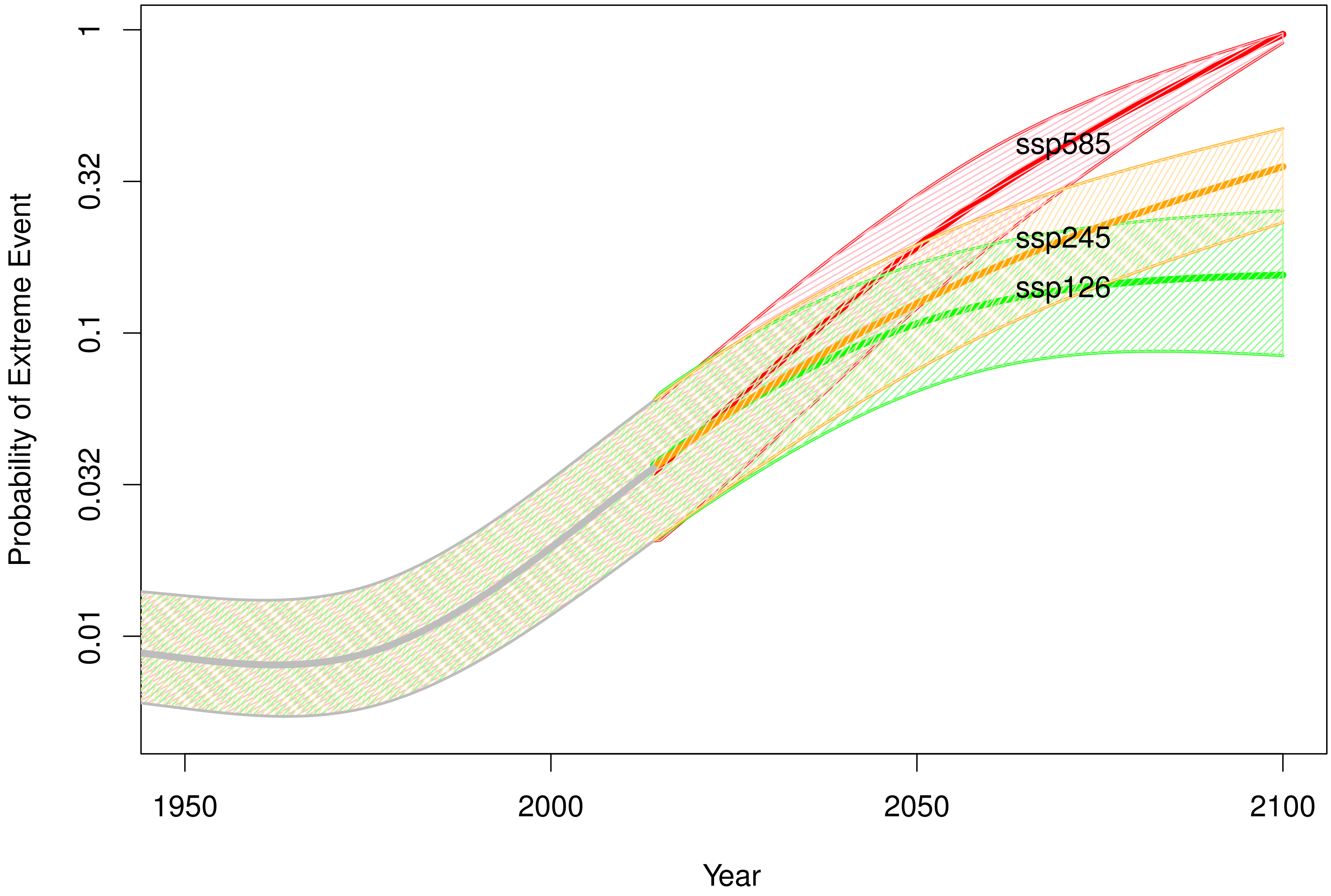
End To End Analysis: Mean Probability of Exceeding 2021 Value for All Stations in Canada

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)



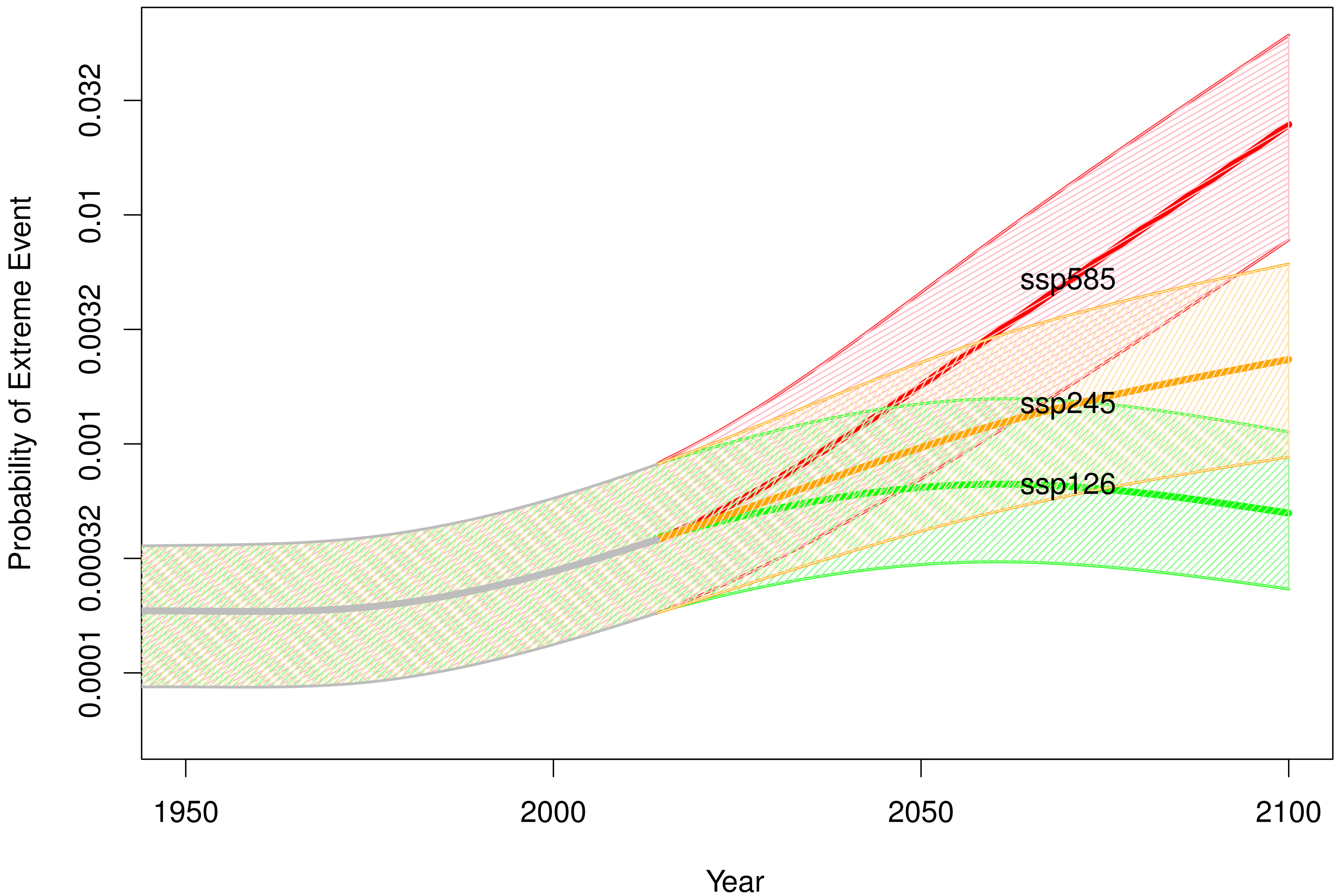
End To End Analysis: Mean Probability of Exceeding 2022 Value for All Stations in U.K.

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. The unique expertise of statisticians is *quantifying uncertainty*

- But there's still a caveat. George Box said, "All models are wrong, but some are useful". In the present context, one half of this quote is definitely correct. The other, I don't know

- Plenty of potential for future research by statisticians!