

# **Robust attribution and projection of extreme heat events to human influence on the climate**

Nicholas J. Leach

St. Cross College  
University of Oxford

*A thesis submitted for the degree of  
Doctor of Philosophy*

Trinity 2022

## **Abstract**

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



# **Robust attribution and projection of extreme heat events to human influence on the climate**



**Nicholas J. Leach**

St. Cross College  
University of Oxford

Supervised by

**Antje Weisheimer  
Myles R. Allen**

A thesis submitted for the degree of  
*Doctor of Philosophy*  
Trinity 2022



# Acknowledgements

**Personal**

**Institutional**



# Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



# Contents

<b>List of Figures</b>	<b>xi</b>
<b>List of Abbreviations</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The question of extreme event attribution . . . . .	2
1.2 Motivating the question . . . . .	2
1.3 Answering the question . . . . .	2
1.3.1 Probabilistic attribution . . . . .	2
1.3.2 Attribution through storylines . . . . .	2
1.3.3 Synthesising the two . . . . .	3
1.3.4 Operationalising attribution . . . . .	3
1.4 Conceptualising attribution . . . . .	3
1.5 What to expect in this thesis . . . . .	3
<b>2 Conventional probabilistic attribution</b>	<b>5</b>
2.1 Chapter open . . . . .	7
2.2 Abstract . . . . .	7
2.3 The 2018 heatwave in Europe . . . . .	7
2.3.1 Defining the event . . . . .	7
2.4 Model simulations & validation . . . . .	8
2.5 Methods . . . . .	9
2.5.1 Estimating the event threshold . . . . .	9
2.5.2 Counterfactual attribution . . . . .	10
2.5.3 Trend-based attribution . . . . .	10
2.5.4 Statistical methods . . . . .	10
2.6 Results . . . . .	10
2.7 Discussion . . . . .	14
2.8 Chapter close . . . . .	15

<b>3 Attribution and projection</b>	<b>17</b>
3.1 Chapter open . . . . .	19
3.2 Abstract . . . . .	19
3.3 Introduction . . . . .	19
3.4 Study design and methods . . . . .	21
3.4.1 Models . . . . .	21
3.4.2 ExSamples experiment design . . . . .	22
3.4.3 Statistical methods . . . . .	27
3.5 Results . . . . .	28
3.5.1 Comparison of HadAM4 and HadGEM3-GC3.05 baseline ensembles . . . . .	28
3.5.2 Projections of future extremes . . . . .	29
3.5.3 Sampling record-shattering subseasonal events . . . . .	37
3.6 Discussion . . . . .	39
3.7 Concluding remarks . . . . .	41
3.8 Chapter close . . . . .	41
<b>4 Partial forecast-based attribution</b>	<b>43</b>
4.1 Chapter open . . . . .	45
4.2 Abstract . . . . .	45
4.3 Introduction . . . . .	45
4.4 The 2019 February heatwave in Europe . . . . .	47
4.4.1 Forecasts of the heatwave . . . . .	49
4.5 Perturbed CO <sub>2</sub> forecasts . . . . .	52
4.6 Attributing the heatwave to diabatic CO <sub>2</sub> heating . . . . .	55
4.7 Discussion . . . . .	58
4.8 Chapter close . . . . .	60
<b>5 Forecast-based attribution</b>	<b>61</b>
5.1 Section . . . . .	62
<b>6 Discussion</b>	<b>63</b>
6.1 An overview of this thesis . . . . .	64
6.2 This thesis in the context of previous work . . . . .	64
6.3 Section . . . . .	64
6.4 Limitations . . . . .	64
6.5 Future research directions . . . . .	65
6.5.1 Assimilating the perturbations . . . . .	65
6.5.2 Alternative extremes . . . . .	65
6.5.3 Alternative forecast models . . . . .	65

6.5.4 Incorporating perturbation uncertainty . . . . .	65
6.5.5 Projections of future extremes . . . . .	65
6.5.6 Impact assessment . . . . .	66
6.6 Concluding remarks . . . . .	66

## Appendices

<b>A Resources</b>	<b>69</b>
<b>B Appendix 2</b>	<b>71</b>
<b>References</b>	<b>73</b>

*x*

# List of Figures

2.1	The 2018 heatwave in Europe: observed mean temperature anomalies over a range of timescales . . . . .	8
2.2	Calculating the heatwave threshold in HadGEM3-A from observations	9
2.3	Factual and counterfactual PDFs of the 2018 heatwave defined over three temporal scales . . . . .	11
2.4	Maps of the estimated change in probability of the 2018 heatwave due to anthropogenic influence on the climate . . . . .	12
2.5	Estimated changes in probability of the 2018 heatwave defined using regional mean temperatures . . . . .	13
2.6	Estimated changes in probability of the 2018 British Isles heatwave across a range of observations and model simulations . . . . .	14
2.7	Historical power spectrum of summer daily mean temperatures over the British Isles across a range of observations and model simulations . . . . .	15
3.1	UKCP PPE 2061-2080 deviations from forced response. . . . .	24
3.2	Synoptic characteristics of the study winters within the UKCP simulations. . . . .	26
3.3	Comparing statistics of DJF mean of daily maximum temperatures averaged over the UK region for the HOT1 winter. . . . .	31
3.4	Comparing statistics of DJF mean of daily maximum temperatures averaged over the UK region for the HOT2 winter. . . . .	34
3.5	Comparing statistics of DJF mean precipitation averaged over the UK region for the WET winter. . . . .	36
3.6	Examining statistics of subseasonal extreme weather events. . . . .	38
4.1	The 2019 February heatwave in Europe: synoptic characteristics & historical context. . . . .	48
4.2	Medium- to extended-range forecasts of the heatwave. . . . .	51
4.3	Global temperature and synoptic-scale dynamical response to CO <sub>2</sub> perturbations. . . . .	54
4.4	Attribution of the direct CO <sub>2</sub> influence on the heatwave. . . . .	57



# List of Abbreviations

<b>CDF</b>	.....	Cumulative distribution function
<b>CI</b>	.....	Confidence interval
<b>CO<sub>2</sub></b>	.....	Carbon dioxide
<b>DJF</b>	.....	December–January–February (meteorological winter)
<b>ECMWF</b>	.....	European Centre for Medium-range Weather Forecasts
<b>EEA</b>	.....	Extreme event attribution
<b>FAR</b>	.....	Fraction of attributable risk
<b>GEV</b>	.....	Generalised extreme value
<b>GMST</b>	.....	Global mean surface temperature
<b>GSAT</b>	.....	Global surface air temperature
<b>JJA</b>	.....	June–July–August (meteorological summer)
<b>PDF</b>	.....	Probability density function
<b>PPE</b>	.....	Perturbed parameter ensemble
<b>SIC</b>	.....	Sea ice concentration
<b>SST</b>	.....	Sea surface temperature
<b>UK</b>	.....	United Kingdom
<b>UKCP</b>	.....	UK Climate Projections
<b>UKMO</b>	.....	UK Met Office
<b>WWA</b>	.....	World Weather Attribution

*xiv*

*Quote*

— author

# 1

# Introduction

In this chapter I introduce the problem of attribution of individual extreme weather events to anthropogenic climate change. I review the current methodologies and frameworks that address this problem, in particular the contrasting storyline and probabilistic approaches to attribution. Although these frameworks are gaining acceptance and maturity, I suggest that a weather forecast-based approach could further increase the trustworthiness of attribution studies. Finally, I provide a conceptual sketch of these various attribution frameworks within a simple non-linear dynamical system.

**Author contributions:** This chapter is based on the the following publication \*

Surname, I1. I2., Surname, I1. I2. (year). **Title.** *Journal*, vol(issue), pages. DOI

## Contents

1.1	The question of extreme event attribution . . . . .	2
1.2	Motivating the question . . . . .	2
1.3	Answering the question . . . . .	2
1.3.1	Probabilistic attribution . . . . .	2
1.3.2	Attribution through storylines . . . . .	2
1.3.3	Synthesising the two . . . . .	3
1.3.4	Operationalising attribution . . . . .	3
1.4	Conceptualising attribution . . . . .	3
1.5	What to expect in this thesis . . . . .	3

---

\*with the author contributing as follows.

## 1.1 The question of extreme event attribution

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 1.2 Motivating the question

Now that I have posed the question, before I move on to how we might answer it, I think it would be useful for me to discuss why we want to answer it. In short: *what's the point of this thesis?*

In 2003, Myles Allen wrote “Liability for climate change” ([1](#)). This commentary is widely acknowledged as the first time the idea that individual extreme events could be attributed to external drivers such as human influence was proposed ([2](#)). Although Allen touched on both methodology and motivation for extreme event attribution, here I shall focus on the latter aspect. The motivation behind extreme event attribution as proposed in “Liability for climate change” is compensation for individual damages caused by climate change or, as Allen puts it,

Will it ever be possible to sue anyone for damaging the climate?

Tort law ([3–12](#)). Loss and damage ([13](#), [14](#)).

([15–17](#))

([18](#))

## 1.3 Answering the question

At this point, I have discussed the question that this thesis is primarily concerned with, and the reasons why it is of broad importance. Now, I shall explore the ways in which previous studies have approached this question, in particular focusing on the probabilistic (often “conventional”, [19](#)) and storyline ([20](#)) frameworks.

### 1.3.1 Probabilistic attribution

Probabilistic attribution seeks, ultimately, to determine the change in probability of an extreme event arising due to some external driver. This was the approach to extreme event attribution proposed in Allen ([1](#)) and first applied in Stott et al. ([19](#)).

### 1.3.2 Attribution through storylines

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no

information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### **1.3.3 Synthesising the two**

### **1.3.4 Operationalising attribution**

## **1.4 Conceptualising attribution**

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## **1.5 What to expect in this thesis**

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.



*Quote*

— author

# 2

## Conventional probabilistic attribution

Here I present a probabilistic extreme event attribution of the 2018 European heatwave. Whilst demonstrating the methodologies behind this framework, I examine how one particular aspect of probabilistic event attribution – the definition of the event – projects strongly onto the quantitative results. In the closing remarks, I reflect on potential issues with the approach taken within the chapter, and suggest ways in which these could be overcome.

**Author contributions:** This chapter is based on the the following publication \*

Leach, N. J., Li, S., Sparrow, S., van Oldenborgh, G. J., Lott, F. C., Weisheimer, A., & Allen, M. R. (2020). **Anthropogenic Influence on the 2018 Summer Warm Spell in Europe: The Impact of Different Spatio-Temporal Scales.** *Bulletin of the American Meteorological Society*, 101(1), S41-S46. <https://doi.org/10.1175/BAMS-D-19-0201.1>

### Contents

2.1	Chapter open	7
2.2	Abstract	7
2.3	The 2018 heatwave in Europe	7
2.3.1	Defining the event	7
2.4	Model simulations & validation	8
2.5	Methods	9
2.5.1	Estimating the event threshold	9
2.5.2	Counterfactual attribution	10

---

\*with the author contributing as follows. Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualisation and Writing – original draft.

2.5.3	Trend-based attribution . . . . .	10
2.5.4	Statistical methods . . . . .	10
2.6	Results . . . . .	10
2.7	Discussion . . . . .	14
2.8	Chapter close . . . . .	15

## 2.1 Chapter open

## 2.2 Abstract

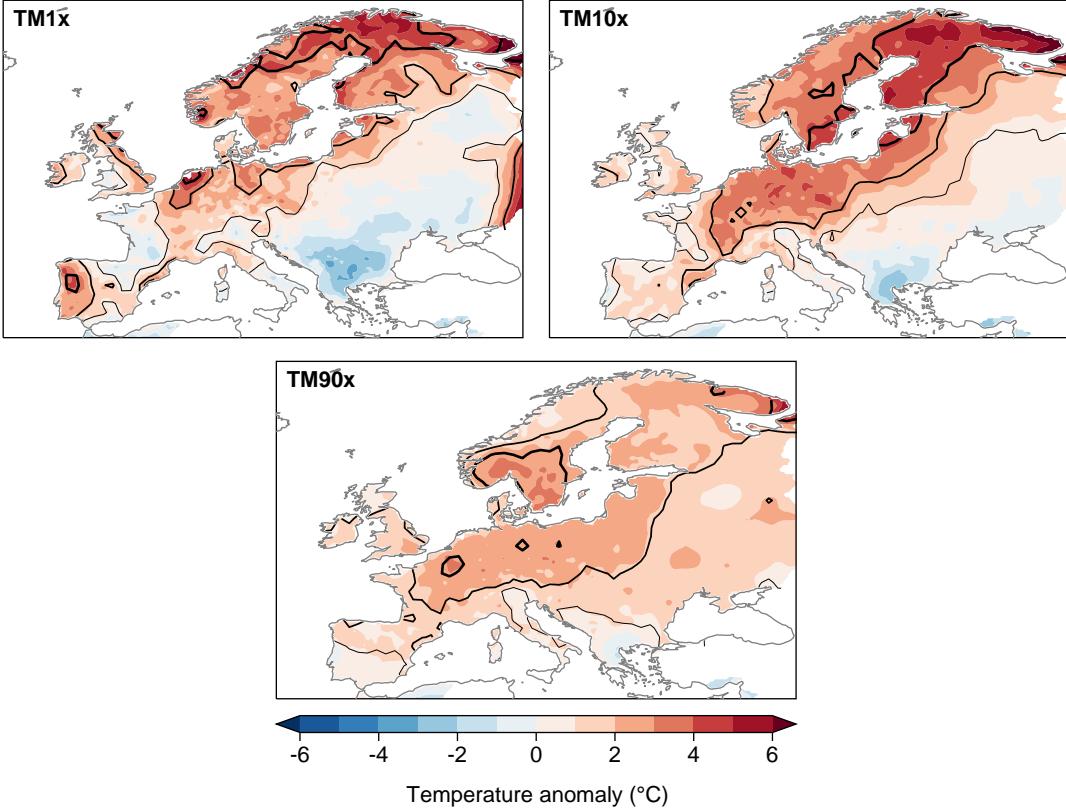
We demonstrate that, in attribution studies, events defined over longer time scales generally produce higher probability ratios due to lower interannual variability, reconciling seemingly inconsistent attribution results of Europe's 2018 summer heatwaves in reported studies.

## 2.3 The 2018 heatwave in Europe

The summer of 2018 was extremely warm in parts of Europe, particularly Scandinavia, the Iberian Peninsula, and central Europe, with a range of all-time temperature records set across the continent (21, 22). Impacts were felt across Europe, with wildfires burning in Sweden (23, 24), heatstroke deaths in Spain (25), and widespread drought (26). During the summer, the World Weather Attribution (WWA) initiative released an analysis of the heat spell (27) based on observations/forecasts and models in specific locations (Dublin, Ireland; De Bilt, Netherlands; Copenhagen, Denmark; Oslo, Norway; Linkoping, Sweden; Sodankyla, Finland; Jokionen, Finland), which concluded that the increase in likelihood due to human induced climate change was at least 2 to 5 times. In December, the U.K. Met Office (UKMO) stated that they found the 2018 U.K. summer temperatures were made 30 times more likely (28, 29). These two estimates appear to quantitatively disagree; however, we show they can be reconciled by considering the effect of using different spatial domains and temporal scales in the event definition. We also demonstrate that prescribed SST model simulations can underrepresent the variability of temperature extremes, especially near the coast, with implications for any derived attribution results.

### 2.3.1 Defining the event

We consider various temperature-based event definitions to demonstrate the impact of this choice in attribution assessments, and assess to what extent human influence affected the seasonal and peak magnitudes of the 2018 summer heat event on a range of spatial scales. The metric we use is the annual maximum of the 1-, 10-, and 90-day running mean of daily mean 2-m temperature (hereafter TM1x, TM10x, and TM90x respectively). We analyze two spatial scales: model grid box and regional. For regional event definitions, the spatial mean is calculated before the annual maximum. Regional domains are taken from Christensen and Christensen (30). Figure 2.1 shows the 2018 anomaly field for each of these metrics. In their study, the WWA used the annual maxima of 3-day mean daily maximum temperatures at specific grid points for its connection to local health effects (31), whereas the UKMO used the JJA mean temperature over the entire United Kingdom in order to answer the question of how anthropogenic forcings have affected the likelihood of U.K. summer seasons as warm as 2018. The same justifications can be used here, although we add that different heat event time scales are important to different groups of people, and as such using several temporal definitions may increase interest in heat event attribution studies. However, we recognize that other definitions than those used here may be more relevant to some impacts observed (such as defining



**Figure 2.1: The 2018 heatwave in Europe: observed mean temperature anomalies over a range of timescales.** Shading indicates mean temperature anomalies for the different temporal-scale heatwave metrics used. Black contours indicate z-scores of the 2018 heatwave for the three metrics based on detrended historical data from E-OBS. The contours indicate scores of 1-, 2-, and 3- $\sigma$ , in order of increasing thickness.

the event in the context of the atmospheric flow pattern and drought that accompanied the heat), and other lines of reasoning for selecting one particular event definition exist (32).

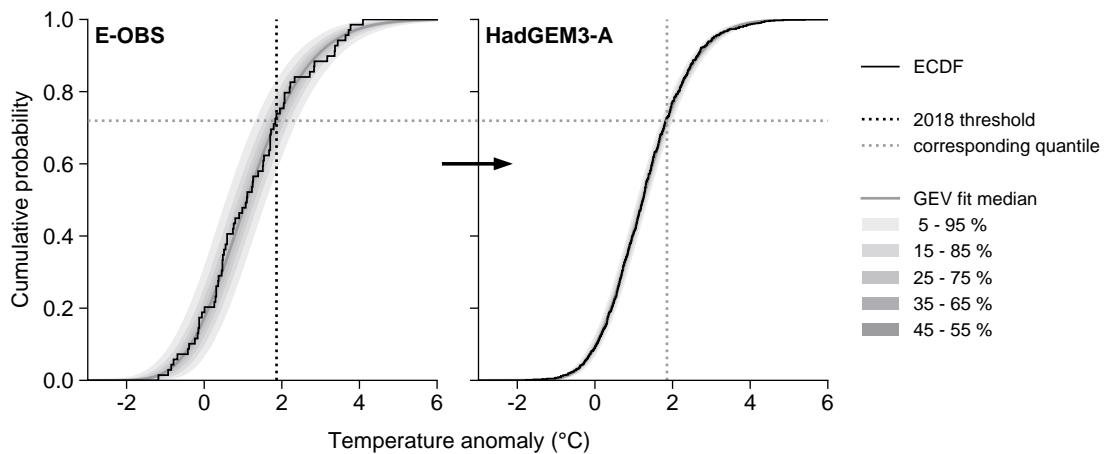
## 2.4 Model simulations & validation

We primarily use three sets of simulations from the UKMO Hadley Centre HadGEM3-A global atmospheric model (33, 34). These are a historical ensemble (1960–2013; Historical), and factual (ACT) and counterfactual (a “natural” world without anthropogenic forcings; NAT) ensembles of 2018. We compare results from this factual-counterfactual analysis with those from a trend-based analysis of Historical, ensembles from EURO-CORDEX (35–37) (1971–2018) and RACMO (38, 39) (1950–2018), and observations from E-OBS (1950–2018). A full model description is provided in the online supplemental information. Initially, we performed our analysis with the weather@home HadRM3P European-25 km setup (40) but found that this model overestimates the variability over all Europe for daily through seasonal-scale event statistics, and so it was omitted.

## 2.5 Methods

### 2.5.1 Estimating the event threshold

We first use historical data to estimate the return time of the 2018 event, and the corresponding temperature threshold in each model. We start by calculating the return period for the 2018 event in E-OBS. Since the distribution of temperature extremes changes as the climate changes, to account for the non-stationarity of the time series we remove the attributable trend by regressing onto the anthropogenic component of globally averaged mean surface temperature (the anthropogenic warming index, based on HadCRUT5; 41, 42). The regression coefficient or trend is shown in the supplemental material in Fig. ES1 (43). We then fit extreme value (EV) distribution parameters to this detrended E-OBS time series, and use these parameters to calculate the estimated return period of the 2018 event. We then find the temperature threshold in the model climatology that corresponds to this return period. We do this by fitting EV parameters to a detrended (by regressing onto the anthropogenic warming index, trends shown in Figs. ES2c7–9) climatological ensemble for each model. For HadGEM3-A, the climatology is Historical plus 15 randomly sampled members of ACT; for RACMO and EURO-CORDEX, the climatology is taken as the entire set of simulations described above. The calculation of model-specific climatological temperature thresholds from the E-OBS temperature threshold is illustrated in Figure 2.2. Using estimated event probabilities rather than observed magnitudes to define the event constitutes a quantile bias correction (44), accounting for model biases in the mean and variability of the temperatures simulated.



**Figure 2.2: Calculating the heatwave threshold in HadGEM3-A from observations.** The heatwave metric used here for illustration is TM1x. Solid black lines indicate empirical CDFs. Solid grey lines indicate GEV distributions fit to the data, with grey shading indicating confidence intervals of the fit. The dotted black line indicates the temperature observed during the 2018 heatwave in E-OBS. Dotted grey lines indicate the best-estimate quantile corresponding to the 2018 event in E-OBS based on the GEV fit, and the temperature of that quantile in the HadGEM3-A historical climatology.

### 2.5.2 Counterfactual attribution

For the main analysis, which we term “counterfactual” attribution (19), we estimate the probability ( $P$ ) of exceeding this climatological temperature threshold in the ACT and NAT ensembles. We do this by fitting EV parameters to each ensemble, and using them to calculate  $P_{\text{ACT}}$  and  $P_{\text{NAT}}$ . The estimated ACT and NAT distributions are shown for each metric in Figure 2.3. We express our results as the probability ratio,  $PR = P_{\text{ACT}}/P_{\text{NAT}}$ , representing the fractional change in probability of the 2018 event in the factual compared to the counterfactual world.

### 2.5.3 Trend-based attribution

We support the counterfactual attribution with a trend-based analysis (45) of E-OBS and all the model ensembles used. This trend-based analysis is based on the climatology alone, and does not require factual and counterfactual simulations. We start with the EV parameters fit to the detrended climatology, and then use the estimated climate change trend between 1900 and 2018 to shift the location of the EV distribution. This shifted distribution then represents the counterfactual distribution (analogous to NAT), and the original distribution represents the factual distribution (analogous to ACT), from which we can calculate probability ratios.

### 2.5.4 Statistical methods

We fit EV parameters using the method of L-Moments (46), modelling TM1x and TM10x using the generalized extreme value (GEV) distribution, and TM90x using the generalized logistic distribution. Uncertainties are calculated using a 10,000 resample nonparametric bootstrap throughout.

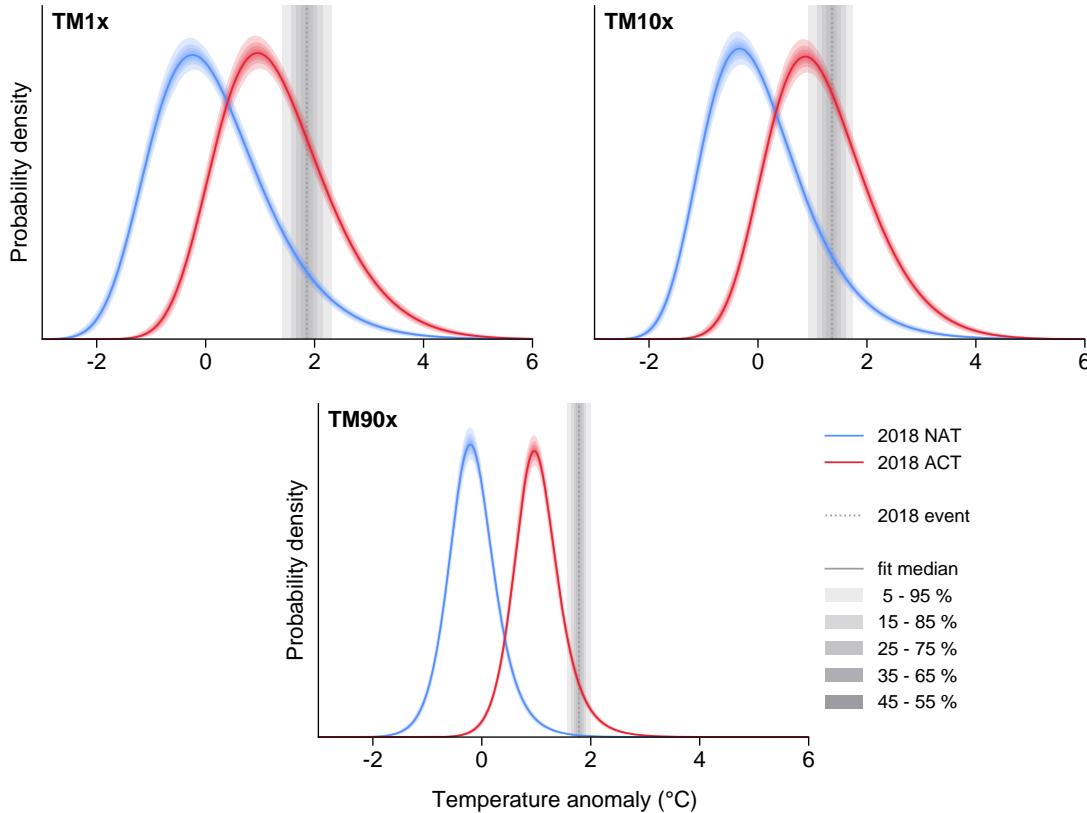
## 2.6 Results

Extreme daily heat events, measured by TM1x, are distributed heterogeneously throughout Europe. This is paralleled in the probability ratios seen in Figure 2.4, with large areas of the Iberian Peninsula, the Netherlands, and Scandinavia experiencing events that were highly unlikely in a climate without anthropogenic influence. A similar result is found on the regional scale in Figure 2.5 with Scandinavia and the Iberian Peninsula respectively experiencing 1-in-150 [25–25,000]<sup>t</sup> and 1-in-30 [9–550] year events in the current climate that were highly unlikely in the natural climate simulated in NAT. The remaining regions recorded maximum daily temperatures expected to be repeated within 4 years. The probability ratios for regional domains are typically larger than single gridboxes within them, though some regions contain clusters of extremely high probability ratios. This result is consistent with Uhe et al. (2016) and Angélil et al. (2018), who showed that increasing the spatial scale over which the event is defined tends to increase the probability ratio.

Extreme 10-day heat events, TM10x, were also widespread in Europe, with the most extreme occurring in Scandinavia (Fig. ES1j). Regionally, the PRs become more uniform (Figure 2.5), although Scandinavia and the Iberian Peninsula still have very high best-estimate PRs of 800 [20–infinite] and 85 [25–40,000] respectively.

---

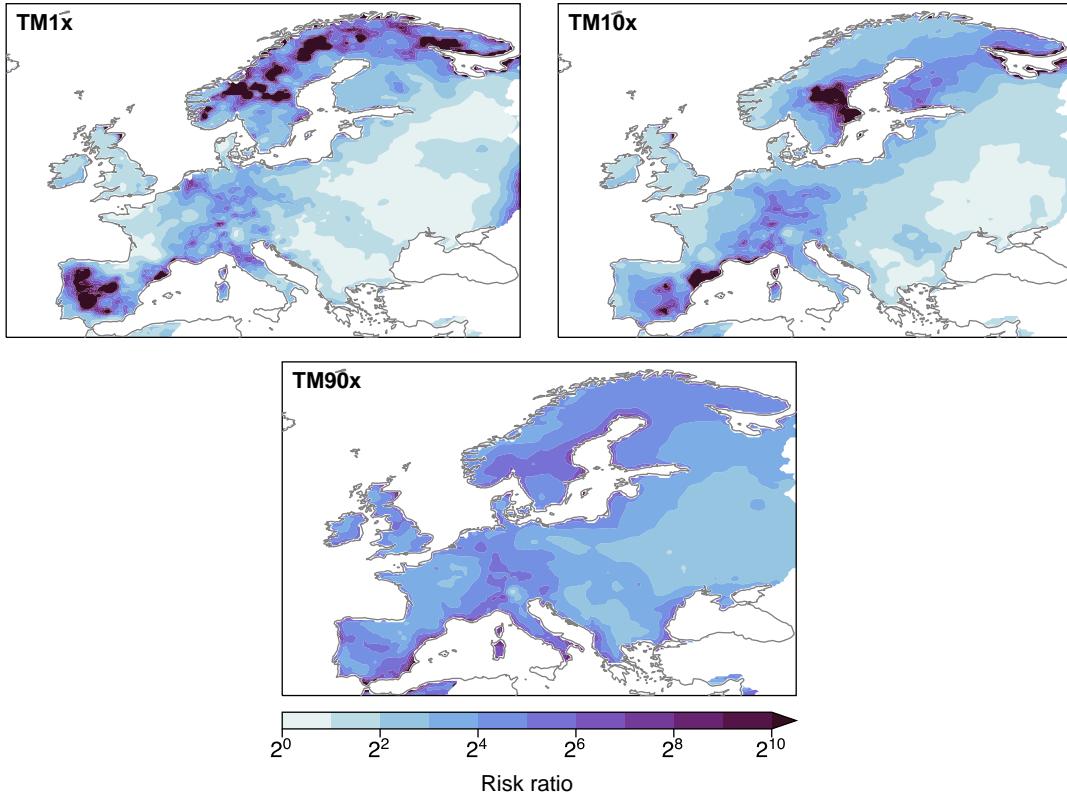
<sup>t</sup>numbers in brackets [] represent a 90% confidence interval.



**Figure 2.3: Factual and counterfactual PDFs of the 2018 heatwave defined over three temporal scales.** The heatwave metric used is given in the title of each panel. Solid red and blue lines indicate best-estimate GEV distributions fit to HadGEM3-A 2018 ACT and NAT ensembles respectively. Dotted grey line indicates 2018 event threshold defined using HadGEM3-A and E-OBS climatology (see Figure 2.2). Shading illustrates confidence intervals.

The PR map for season-long heat events measured by TM90x is more uniform throughout Europe (Figure 2.4). Scandinavia, the British Isles, France, and central and eastern Europe all experienced on the order of 1-in-10 year events (Fig. ES1I). The corresponding best-estimate PRs are between 10 and 100 for all regions (Fig. 1d), including those with lower return periods.

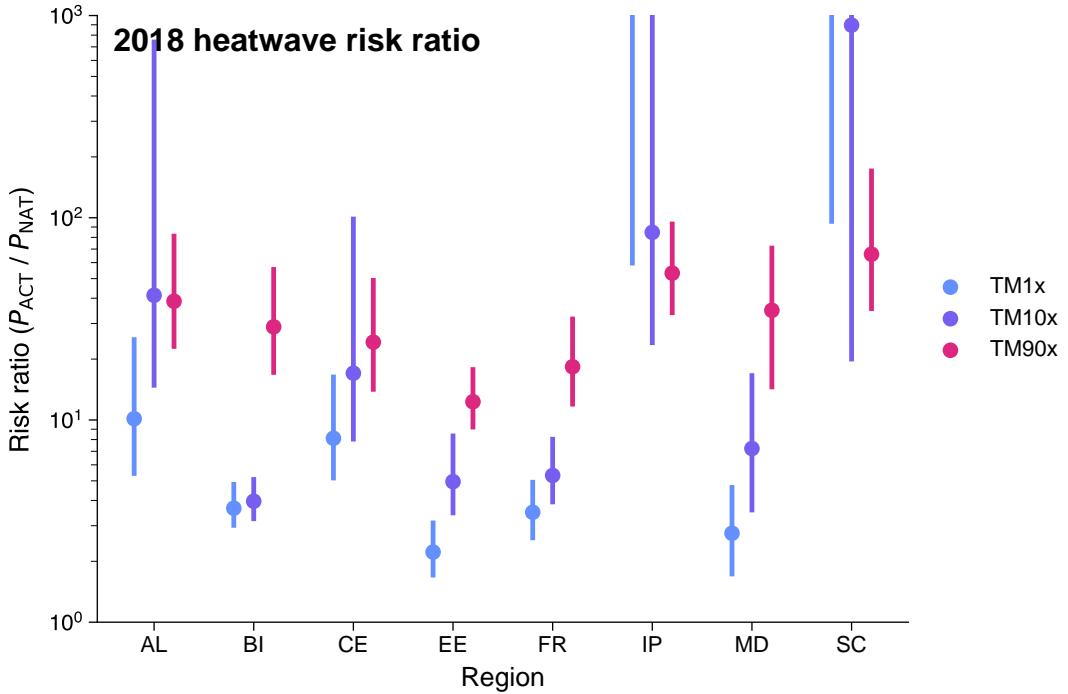
A trend-based analysis yields similar results, with PRs for the British Isles region shown in Figure 2.6, though we note that for HadGEM-3A this results in generally higher PRs than the corresponding counterfactual analysis, due to the attributable anthropogenic trend in the climatology being greater than the mean difference between the ACT and NAT ensembles. For the vast majority of the regions and metrics analysed here, the trend-based observational, trend-based model, and counterfactual model estimates of the return period are consistent with one another, and Figure 2.5 is a good representation of the synthesis of these different approaches and data sources. However, there are a few regions with notable discrepancies. The uncertainty in the E-OBS observed Scandinavia region TM1x trend are large enough that a 90 % confidence interval is not able to rule out a negative trend. Hence the corresponding probability ratio confidence interval includes values of less than 1 (i.e. that TM1x events at least as hot as the 2018 event have been



**Figure 2.4: Maps of the estimated change in probability of the 2018 heatwave due to anthropogenic influence on the climate.** The heatwave metric used is given in the title of each panel. Shading illustrates risk ratio of 2018 event at each gridpoint computed using HadGEM3-A ACT and NAT ensembles.

made less likely by climate change). This interval is large enough that it does still overlap with all the model-derived estimates, all of which suggest that the probability ratio is very likely greater than 70. This very large interval may arise due to natural variability affecting the relatively small sample size. Synthesizing these strands of information, we suggest that such daily extreme heat events over Scandinavia have been made more likely by climate change, but we are cautious about drawing very firm conclusions. The other clear discrepancy is for the TM90x metric British Isles results, shown in Figure 2.6. Despite good agreement between all other approaches and sources, the trend-based HadGEM3-A estimate is an order of magnitude higher and does not overlap with the others. This appears to be due to the variability of British Isles temperatures on this ~seasonal timescale being underestimated by this model, even though the estimated trend closely matches the other models and observations. We discuss this further below.

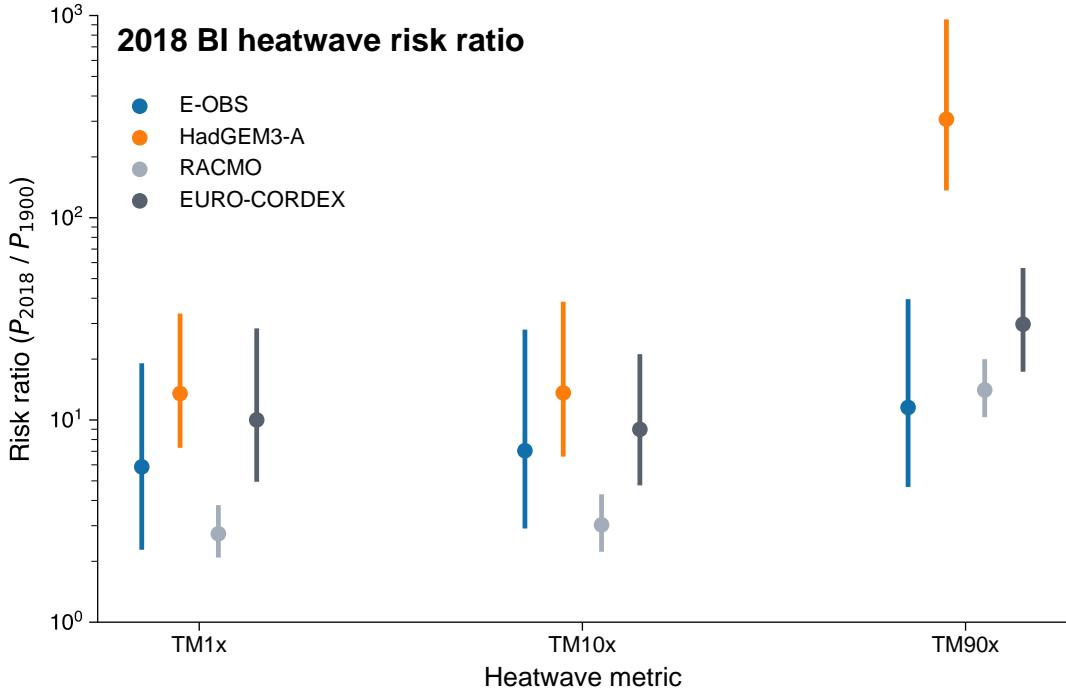
The PR increases with the event statistic time scale for the majority of grid points and regions, demonstrated in Figures 2.4 and 2.5. Figure 2.3 illustrates the cause using the British Isles region: as the time scale increases, the variance in the event metric decreases, while the mean shift between the factual and counterfactual distributions remains comparable. The similarity in attributed anthropogenic trend for the three time scales is also true in the observations and other models. The decrease in variance usually results in higher PRs, given a particular event return time, for the longer time scales.



**Figure 2.5: Estimated changes in probability of the 2018 heatwave defined using regional mean temperatures.** Color indicates heatwave metric. Dots indicate central risk-ratio estimate and bars indicate 90 % confidence interval.

There are exceptions due to the bounded upper tail of a GEV distribution with a negative shape parameter, resulting in the very high estimated PRs for TM1x in Scandinavia, the Iberian Peninsula, and the Netherlands (Figure 2.5). Now focussing on the British Isles region, Figure 2.3 also shows another reason why the TM90x metric probability ratios are much higher than the TM10x or TM1x results: in addition to the decreased variance in the TM90x metric, the 2018 event was more unusual when measured with this metric (a return period of 10.3 [5.7–20] years) compared to the two shorter timescale metrics (return periods of 2.5 [1.7–3.8] and 3.6 [2.4–6.0] for TM10x and TM1x respectively). These two factors (reduced variance and rarer event) result in best-estimate probability ratios of 3.7 [2.9–4.9] for TM1x and 29 [17–57] for TM90x. We therefore suggest that changes in the variance of the event metric as the time scales used changes largely reconciles the differences between the “2 to 5” and “30” times increases in likelihood found by the WWA and UKMO reports, with other methodological factors, such as the spatial scale used in the event definition, playing a more minor role as we have demonstrated for the British Isles.

As mentioned above, the trend-based HadGEM3-A analysis appears to overestimate the probability ratio of the 2018 event when considering the other approaches taken here (Figure 2.5). This is due to an important deficiency in the model: the model distributions are narrower than the observed distributions for this heatwave metric, meaning the model has lower variability in temperatures on seasonal timescales than the real world. This reduced variance has a significant impact on attribution results (47) and means that the PRs for the British Isles presented here, especially for TM90x, are likely to be overestimated. Underrepresented variability often occurs in prescribed-SST models (48,

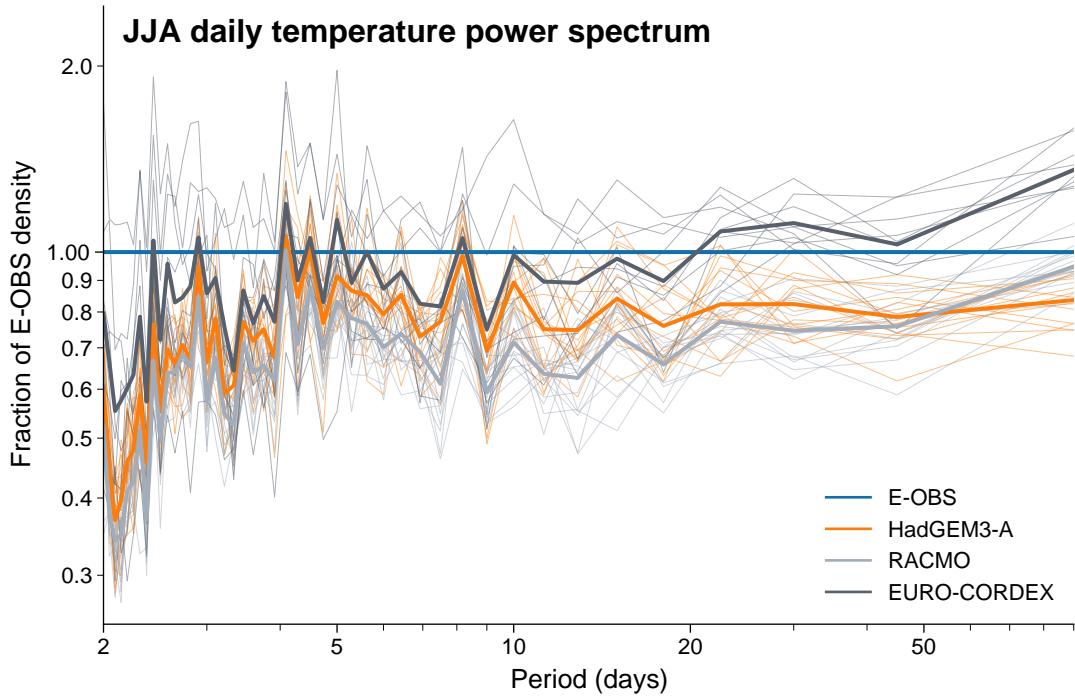


**Figure 2.6: Estimated changes in probability of the 2018 British Isles heatwave across a range of observations and model simulations.** Here, risk-ratios are estimated using a trend-based analysis. Color indicates data source. Dots indicate central risk-ratio estimate and bars indicate 90 % confidence interval.

49) and is present in HadGEM-3A for many coastal gridboxes in Europe. Figure 2.7 shows the power spectrum of JJA summer temperatures over the British Isles, indicating that HadGEM3-A has broadly similar spectral characteristics to E-OBS, but underrepresents the intraseasonal 2-m temperature variability at almost all frequencies, which will likely result in overestimated PRs. Power spectra for other model ensembles are shown for comparison, demonstrating that only the fully bias-corrected EURO-CORDEX ensemble has variability characteristics and magnitude that closely match the observations.

## 2.7 Discussion

Our analysis highlights a key property of extreme weather attribution: the variance of the event definition used, both in terms of the statistic itself and its representation within any models used. The use of longer temporal event scales in general increases both the spatial uniformity and magnitude of the probability ratios found, consistent with Kirchmeier-Young et al. (50), due to a decrease in variance compared to shorter scales. The difference in temporal scale between two reports concerning the 2018 summer heat is sufficient to explain the large discrepancy in attribution result between them. We find that several European regions experienced season-long heat events with a present-day return period greater than 10 years. The present-day likelihood of such events occurring is approximately 10 to 100 times greater than a “natural” climate without human influence.



**Figure 2.7: Historical power spectrum of summer daily mean temperatures over the British Isles across a range of observations and model simulations.** Power spectra are estimated as periodograms averaged over all historical years available for each data source, expressed as a fraction of the E-OBS power at each frequency. Color indicates data source. Thick lines show ensemble mean power for each source. Thin lines show individual ensemble members for each source.

The attribution results also show that the extreme daily temperatures experienced in parts of Scandinavia, the Netherlands, and the Iberian Peninsula would have been exceptionally unlikely without anthropogenic warming. The prescribed-SST model used primarily here tends to underestimate the variability of temperature extremes near the coast, which may lead to the attribution results overstating the increase in likelihood of such extremes due to anthropogenic climate change (47). We aim to properly quantify the impact of the underrepresented variability in further work. Although here we have used an unconditional temperature definition for consistency with the studies we aimed to reconcile, we plan to further investigate the effect of including both the atmospheric flow context and other impact-related variables such as precipitation in the event definition, and address issues models might have with realistically simulating the physical drivers of heatwaves.

## 2.8 Chapter close



*Quote*

— author

# 3

## Attribution and projection

In this chapter, I explore the close links between attribution of extreme weather events and their projection with climate change. I study a novel set of large-ensemble atmosphere-only model experiments to show that such large-ensembles are necessary to generate samples of the most extreme weather events, an understanding of which is crucial for climate change adaptation. In the closing discussion, I consider how forecast-based attribution could be leveraged to provide similar samples of specific future extreme weather events.

**Author contributions:** This chapter is based on the the following publication \*

Leach, N. J., Watson, P. A. G., Sparrow, S. N., Wallom, D. C. H., & Sexton, D. M. H. (2022). **Generating samples of extreme winters to support climate adaptation.** *Weather and Climate Extremes*, 36(), 100419. <https://doi.org/10.1016/j.wace.2022.100419>

### Contents

3.1	Chapter open . . . . .	19
3.2	Abstract . . . . .	19
3.3	Introduction . . . . .	19
3.4	Study design and methods . . . . .	21
3.4.1	Models . . . . .	21
3.4.2	ExSamples experiment design . . . . .	22
3.4.3	Statistical methods . . . . .	27
3.5	Results . . . . .	28
3.5.1	Comparison of HadAM4 and HadGEM3-GC3.05 baseline ensembles . . . . .	28

---

\*with the author contributing as follows. Data curation, Formal analysis, Investigation, Methodology, Visualization and Writing – original draft.

3.5.2	Projections of future extremes . . . . .	29
3.5.3	Sampling record-shattering subseasonal events . . . . .	37
3.6	Discussion . . . . .	39
3.7	Concluding remarks . . . . .	41
3.8	Chapter close . . . . .	41

## 3.1 Chapter open

## 3.2 Abstract

Recent extreme weather across the globe highlights the need to understand the potential for more extreme events in the present-day, and how such events may change with global warming. We present a methodology for more efficiently sampling extremes in future climate projections. As a proof-of-concept, we examine the UK's most recent set of national Climate Projections (UKCP18). UKCP18 includes a 15-member perturbed parameter ensemble (PPE) of coupled global simulations, providing a range of climate projections incorporating uncertainty in both internal variability and forced response. However, this ensemble is too small to adequately sample extremes with very high return periods, which are of interest to policy-makers and adaptation planners. To better understand the statistics of these events, we use distributed computing to run three ~1000-member initial-condition ensembles with the atmosphere-only HadAM4 model at 60km resolution on volunteers' computers, taking boundary conditions from three distinct future extreme winters within the UKCP18 ensemble. We find that the magnitude of each winter extreme is captured within our ensembles, and that two of the three ensembles are conditioned towards producing extremes by the boundary conditions. Our ensembles contain several extremes that would only be expected to be sampled by a UKCP18 PPE of over 500 members, which would be prohibitively expensive with current supercomputing resource. The most extreme winters we simulate exceed those within UKCP18 by 0.85K and 37% of the present-day average for UK winter means of daily maximum temperature and precipitation respectively. As such, our ensembles contain a rich set of multivariate, spatio-temporally and physically coherent samples of extreme winters with wide-ranging potential applications.

## 3.3 Introduction

Weather extremes are one of the most damaging hazards that society faces at the present-day (51). Many studies have now found that anthropogenic climate change is increasing the frequency and/or magnitude of certain types of extreme weather, including heatwaves, extreme rainfall and droughts (52). This has therefore resulted in a need to plan how society can adapt to the more frequent or severe weather extremes projected to occur under continued greenhouse gas emissions (43, 53, 54). In order to plan effectively, we must first understand and quantify how extreme weather events are projected to change into the future.

In the United Kingdom (UK), a key part of this understanding has been informed by the UK Climate Projections (UKCP) project. The most recent iteration of UKCP (UKCP18) was released in 2018 (55, 56) and included a number of novel climate model ensembles: a set of transient global simulations from coupled climate models, with 15 simulations from a single-model perturbed parameter ensemble (PPE) and 13 additional simulations from CMIP5 models; a set of 12 regional climate model simulations; and a set of 12 convection permitting model projections. In this study, we focus on the PPE of 15 global simulations, and our analysis and results build upon the information provided by these runs.

In particular, we are interested in how effectively the UKCP18 PPE has sampled extreme weather during the UK winter, and in exploring methods for improving the

sampling of extremes that could inform the design of future projections. To this end, we aim to provide proof-of-concept of a methodology for generating large ensembles of extreme winters. The key advantage is that our ensembles provide multivariate spatially and physically coherent scenarios of extreme weather with high return periods for use in impacts assessment.

We first select three exceptional UK winters from the UKCP18 PPE that occurred between 2061 and 2080 (henceforth the “study winters”). We then use the sea surface temperature (SST) and sea ice (SIC) fields from these winters to force very large perturbed initial-condition ensembles using the HadAM4 model, which has been implemented to run in the distributed computing system climateprediction.net at the same horizontal resolution as the UKCP18 global simulations. This allows very large ensembles to be produced and is possible because HadAM4 requires less computational resources. These ensembles are intended to provide numerous extreme samples, hence are called the “ExSamples” ensembles.

This provision of many samples of extremes is similar to the UNSEEN method for quantifying weather extremes (57, 58). UNSEEN uses seasonal hindcast ensembles to estimate the likelihood of “unprecedented” extreme events with considerably more confidence than possible from the observational record in isolation. The key similarity between UNSEEN and the approach taken here is that both are methods that aim to drill into the uncertainty surrounding the most extreme events by providing very large ensembles of such extremes using a dynamical model. However, there are key differences: UNSEEN uses coupled simulations that are conditioned solely on the predictable component of the weather at the time the model was initialised by observations, while in ExSamples, the model is atmosphere-only and conditioned both on perturbed initial conditions and lower boundary forcing from a climate projection. Another difference lies in the distributed computing system used here, which enables 1000+ member ensembles of a single winter to be produced; compared to the O(100) members produced by operational seasonal forecasting centers.

We compare the statistics of weather extremes in these ExSamples ensembles to both the corresponding extreme study winter, and to the whole UKCP18 PPE 2061-2080 climate distribution in order to answer several science questions:

- Is the atmosphere-only model able to produce equal magnitude extremes to those within the study winters from the UKCP18 PPE? If the study winter lies outside the atmosphere-only model distribution, this suggests the importance of coupling to a dynamic ocean and other differences between the models for producing extremes.
- Were the study winters truly exceptional, or could they have been even more extreme?
- To what extent did the SSTs and SIC during the study winters condition the extreme response?
- Is carrying out this type of experiment using a computationally cheaper, but less modern, atmosphere-only model a better methodology for sampling extremes than increasing the size of the UKCP18 PPE?

In this paper, we first describe the models used, experiment design and statistical methodologies performed within the study. We then present the results of our experiments, first comparing the climate distributions of the two models over a present-day baseline

period to assess whether there are any significant biases between them. Taking any biases into account, we compare the projections from our three future ensembles to the UKCP18 PPE, focussing on how the extreme tail of the climate distribution is sampled. This comparison allows us to explore the sampling advantage given by, and influence of, the SST and SIC. The very large ensembles created also allow us to examine the influence of the large scale dynamics present during the study winters using a circulation analog approach. We then use a single ensemble member case study to highlight the importance of large ensembles for sampling unprecedented extreme events that cannot always be statistically extrapolated from smaller ensembles (59, 60). Finally, we discuss the insights provided by these experiments, and how they might inform the design of future projections; also suggesting directions for future research that could further improve our approach.

## 3.4 Study design and methods

### 3.4.1 Models

#### HadGEM3-GC3.05 global climate model

In addition to the novel ExSamples ensembles, we also analyse UKCP18 global PPE simulations of the RCP8.5 emission scenario (61). This PPE is based on the global HadGEM3-GC3.05 coupled ocean atmosphere model (56, 62). This combines an 85 vertical level atmosphere model at 5/6° zonal and 5/9° meridional resolution (N216, ~60 km at mid-latitudes) with a 75 level ocean model at ORCA025 (1/4°) horizontal resolution. The aim of this PPE is to explore a range of plausible model responses to climate change. The parameters were selected on the basis of the credibility of the model response on both weather and climate timescales (63–66). In this study we use both the final product 15-member PPE and a 10-member subsample. The 10-member subsample consists of the 12 members that compose the accompanying UKCP18 regional climate model projections (56), minus two members that displayed a significant weakening of the Atlantic Meridional Overturning Circulation (67). Henceforth, we shall refer to the HadGEM3-GC3.05 simulations analysed here as the “UKCP18 PPE”. Unless stated otherwise, this refers to the 15-member PPE.

#### HadAM4 N216 atmospheric model

The novel simulations presented here are performed by the global HadAM4 atmosphere and land surface model (68, 69). Like its predecessor, HadAM3 (70), it includes prognostic cloud, convection and gravity-wave drag parameterisation schemes, a radiation scheme that treats water vapour and ice crystals separately, and a land surface scheme. The updates in HadAM4 include a mixed-phase precipitation scheme, parameterisation of ice cloud particle size and the radiative effects of non-spherical ice particles, and a revised boundary layer scheme. The version used here incorporates an upgrade to the spatial resolution (71, 72), which matches the horizontal resolution of the HadGEM3-GC3.05 simulations analysed here. HadAM4 has 38 vertical levels; and here the sea surface temperature (SST) and sea ice fraction (SIC) boundary conditions are taken from specific years and members of the HadGEM3-GC3.05 UKCP18 PPE simulations.

A key aspect of the HadAM4 simulations described here are that they are performed on the personal computers of volunteers using the climateprediction.net distributed computing

system (73–75). This system has been used previously to run a range of Hadley Center Unified Model variants (76), including a coupled atmosphere-slab ocean model (77), a fully coupled model (78) and an atmosphere-only model (79) similar to HadAM4. The near thousand member ensembles presented here would be prohibitively expensive to run using a standard supercomputer, and so we are only able to run the bespoke experiments presented in this study because of this distributed computing system, and the volunteers involved. However, the constraints of this system strongly motivate the choice of HadAM4: it is sufficiently memory-efficient that it can be run on personal computers at the same horizontal resolution as the state-of-the-art HadGEM3-GC3.05 model.

Henceforth, we shall refer to the HadAM4 simulations presented here as the “ExSamples” ensembles. A complete description of the ExSamples ensembles, including the selection of the prescribed SST/SIC, is given below in “Experiment design”.

### 3.4.2 ExSamples experiment design

ExSamples covers six distinct sets of simulations: three future winter and three baseline period ensembles. The process behind generating each future and corresponding baseline ensemble is as follows:

1. Select a single winter from within the UKCP18 PPE over the 2061-2080 period. This winter is chosen on the basis of being particularly “extreme”; more detail on how we selected the three future winters is given below in “Selecting the three ‘extreme’ study winters”. The 2061-80 period is used as we wanted to test this proof-of-concept with a large underlying climate change signal; and this is the period for which there is additional UKCP18 data available: 12km regional and 2.2km convection-permitting model projections (56, 80).
2. Use the SSTs and SICs from this winter to force HadAM4 over the November - March period (the November of each simulation is used to spin-up the simulation and is discarded prior to analysis). An ensemble is created from the boundary conditions for this single winter through initial-condition perturbations. Due to the nature of the (ongoing) distributed computing system used to run the model (73, 77), our target final ensemble size is 1500 members conditioned on the SST/SIC from a single winter, and in this study we analyse all the members that are complete at the time of writing and pass our quality control checks, which ranges from 883 to 1036 over the three ensembles (81).
3. Create a corresponding HadAM4 baseline ensemble by using winter SSTs and SICs from the same UKCP18 member as the selected winter over the period 2007-2016. For each of the ten years, an ensemble of 50 members is generated using initial-condition perturbations. This results in a target baseline ensemble size of 500 members per future winter ensemble, conditioned on SST/SICs from 10 present-day winters. Although the difference in size between the future and baseline ExSamples ensembles is not relevant in this study, it may be for specific user applications.

### Motivation of the experiment design

In this section we outline the motivation behind our experimental design, with a particular focus on the differences between the internal variability sampled by a coupled model, and

sampled by an atmosphere-only model. The coupled PPE in UKCP18 samples a series of events including the most extreme ones, that arise from the response to anthropogenic forcing plus coupled internal variability. The latter is due to a combination of internal variability in the ocean, the impact this has on the atmosphere, and internal variability generated within the atmosphere itself (82). So an extreme deviation about the long-term forced trend in a coupled simulation might have occurred solely due to atmospheric internal variability but it is reasonable to expect that it is more likely than other years to have had a contribution from ocean internal variability. Therefore, by picking three winters with the largest deviations from the long-term climate trend, we hope to capture more winters where the ocean has strongly influenced the extreme. In years where there is an appreciable influence from ocean internal variability, which will be manifest in the simulated SST and SIC patterns along with the long term forced response of the ocean to anthropogenic forcing, then there is more potential for there to be an additional effect from atmospheric internal variability to produce greater extremes. Therefore an initial-condition ensemble of atmosphere-only simulations forced by SSTs, SIC and anthropogenic forcing from a study winter, where members differ only by atmospheric internal variability, can be used to distinguish winters where the ocean internal variability has played an important role from ones where the ocean has played little role. In the former case, we would expect to sample extremes beyond the UKCP18 extreme more often than we would by chance from atmospheric internal variability around the long term forced response.

## Definitions of key terms

There are several technical definitions we use throughout this study, which we will define in this section.

Firstly, a “raw value” is the simulated value straight from the model, as found within the relevant data product.

“Anomalies” are these raw values set relative to the average absolute value over some reference period, in order to remove any mean model biases. For the ExSamples simulations, we define anomalies as the raw values minus the average over the corresponding 2007-2016 baseline ensemble members. For the UKCP simulations, we define anomalies as the raw values minus the 1997-2026 reference period mean for each PPE member. This longer 30-year period is used to reduce the impact of inter-decadal variability that may be present in the time series of each member. For precipitation, we show results in terms of the “percent change” to compensate for differences in average rainfall intensity between the two models used. Percent changes are calculated as anomalies divided by the average raw value over the reference period (times 100 %).

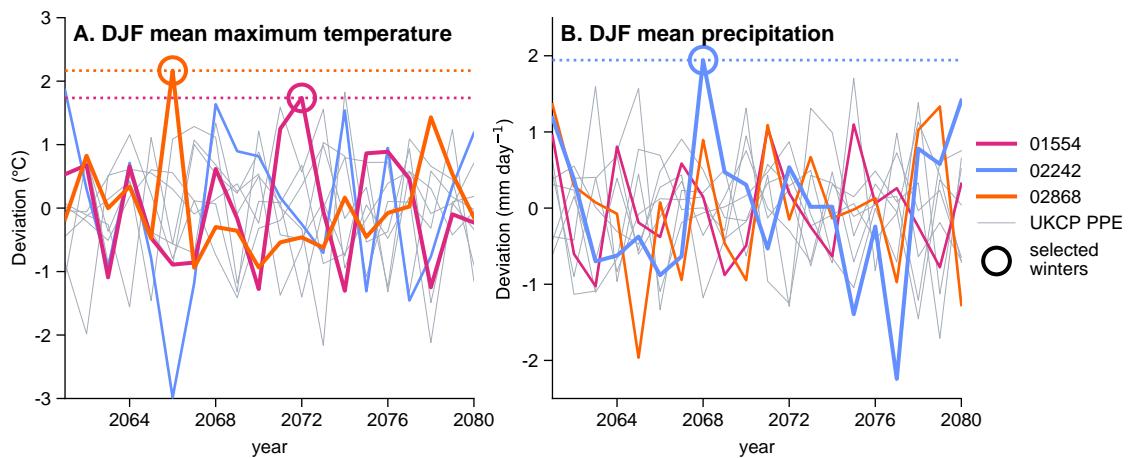
Finally, we use “deviations” in the context of the UKCP PPE to refer to the raw values relative to a long-term trend. Deviations are calculated as the residual of a simple linear regression computed over time for each PPE member (ie. over the 2061-2080 period). Deviations therefore represent a basic estimate of the variability about a long-term forced trend. Hence we use deviations to measure how unusual a particular simulated winter within the UKCP18 PPE is compared to others when a forced trend that may vary across ensemble members is present; and also to generate time series that can be fitted using statistical models that assume the underlying process is stationary (though we note that non-stationary statistical models could also be used). Deviations of the UKCP18 PPE also provide the closest simple comparison to the atmosphere-only ExSamples ensembles which only sample atmospheric internal variability.

## Selecting the three “extreme” study winters

To generate our future ExSamples ensembles, we needed to select three “extreme” winters from the UKCP18 PPE projections. We considered winters from the 10-member subsample over the period 2061-2080, giving a total of 200 candidate winters for selection. The 10-member subsample was used such that the ExSamples ensembles generated here would be able to be directly compared to the UKCP18 regional climate and convection permitting model projections if desired in the future.

The variables we used to compare how “extreme” each candidate winter was were the winter (DJF) mean of daily maximum temperatures, and winter mean precipitation, each averaged over the UK land region. Since the UKCP18 PPE displays significant forced trends in climate over the 2061-2080 period and based on the thinking behind the experimental design, we used the deviations of each candidate winter as the basis for our selection; if we used anomalies we would naturally bias our selection towards the end of the period.

Motivated by the recent winter extremes of the record hottest UK winter day of February 26th 2019 and the record wet winter month of February 2020, we aimed to select two “hot” winters and one “wet” winter. However, the method could be applied to the winters with the coldest or driest deviations. As shown in Figure 1, there is one clear candidate for each type of extreme: UKCP18 PPE member 02868 (ID numbers as [65](#)) year 2066 as a hot winter; and member 02242 year 2068 for the wet winter. The next most extreme hot winters shown in Figure 1A all had similar deviations, so we distinguished between them on the basis of their anomalies, choosing member 01554 year 2072, which has the highest anomaly of any of the candidate winters.



**Figure 3.1: UKCP PPE 2061-2080 deviations from forced response.** **A,** DJF mean of daily maximum temperatures averaged over the UK region. Coloured lines indicate the three UKCP runs from which the study winters were chosen. The study winters are circled and dotted horizontal lines indicate the deviation of each study winter. The ensemble member id of the three runs is given in the legend. **B,** as **A**, but for DJF mean precipitation.

Table 1 provides a summary of the study winters selected. For clarity, we refer to the ExSamples ensembles by the abbreviations given in the final column of Table 1 followed by “ensemble” (so the ensemble that uses the SST/SIC from UKCP18 member 02868 year 2066 is “HOT1 ensemble”, and the corresponding baseline ensemble is “HOT1-B ensemble”). We use “aggregate baseline ensemble” to denote the aggregate of all three baseline ensembles. We refer to the corresponding winters as the ensemble abbreviation followed by “winter”. Finally, we refer to the UKCP18 PPE ensembles as “UKCP” followed by the period the samples are taken from.

	Boundary condition (study winter) detail			Abbreviation
	UKCP18 member	Year	Extreme type	
<b>Future projections</b>	02868	2066	HOT	HOT1
	01554	2072	HOT	HOT2
	02242	2068	WET	WET
<b>Baseline ensembles</b>	02868	2007-2016	-	HOT1-B
	01554	2007-2016	-	HOT2-B
	02242	2007-2016	-	WET-B

**Table 3.1:** Summary of experiments performed for ExSamples project.

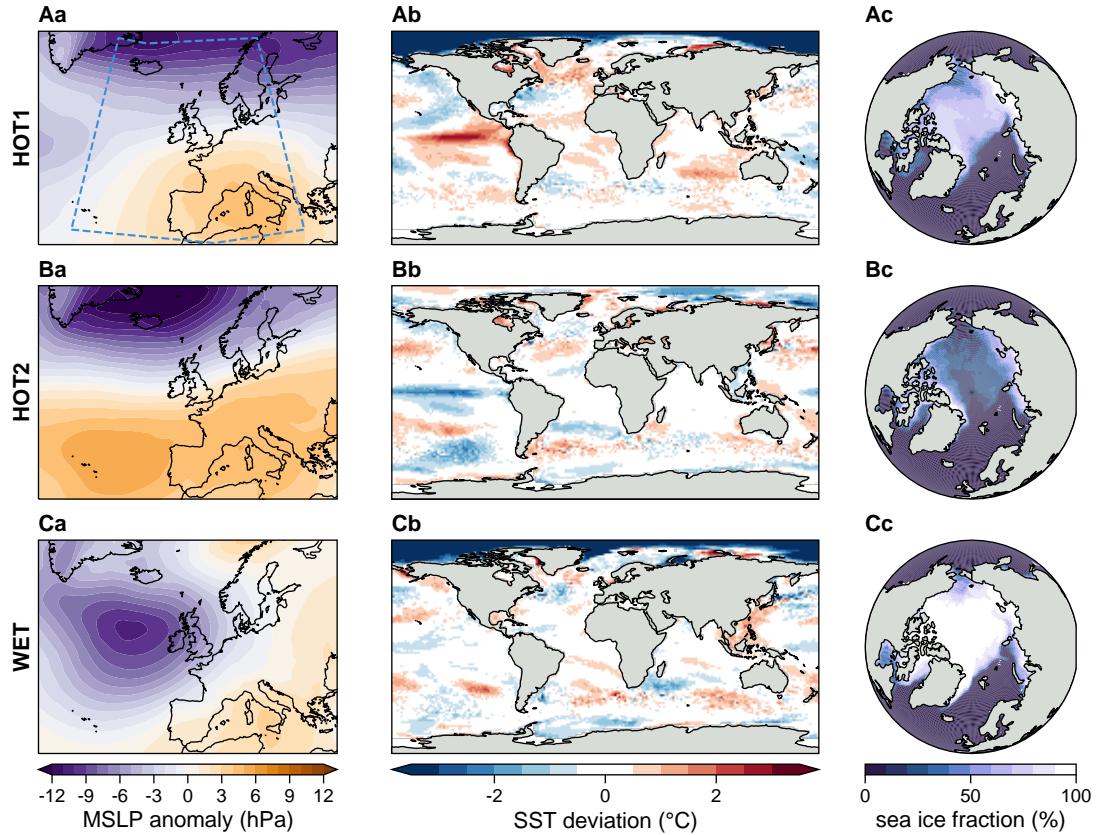
### Synoptic characterisation of the study winters

Here, we briefly describe the broad synoptic characteristics of each of the three future winters selected. Figure 2 shows three key characteristics: mean sea level pressure (MSLP) anomalies over the UK; SST deviations; and Arctic SICs. They display a wide range of meteorological and climatological features: none of the extreme winters selected are caused by very similar large-scale features.

The HOT1 winter displays a strong positive NAO pattern. Over the UK the flow is even more zonal, and has a weaker gradient; the positive NAO pattern is also weaker. In terms of the 30 weather patterns derived by (83), this winter shares similarities with several weather patterns, including those they numbered 20 and 23. These two patterns have been shown to be conducive to producing record temperatures on daily timescales (84). During this winter, the El Nino Southern Oscillation (ENSO) pattern of global SST variability was in a strong positive phase , alongside moderately positive Atlantic Multidecadal Variability and negative phase Pacific Decadal Oscillation (85). This extreme winter shows some loss of Arctic sea ice compared to the present day, though it is still mostly intact - the mean Arctic sea ice fraction is approximately 70 %.

The HOT2 winter displays a similar MSLP pattern to the first hot winter. The mean large scale flow over the whole winter is closest to weather pattern 20 of Neal et al. (83): a strong positive NAO with associated pressure high off the west coast of Spain. This weather pattern is associated with warm and wet weather over the UK (84, 86–88). There is a weak La Nina (negative) ENSO phase ; which has previously been linked to an increased likelihood of positive NAO (89–92). No other modes of SST variability are present. With regards to SIC, this particular PPE member has virtually lost all winter Arctic sea ice by 2072. It has been suggested that Arctic sea ice loss may be linked with more persistent mid-latitude weather patterns (93, 94), though this is still a subject of active scientific interest (95–97).

The WET winter displays a strong cyclonic south westerly flow with a low west of Ireland; classified as weather pattern 29. This pattern is associated with generally warm and wet weather. ENSO is in a neutral phase during this winter; and there are no other modes of SST variability in significantly positive or negative phases. Of the three study winters, this one has the smallest change in sea ice relative to the present-day; Arctic sea ice is almost entirely intact over the winter.



**Figure 3.2: Synoptic characteristics of the study winters within the UKCP simulations.** The row titles indicate the study winter. **a**, DJF mean MSLP anomalies for each winter. **b**, DJF mean SST deviations for each winter. Deviations are calculated for each gridpoint timeseries over 2061-2080. **c**, DJF mean Arctic sea ice fraction for each winter. The blue dashed line in **Aa** indicates the area used for analog subsampling described in 2.3.2.

### 3.4.3 Statistical methods

#### Estimating distributions of extremes

We estimate distributions using the method of L-moments (46, 98, 99). We use L-moments for their computational efficiency and stability. Uncertainties in the fit distributions, their CDFs and corresponding return periods are calculated using a 10,000 resample nonparametric bootstrap. The specific distributions used for each variable analysed are described in the following paragraphs.

For mean DJF daily maximum temperatures (TXm) and mean DJF precipitation rate (PRm), we use a generalised pareto distribution (100, 101) fit to the upper quartile of the sample population. When estimating CDFs and corresponding return periods from the fit, if the value in question lies below the upper quantile, we use the empirical CDF.

For maximum DJF daily maximum temperatures (TXx), we use a generalised extreme value distribution fit to the sample population.

For maximum DJF daily mean precipitation rate (PRx), we use a generalised logistic distribution (46) fit to the sample population. A generalised logistic distribution is used since the tail of the UKCP18 PPE 2061-2080 deviations population is clearly heavier than estimated by best-fit generalised extreme or generalised pareto distributions; we note that this approach to modelling block maxima of daily rainfall has some precedent in the literature (102, 103). This issue is not a feature of the L-Moments estimator used: a maximum likelihood estimator yields near-identical results. It is possible that the apparent discrepancy with the generalised extreme value distribution arises from the number of independent precipitation events per season not being near enough to the asymptotic limit (independent event count  $\rightarrow \infty$ ) for classical extreme value theory to be appropriate, as noted previously for annual daily rainfall maxima (104), though further work is needed to determine this conclusively.

#### Analog construction

In order to assess the dynamical contributions to the extreme weather simulated during the study winters, we use an MSLP analog approach (105–107). For each future ExSamples ensemble (and each corresponding baseline ensemble), we create a subsample of analogs composed of ensemble members that have a root mean square error (Euclidean distance) of less than 3 hPa from the UKCP18 PPE study winter average MSLP over the domain enclosed by the dashed blue lines in Figure 2Aa (-30:20° E; 35:70° N). This domain was the best for explaining variance in UK temperatures and close to best for UK precipitation of those investigated by (83). We used a 3hPa threshold as this was the tightest constraint that resulted in analog ensembles large enough to infer statistics from with any degree of certainty ( $>20$  members in each case). The MSLP distance based subsampling results in an ensemble of analogs in which the mean large scale flow during the winter very closely matches the study winter. We can then use these ensembles of analogs to estimate the dynamical contribution and associated uncertainty to the extreme weather.

## 3.5 Results

### 3.5.1 Comparison of HadAM4 and HadGEM3-GC3.05 baseline ensembles

Before we can robustly compare the projections within the UKCP18 PPE and ExSamples ensembles, we must first quantify any differences between the representations of UK climate within the HadAM4 and HadGEM3-GC3.05 models. We do this by comparing the 15-member UKCP18 PPE over 2007-2016 ( $15 * 10y = 150$  samples total) with each of the three 2007-2016 ExSamples baseline ensembles ( $\sim 50 * 10y = 500$  samples each) in turn, and their aggregate ensemble. Here we quantify whether the simulated climates differ using a two-sample Kolmogorov-Smirnov (K-S) test (108–111) at the 5 % significance level on the anomalies of the variable in question unless stated otherwise. We use anomalies here since our main results are presented using anomalies to account for any model mean biases (and biases between different UKCP18 PPE members), but note if there are significant differences between the two model climate means. Verifying the accuracy of these models against reality lies outside of the scope of this paper, but has already been studied for both the UKCP18 PPE (56) and HadAM4 (71, 72).

For both mean and maximum DJF daily maximum temperatures over the UK (TXm and TXx respectively), the UKCP 2007-2016 and ExSamples baseline distributions are highly comparable (Figures 3, 4, S4, S7, S8, S9). None of the three (nor their aggregate) ExSamples baseline ensemble distributions are statistically significantly different from the corresponding UKCP baseline ensemble distributions for either TXm or TXx anomalies. The ExSamples aggregate baseline ensemble mean biases are +0.06 K and +0.18 K compared to the UKCP18 PPE for TXm and TXx respectively. We note that this lack of a statistically significant difference does not imply that the two model ensembles are drawn from identical underlying distributions.

For mean DJF precipitation rate over the UK (PRm), we do find clear differences in the behaviour of the models. The ExSamples baseline ensembles have a reduced winter average rainfall intensity compared to the UKCP18 PPE: a 16 % ( $0.61 \text{ mm day}^{-1}$ ) lower ensemble mean. They also have a slightly increased spread in winter rainfall. We note that these differences in simulated UK climate do not appear to be the result of differences in the large-scale dynamics of the two models over the Euro-Atlantic sector; as investigated using a Principal Component (PC) Analysis in the Supplementary Information. Summarising this analysis: we find three DJF mean MSLP PCs dominate the variance of UK rainfall explained by the PCs in the UKCP18 PPE; the distributions of these PCs is near-identical in the ExSamples baseline ensembles and the UKCP18 PPE. Despite the difference in spread, none of three ExSamples baseline ensemble distributions are statistically significantly different from the UKCP18 baseline ensemble distribution for absolute PRm anomalies; nor is their aggregate. However, due to this discrepancy in mean rainfall intensity between the two models, we measure projected PRm in percent changes rather than anomalies, both in the figures presented and analysis carried out. After converting to percent changes, the differences in the spread of the distributions becomes relatively larger (Figure 5) and the distributions of percentage anomalies are statistically significantly different. This does not appear to arise from the specific sets of lower boundary conditions used in ExSamples: there are no statistically significant differences between any of the three ExSamples baseline ensembles.

Despite the differences in PRm, the two models show little difference in their simulated

distributions of the DJF maximum of daily mean precipitation averaged over the UK (PRx). The difference in mean PRx between all the ExSamples baseline ensembles and the UKCP18 PPE is only 4 % ( $0.99 \text{ mm day}^{-1}$ ). None of the three (nor their aggregate) ExSamples baseline ensembles are statistically significantly different from the UKCP 2007-2016 distribution for PRx anomalies.

### 3.5.2 Projections of future extremes

In this section we examine the future ExSamples ensembles and compare them to the UKCP18 PPE projections. Since we are largely concerned with winters that are extreme as a whole, rather than isolated extreme weather events within the winters (consistent with our methodology for selecting the three study winters), we analyse “hot” winters through DJF-mean temperatures and “wet” winters through DJF-mean precipitation.

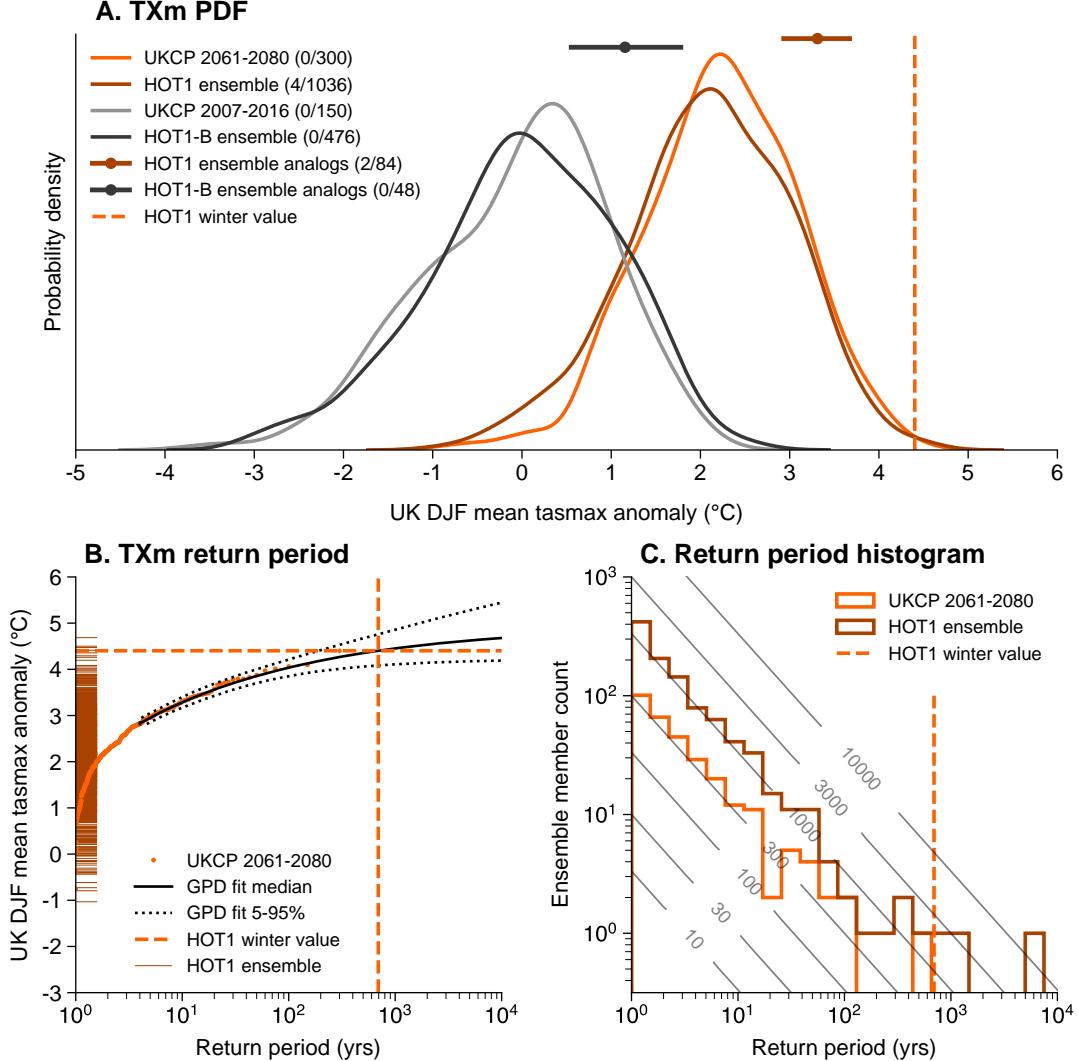
#### HOT1

We first address the primary question: was the atmosphere-only HadAM4 model able to capture the magnitude of the extreme simulated in the study winter by the coupled HadGEM3-GC3.05 model? Yes - there are four within the HOT1 ensemble that exceed the TXm value of the study winter, as shown in Figure 3.

However, the prescribed SST/SIC within the HOT1 simulations do not appear to have conditioned this ensemble towards producing more extremes than would be expected from an (unconditioned by construction) UKCP18 PPE of the same (increased) size. This is clearly seen in Figure 3: the distributions of the HOT1 and UKCP 2061-2080 ensembles are very similar in the PDF subplot; and the ExSamples return period sample histogram follows the “1000 member” expectation line closely. We can conclude that despite the HOT1 winter being an exceptional extreme within the context of the UKCP18 PPE, the associated SST and SICs did not pre-condition the winter towards (nor away from) such an extreme.

In order to compare the conditioning (effectively the “sampling advantage”) across the three ensembles, we examine the relative exceedance risk of three different extreme thresholds set by the following UKCP18 PPE distribution quantiles: 0.9, 0.95 and 0.99; representing 1-in-10, -20 and -100 year extremes. We do this for both the TXm and PRm variables. We first calculate the threshold values that correspond to the given extremes using the UKCP 2061-2080 deviations statistical fit (ie. the black line in Figure 3B). We then calculate the fractions of the UKCP 2061-2080 and ExSamples ensembles that lie above these thresholds. We present the results in Table 2 in terms of the relative risk of the given extreme in the ExSamples ensemble compared to the UKCP ensemble. This is calculated as the fraction of the ExSamples ensemble that exceeds the threshold divided by the corresponding fraction of the UKCP ensemble, analogous to the “risk ratios” often used in extreme event attribution studies (19, 112). This relative risk provides a measure of how many more samples of extremes of a particular return period we would expect to see in the ExSamples ensembles compared to a UKCP18 PPE-style ensemble of equal size. The quantitative results in Table 2 support the picture provided by Figure 3: the HOT1 ensemble was not conditioned towards producing any more extremes than expected from the unconditioned UKCP 2061-2080 ensemble (for several thresholds it actually appears to have been marginally conditioned away from producing extremes).

While the boundary conditions did not have any impact on the likelihood of an extreme winter, the large-scale dynamical situation of the study winter did. According to the analogs within the HOT1 ensemble, this specific dynamical situation increased the chance of a 1-in-100 year winter (based on the UKCP 2061-2080 statistical fit in Figure 3B) by a factor of 6.2 [5.3 , 6.9]. A similar level of dynamical conditioning is seen in the baseline ensemble. The analog-based subsampling also suggests that the prescribed SST/SIC may actually make the dynamical situation of the study winter less likely to occur than expected from the baseline climatological rate: the proportion of analogs in the HOT1 ensemble is 20 % lower than in the HOT1-B ensemble. Note that this change in analog frequency is not significant at the 5 % level. This change is reflected in the HOT1 ensemble mean MSLP anomalies, which are negative southwest of the UK and positive northwest of the UK (the opposite pattern to the study winter).



**Figure 3.3: Comparing statistics of DJF mean of daily maximum temperatures (TXm) averaged over the UK region for the HOT1 winter.** **A**, PDFs of baseline and future ensembles. The light orange PDF shows UKCP 2061-2080 deviations, with the distribution mean set to the ensemble mean anomaly between 2007-2016 and 2061-2071. The dark orange PDF shows HOT1 ensemble anomalies. The light grey PDF shows UKCP 2007-2016 anomalies. The black PDF shows HOT1-B ensemble anomalies. The dashed vertical light orange line indicates the HOT1 winter deviation. The dark orange and black dotted bars indicate the mean and likely range (16-84 %) of corresponding analog subsamples. The bracketed values in the legend indicate the number of ensemble members that exceed the HOT1 winter threshold over the total number of ensemble members. **B**, return period diagram. The light orange dots show the empirical CDF of UKCP PPE 2061-2080 deviations. The solid black line shows the median generalised pareto distribution fit. The dotted black lines indicate a 5-95 % credible interval of the distribution fit. The dark orange dashes along left y axis indicate positions of HOT1 ensemble anomalies. **C**, histograms of sampled return periods. The light orange line indicates the UKCP 2061-2080 deviations histogram, and the dark orange line the HOT1 ensemble anomalies. The dashed light orange line indicates the best-estimate return period of the HOT1 winter deviation. Grey contours indicate the expected histogram curve arising from a sample of size given by the contour labels.

DRAFT Printed on May 17, 2022

We note that the sampled return periods are calculated using the best-estimate fit distribution shown in the return period diagram; hence the curves in **C** and **A** are related by the transfer function indicated by the solid black line in **B**.

Study winter	Variable	UKCP18 quantile (return period)		
		0.9 (1-in-10 year)	0.95 (1-in-20)	0.99 (1-in-100)
<b>HOT1</b>	TXm	0.9 [0.86 , 0.96]	0.84 [0.77 , 0.97]	0.97 [0.75 , 2.32]
	PRm	1.02 [0.95 , 1.08]	0.98 [0.85 , 1.03]	2.03 [1.0 , 3.78]
<b>HOT2</b>	TXm	4.25 [3.95 , 4.64]	5.71 [4.97 , 6.05]	9.97 [7.34 , 24.8]
	PRm	2.93 [2.5 , 3.22]	3.6 [3.17 , 3.81]	10.08 [4.5 , 16.19]
<b>WET</b>	TXm	3.75 [3.61 , 4.06]	4.3 [3.67 , 4.7]	5.02 [3.53 , 10.14]
	PRm	3.96 [3.42 , 4.22]	4.7 [4.22 , 4.94]	11.75 [6.17 , 17.14]

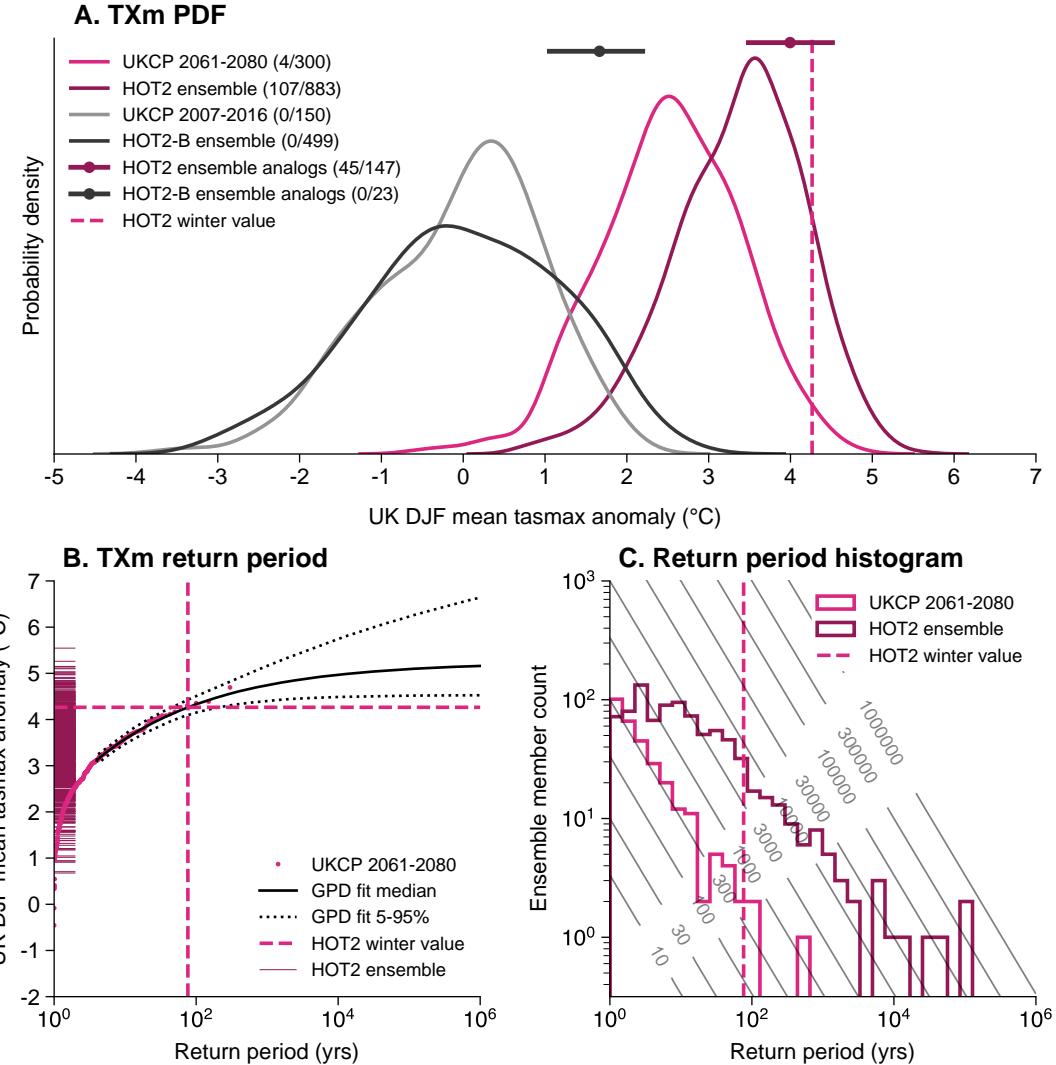
**Table 3.2:** Ratio of exceedance likelihood of three extreme thresholds between the ExSamples future ensembles and the UKCP18 PPE 2061-2080 deviations. Square brackets indicate a 90 % CI.

## HOT2

Again, the magnitude of the extreme in the study winter was captured within the HOT2 ensemble.

The HOT2 ensemble produced more extremes than would be expected from a UKCP18 PPE ensemble of the same size (Figure 4A, C, Table 2), suggesting that it was conditioned towards such extremes by the prescribed SST/SIC. We can see from Figure 4C that the HOT2 ensemble samples extremes that we would only expect to see within an unconditional UKCP18 PPE-type ensemble of total sample size 10,000 (for the period 2061-2080, this would be 500 members \* 20 years = 10,000 samples). Table 2 supports the picture that the HOT2 ensemble was significantly primed towards producing extremes: the relative risk of a 1-in-100 year event was 10 times greater in the HOT2 ensemble than the UKCP18 PPE for both hot (TXm) and wet (PRm) extremes.

In addition to the SST conditioning, the dynamical situation of the study winter also made an extreme season more likely, as shown by the horizontal lines representing the likely range of the analog subsamples in Figure 4A. Based on the number of analogs sampled, the frequency of this particular large-scale flow was increased by a factor of 3.6 [2.6 , 5.4] relative to the climatological frequency estimated using the ExSamples baseline ensemble, which may be due to the prescribed boundary conditions. This would fit within the canonical picture that the negative La Nina ENSO phase is associated with positive NAO ([89](#), [113](#)).



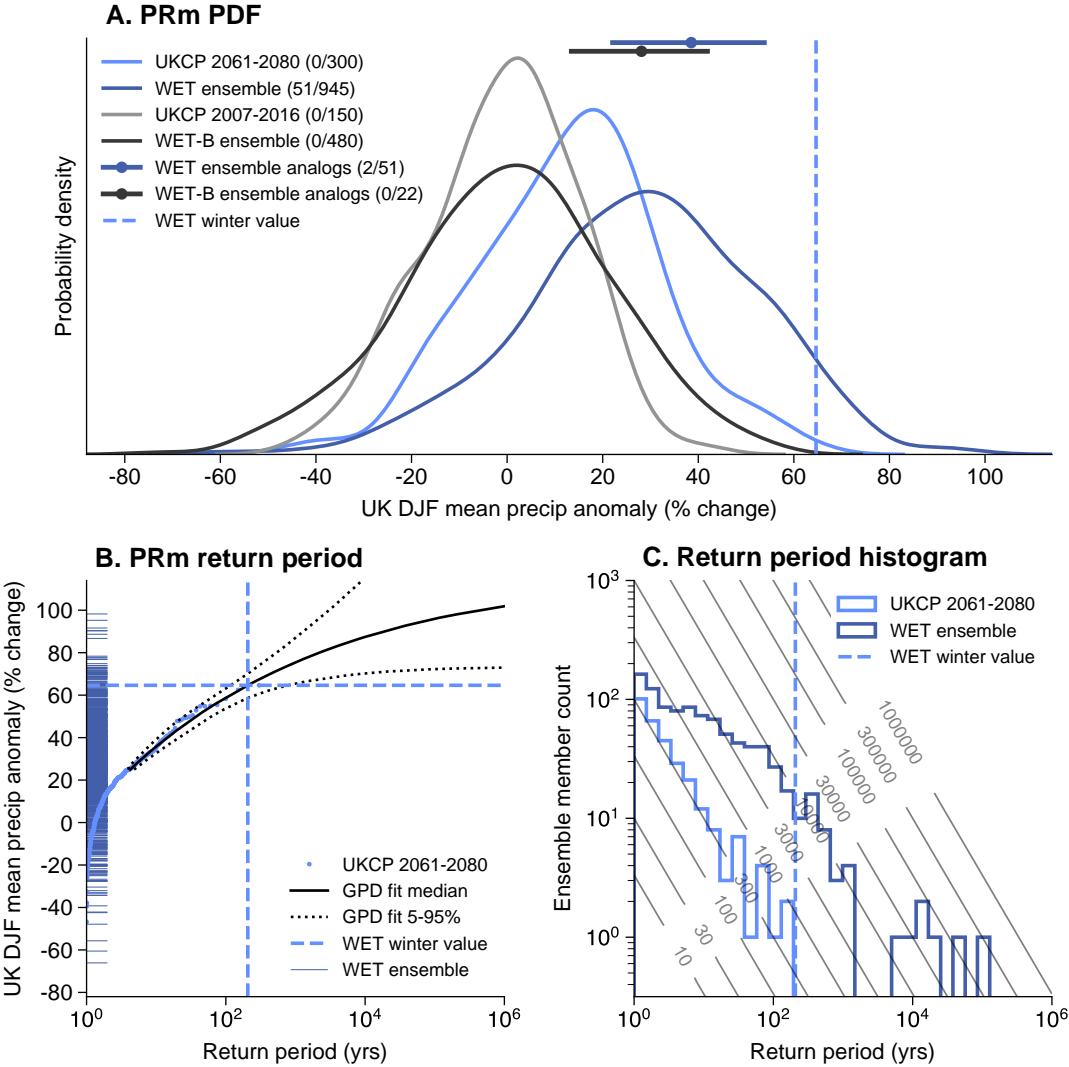
**Figure 3.4: Comparing statistics of DJF mean of daily maximum temperatures (TXm) averaged over the UK region for the HOT1 winter. As Figure 3, but for the HOT2 winter.**

## WET

Finally, we examine the WET winter extreme. As in both hot winters, the magnitude of the extreme within the study winter lies within the range of the WET ensemble.

As in the HOT2 ensemble, the prescribed SST/SIC have conditioned the WET ensemble towards producing more wet extremes than would be expected from an unconditioned ensemble, as shown by the histogram of sampled return periods and shifted PDF compared to the UKCP 2061-2080 PDF in Figure 5. This is consistent with the quantitative estimates in Table 2, which suggest that the WET ensemble was 5 times more likely to produce a 1-in-20 year wet (PRm) extreme, and 12 times more likely to produce a 1-in-100 year extreme.

An analog-based dynamical analysis shows that, once again, the large-scale circulation pattern present in the study winter was important for the development of the extreme rainfall that was simulated, consistent with previous weather pattern studies (Richardson et al., 2018, 2020). Interestingly, conditioning on the study winter dynamics appears to have a smaller influence on the WET ensemble than on the corresponding baseline: the difference between the distributions implied by the PDF and by the dotted bar is much greater for the baseline simulations (black) than for the future simulations (dark blue) in Figure 5A. This may be due to the SST/SIC conditioning in the future ensemble.

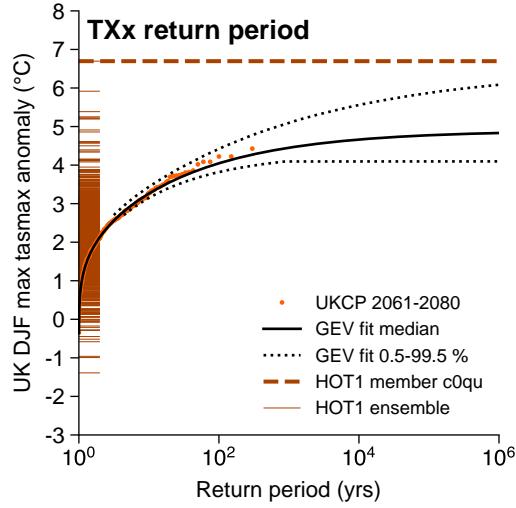


**Figure 3.5: Comparing statistics of DJF mean precipitation (PRm) averaged over the UK region for the WET winter.** As Figure 3, but of DJF mean precipitation averaged over the UK region for the WET winter.

### 3.5.3 Sampling record-shattering subseasonal events

Although this study is largely concerned with extremes that occur on seasonal timescales, the novel large ensembles created here also provide a set of extremes occurring on shorter weather timescales. Such extreme weather events are of particular importance for decisions surrounding adaptation to climate change. The “H++” scenario concept has been developed to inform such adaptation decisions by considering plausible low likelihood but high impact events that might test the limits to adaptation (114–116). Here we consider how the ExSamples methodology could be used to supplement the UKCP18 PPE with regard to such H++ scenarios by examining a particular ExSamples ensemble member as a case study.

This case study is an example of extreme DJF maximum of daily maximum temperatures averaged over the UK ( $\text{TX}_x$  as previously defined). Figure 6 shows a return period diagram of UKCP 2061-2080  $\text{TX}_x$  deviations (centered on the mean anomaly for 2061-2071 over 2007-2016), plus a fitted generalised extreme value distribution (GEV) and associated uncertainty. GEVs are often used to statistically model block maxima of climate variables; and therefore infer information about the likelihood of such extreme events (117). However, this statistical approach appears to have inadequately accounted for the risk of very high impact events, an issue noted previously by Sippel et al. (118). The dashed dark orange line in Figure 6 shows the  $\text{TX}_x$  for HOT1 ensemble member c0qu, which lies considerably above (by  $2.3\text{ }^{\circ}\text{C}$ ) any UKCP18 PPE samples. This event is roughly 5 standard deviations above the mean of the UKCP18 deviations distribution shown in Figure 6. This is an example of a potential “record-shattering” event as discussed by Fischer et al. (2021). Since the particular GEV fitted to the UKCP18 deviations is type III (100), it sets an upper bound on  $\text{TX}_x$ , consistent with previous studies of extreme heat events (60). However, in a 100,000 member resample bootstrap, the UKCP inferred GEV upper bound is only above this most extreme member in 0.3 % of resamples. This does not appear to be due to a mean bias between the two models: they display near-identical climatological distributions of  $\text{TX}_x$  over the baseline period. However, there are a number of reasons that may explain why this extreme lies well outside the confidence intervals from this statistical extreme value analysis of the UKCP18 PPE. These include: the SST/SIC pattern prescribed being highly conducive to these kinds of hot weather extremes noting that the extreme value analysis is not conditioned on SST/SIC patterns; potential differences in the tails of the  $\text{TX}_x$  distributions simulated by HadAM4 and HadGEM3-GC3.05; and differences in the response of those tails to climate change. We note that this exceptional  $\text{TX}_x$  extreme arises from a very similar set of meteorological circumstances (not shown) to the record-breaking winter temperature extreme that occurred over Europe in 2019 (84, 119). However, we believe that the key point to take away from this is not necessarily the specific estimated likelihood of these extreme weather events, but that the methodology used here could help to provide multivariate spatially, temporally and physically coherent examples of the kinds of H++ scenarios used to consider the limits to adaptation.



**Figure 3.6: Examining statistics of subseasonal extreme weather events.** As return period diagram of Figure 3, but of DJF maximum of maximum daily temperatures averaged over the UK region for the HOT1 winter. The statistical model indicated by the solid and dotted black lines is a generalised extreme value distribution fit over the entire population of UKCP PPE 2061-2080 deviations, which are shown as light orange dots. Note the dotted lines indicate a 0.5 - 99.5 % CI in this instance. The dashed dark orange line shows the value of the most extreme member within the HOT1 ensemble.

## 3.6 Discussion

The first science question we aimed to answer through our experiments is also the most straightforward: is the atmosphere-only HadAM4 model able to simulate the highest extremes observed in the UKCP18 HadGEM3-GC3.05 PPE, or do the differences between the models preclude HadAM4 from producing such events? The answer to this is a confident yes. We have found that HadAM4 is not only able to closely reproduce the present-day climate statistics of the more complex model (after correcting the bias in seasonal mean rainfall, which may be due to model parameterisation), but is able to produce winters just as extreme as the selected study winters when driven by the SST and SICs from those winters.

The question that naturally follows on from this is: were the selected winters genuinely exceptional events, or could they have been more extreme? Despite the fact the selected winters were already far into the tails of the projected climate distribution from UKCP18, the SST/SIC forced ExSamples experiments show that higher extremes are possible. In the two winters pre-conditioned by the SST and SIC patterns, there were more higher extremes than in the winter where the ocean pattern did not contribute to the extreme. Since the ExSamples ensembles are forced by the same lower boundary conditions as the study winters, they cannot be used to determine the unconditional likelihood of these higher extremes, but they do provide plausible and physically consistent scenarios in which such higher extremes might be generated.

We suggest that the ExSamples methodology is more efficient at sampling extremes than the simplest alternative approach of increasing the UKCP18 PPE size. We have found that overall, for both hot and wet extremes, on both seasonal and daily timescales, the future ExSamples ensembles were able to produce many more samples of extreme winters than would be expected if we simply increased the UKCP18 2061-2080 ensemble to be the same size as the ExSamples ensembles. Across the three future ExSamples ensembles, for mean temperature we sampled 44 winters above the most extreme winter in UKCP18, and 106 for mean precipitation (using re-centered deviations to define the UKCP18 maxima as shown in Figures 3-5, S4-S6). However, there is an important caveat to bear in mind here: the SST/SICs taken from the selected study winters clearly “primed” the corresponding ExSamples ensembles towards producing relatively more extremes in two of the three cases (HOT2 and WET), but not in the third (HOT1). For the two primed study winters, the benefits of the ExSamples methodology is clear: we get many more samples of extreme winters than would be expected from an unconditioned ensemble of the same size (like the UKCP18 PPE). In particular, the HOT2 ensemble produces 10 times more samples of 1-in-100 year TXm and PRm events than would be expected for an equal-size UKCP18 PPE (from Table 2). For the third study winter the overall benefits to sampling efficiency are less clear. However, this winter generated a TXx extreme that far exceeds anything seen in the UKCP18 PPE (and indeed anything that would be expected to be seen even if the UKCP18 PPE was considerably larger, based on a statistical extreme value analysis).

In addition to the methodology presented here, the future ExSamples ensembles explored here represent a data set that may be of considerable interest to the wider scientific community, since they provide multivariate spatially coherent information for climate projections of very high return period extremes. These ensembles, and in particular the physically plausible simulations of extremes within, could be used in the context of “H++ scenarios” to explore and understand the potential impacts of climate

change, and the limits to adaptation planning (116). The efficiency with which we have been able to sample extremes with the ExSamples methodology means that we can provide a much richer set of future extreme winter events than exist within the UKCP18 PPE. This rich set of events could be used, for example, by impact modelling, to more fully explore the space of impacts that may arise from climate change.

A final topic that this study touches on is the use of atmosphere-only versus coupled models (48, 49, 120, 121). Here, we have explored both present-day baseline and projected climates from a coupled model (HadGEM3-GC3.05) and a comparable atmosphere-only model (HadAM4). Whilst atmosphere-only simulations have been found to have lower variability of ocean surface air temperature (120) and could potentially exhibit lower variability in other quantities, we have not found this to be the case for the mean UK temperature and precipitation studied here (though definitive proof of this would require us to repeat the ExSamples exercise with the coupled model). For the baseline period, the atmosphere-only model did not systematically underestimate the internal variability of the seasonal (or daily) timescale extreme variables considered here (Figures 3-5, S4-S12). Since we only have ExSamples future ensembles for three different sets of SST/SIC conditions, it is more difficult to quantify whether the projected internal variability is significantly different from the coupled model simulations, but the climate distributions of the relatively unconditioned HOT1 ensemble suggest that this is not the case.

If the ExSamples methodology were to be repeated, for the purpose of sampling additional extremes, being able to pre-select study winters (ie. lower boundary conditions) that condition the resulting ensembles towards extremes would be of considerable value. Here, we simply chose three of the most extreme winters within the UKCP18 PPE, expecting that these would be more likely to produce extremes than a randomly selected winter. This turned out to be the case for two of the winters we chose, but not the third. Understanding what features of the prescribed SST and SIC patterns caused the ensembles to be conditioned towards extremes would be a very useful direction for further study to take. If future research were able to provide evidence of such features, then we could pre-select study winters more intelligently, and therefore sample extremes even more efficiently. There has been some previous work done on the subject of how SST patterns affect seasonal mid-latitude weather that could potentially be used in this manner (Baker et al., 2019). On a related note, our methodology could be used to understand real extremes in the present-day by driving the model with observed rather than simulated SST/SICs. This would allow some exploration of whether extremes that have already occurred might have been even more extreme.

Another research direction that could be taken would be to attempt to extract additional information from the existing set of events provided by the ExSamples ensembles presented here. Although the ~60 km (N216) resolution of both the ExSamples ensemble and UKCP18 PPE is very competitive within the context of the current generation of climate models (122, 123), it is still relatively coarse for providing assessments of weather events on small spatial or temporal scales. For example, catchment-scale hydrological modelling would require much higher spatial resolutions (124). Hence, we suggest that the ExSamples ensembles could be statistically downscaled (or dynamically downscaled using a regional model if suitable model output was stored to drive these models) in order to provide information that is more relevant for localised climate change adaptation planning. Such downscaling could result in an extensive set of extreme local scenarios to complement the raw model output that provides a corresponding set of extreme national scenarios. For downscaling to be trustworthy, the large-scale dynamical features of the

input simulations must be an accurate representation of reality. The analysis that we have performed here suggests this is the case: as demonstrated in the Supplementary Information, the large-scale dynamics over the Euro-Atlantic sector within HadAM4 very closely replicates those within HadGEM3-GC3.05.

## 3.7 Concluding remarks

In this study we have presented a new set of ~1000-member ensembles of simulations from the HadAM4 atmosphere-only model, run on the personal computers of volunteers using a distributed computing system, to allow the study of extreme weather events. The lower boundary conditions of these ensembles were taken from three of the most extreme winters within the UKCP18 PPE between 2061-2080, and they therefore represent a comprehensive sampling of atmospheric internal variability conditioned on the prescribed SST, SIC and anthropogenic forcings. Corresponding ensembles for a 2007-2016 baseline period were also run to enable the HadAM4 model to be verified against the coupled HadGEM3-GC3.05 model used in UKCP18.

We find that the HadAM4 ensembles are able to simulate winters with temperature and precipitation anomalies that exceed the magnitudes of the most extreme examples within the UKCP18 PPE. Conditioning from the prescribed SST/SICs present in two of the three ensembles resulted in significantly more extremes being sampled by these ensembles than would be expected from a UKCP18 PPE-style ensemble of the same size: around 10 times more 1-in-100 year extremes.

The computational efficiency with which our methodology was able to sample such extremes provides a compelling argument for how it could be used to support future climate projection efforts. The ensembles that we have presented here could themselves be used to provide multivariate spatially, temporally and physically coherent examples of extreme weather in the context of H++ scenarios and for adaptation planning. Although we have focussed on the UK in this study, our methodology could be applied to other regions, subject to proper model validation (56, 72).

## 3.8 Chapter close



*Quote*

— author

# 4

## Partial forecast-based attribution

This chapter contains much of the conceptual description of, and motivation for, forecast-based attribution. Using the well-predicted February 2019 heatwave as a case study, I carry out forecasts with the operational medium-range ECMWF model in which I have instantaneously perturbed the CO<sub>2</sub> concentration at initialisation. These perturbed forecasts allow me to estimate the direct contribution of diabatic heating due to CO<sub>2</sub> to the heatwave. This partial attribution provides a proof-of-concept of the forecast-based approach, and I close with a discussion of how I could perform a more complete estimate of anthropogenic influence on a specific extreme event in following work.

**Author contributions:** This chapter is based on the the following publication \*

Leach, N. J., Weisheimer, A., Allen, M. R., & Palmer, T. (2021). **Forecast-based attribution of a winter heatwave within the limit of predictability.** *Proceedings of the National Academy of Sciences*, 118(49), . <https://doi.org/10.1073/pnas.2112087118>

### Contents

4.1	Chapter open	45
4.2	Abstract	45
4.3	Introduction	45
4.4	The 2019 February heatwave in Europe	47
4.4.1	Forecasts of the heatwave	49
4.5	Perturbed CO <sub>2</sub> forecasts	52
4.6	Attributing the heatwave to diabatic CO <sub>2</sub> heating	55
4.7	Discussion	58
4.8	Chapter close	60

---

\*with the author contributing as follows. Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualisation and Writing – original draft



## 4.1 Chapter open

## 4.2 Abstract

Attribution of extreme weather events has expanded rapidly as a field over the past decade. However, deficiencies in climate model representation of key dynamical drivers of extreme events have led to some concerns over the robustness of climate model-based attribution studies. It has also been suggested that the unconditioned risk-based approach to event attribution may result in false negative results due to dynamical noise overwhelming any climate change signal. The “storyline” attribution framework, in which the impact of climate change on individual drivers of an extreme event is examined, aims to mitigate these concerns. Here we propose a novel methodology for attribution of extreme weather events using the operational ECMWF medium-range forecast model that successfully predicted the event. The use of a successful forecast ensures not only that the model is able to accurately represent the event in question; but also that the analysis is unequivocally an attribution of this specific event, rather than a mixture of multiple different events that share some characteristic. Since this attribution methodology is conditioned on the component of the event that was predictable at forecast initialisation, we show how adjusting the lead time of the forecast can flexibly set the level of conditioning desired. This flexible adjustment of the conditioning allows us to synthesise between a storyline (highly-conditioned) and a risk-based (relatively unconditioned) approach. We demonstrate this forecast-based methodology through a partial attribution of the direct radiative effect of increased CO<sub>2</sub> concentrations on the exceptional European winter heatwave of February 2019.

## 4.3 Introduction

Attribution of extreme weather events is a relatively young field of research within climate science. However, it has expanded rapidly from its conceptual introduction (1) over the past twenty years; it now has an annual special issue in The Bulletin of the American Meteorological Society (125). Extreme event attribution is of particular importance for communicating the impacts of climate change to the public (126, 127), since the changing frequency of extreme weather events due to climate change is an impact that is physically experienced by society. As a result of this rapid expansion, there now exists a large number of different methodologies for carrying out an event attribution (128). Many of these rely on large ensembles of climate model simulations, the credibility of which has been questioned by recent studies (47, 129, 130). A particular issue is the dynamical response of the atmosphere to external forcing, which is highly uncertain within these models (131). As attribution studies try to provide quicker results, with an operational system a clear aim, it is vital that any such system provides trustworthy results. In this study we propose a “forecast-based” attribution methodology using medium-range weather forecasts which could provide several key advantages over traditional climate model-based approaches. Firstly, if an event is predictable within a forecasting system, we know that that system is capable of accurately representing the event. Secondly, we know that any attribution performed is unequivocally an attribution of the specific event that occurred; unlike in unconditioned climate model simulations. Finally, weather forecasts are run routinely by many different national and research centers. The models used are generally

state-of-the-art and extensively verified. We propose that the attribution community could and should take advantage of the massive amount of resources that are put into these forecasts by developing methodologies that use the same type of simulation. Ideally, the experiments required for attribution with forecast models would be able to be run with little additional effort on top of the routine weather forecasts; in this way they might provide a rapid operational attribution system. We discuss these ideas further throughout the text.

There have been several studies that propose or perform methodologies related to the forecast-based attribution demonstrated here. Hoerling et al. (20) used two seasonal forecast ensembles to examine the predictability of the 2011 Texas drought/heatwave within a comprehensive attribution analysis involving several different types of types of climate simulation. Meredith et al. (132) used a triply nested convection-permitting regional forecast model to investigate the role of historical SST warming within an extreme precipitation event. They conditioned their analysis on the large-scale dynamics of the event through nudging in the outermost domain. More recently, Van Garderen et al. (133) employed spectrally nudged simulations to assess the contribution of human influence on the climate over the 20th century on the 2003 European and 2010 Russian heatwaves. Possibly the most similar studies to the one presented here are a series of studies by Hope and colleagues (134–136). They used a seasonal forecast model to assess anthropogenic CO<sub>2</sub> contributions to record-breaking heat and fire weather in Australia. Two more similar studies carried out forecast-based hurricane attribution studies (137, 138). Tropical cyclones are a natural candidate for forecast-based methodologies due to the high model resolution required to represent them accurately, if at all. A final distinct, but related study is Hannart et al. (139), which proposes the use of Data Assimilation for Detection and Attribution (DADA). They suggest that operational causal attribution statements could be made in a computationally efficient manner using the kind of data assimilation procedure carried out by weather centers (to initialise forecasts) to compute the likelihood of a particular weather event under different forcings (these would be observed and estimated pre-industrial forcings for conventional attribution). Our forecast-based framework differs from these other studies in several regards. Firstly, we use a state-of-the-art forecast model to perform the attribution analysis of the event in question; rather than to solely assess the predictability of the event. We use free-running coupled global integrations here, allowing the predictable component at initialisation to dynamically condition the ensemble; as opposed to nudging our simulations towards the dynamics of the event, using nested regional simulations, or using the highly observationally constrained output of data assimilation procedures. A final key difference is that here we present an attribution of the direct radiative effect of CO<sub>2</sub> in isolation, though we hope that our approach could be extended in the future to provide an estimate of the full anthropogenic contribution to extreme weather events as in these other studies. We argue that the relative simplicity in the validation, setup and conditioning of our simulations is desirable from an operational attribution perspective; and flexible across many different types of extreme event.

We begin by introducing the chosen case study, the 2019 February heatwave in Europe, describing its synoptic characteristics and formally defining the event quantitatively. We then demonstrate the predictability of the event within the ECMWF ensemble prediction system, showing that this operational weather forecast was able to capture both the dynamical and thermodynamical features of the event. In [Perturbed CO<sub>2</sub> forecasts](#), we outline the experiments we have performed in order to quantitatively determine the direct CO<sub>2</sub> contribution to the heatwave. We then provide quantitative results from these experiments, and finally conclude with a discussion of the strengths and potential

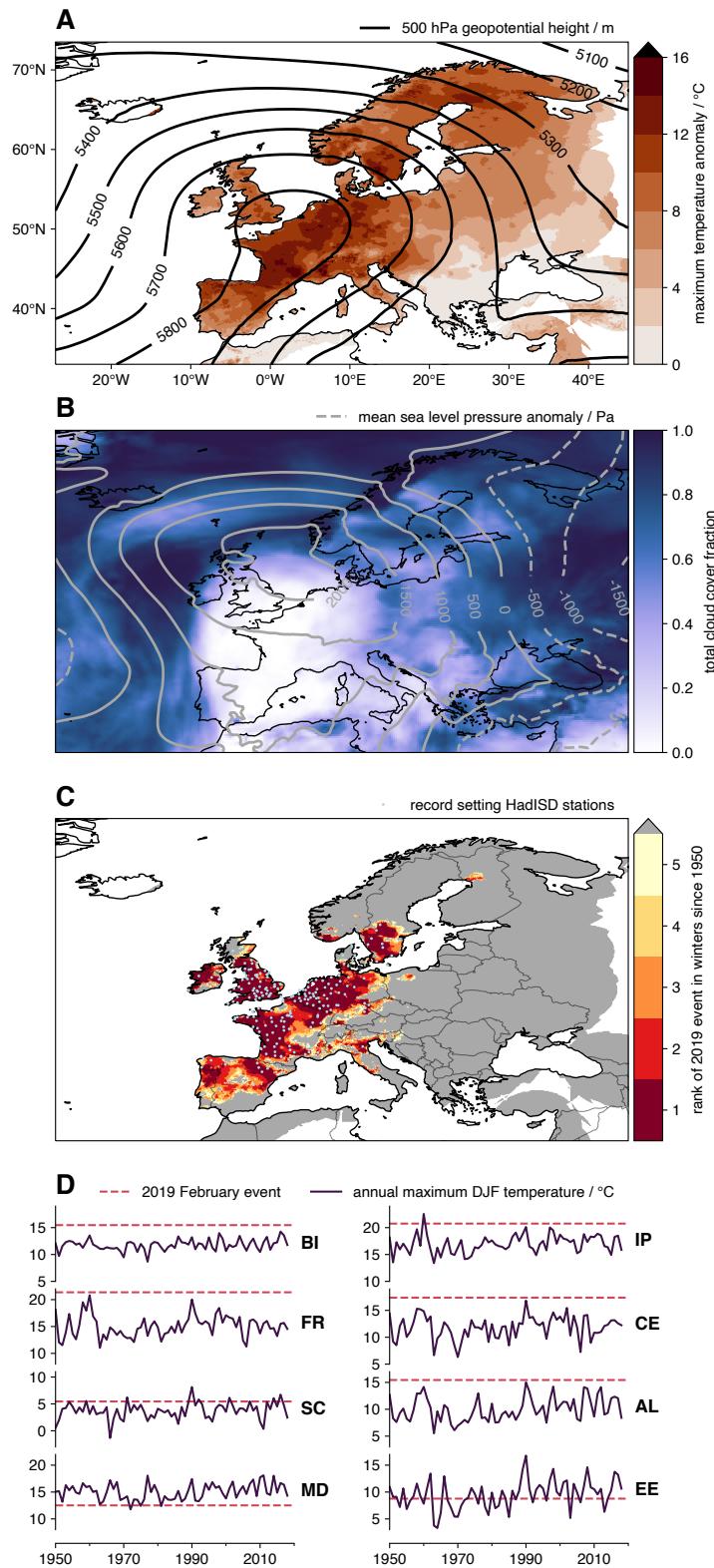
issues of our forecast-based attribution methodology, including our proposed directions for further work.

## 4.4 The 2019 February heatwave in Europe

Between the 21st and 27th February 2019, climatologically exceptional warm temperature anomalies of 10-15 were experienced throughout Northern and Western Europe (119), as shown in Fig. 1A. In particular, the 25th - 27th February saw record-breaking temperatures measured at many weather stations and over wide areas of Iberia, France, the British Isles, the Netherlands, Germany and Southern Sweden, as shown in Fig. 1C (140). Fig. 1D, comparing the regional mean maximum temperatures during the 2019 heatwave with timeseries of winter mean maximum temperatures between 1950 and 2018, illustrates just how unusual and widespread the event was. This heat was associated with a characteristic flow pattern: a narrow tilted ridge extending from north-west Africa out to the southern tip of Scandinavia, advecting warm subtropical air north-east (141), as shown in the geopotential height field in Fig. 1A. This dynamical driver was accompanied by another synoptic feature that further enhanced the warming: widespread clear skies between the 25th - 27th, shown in Fig. 1B. These clear skies resulted in a widespread and persistent strong diurnal cycle, reaching 20 in some locations. Further details of the meteorological mechanisms and historical context of the heatwave are provided in refs. (84, 119, 142).

In order to quantify the direct impact of CO<sub>2</sub> on the heatwave in question within this study, we need to characterise the heatwave in an “event definition”. The choice of event definition is subjective but can impact on the quantitative results of an attribution study significantly (50, 143, 144). The most remarkable feature of the February 2019 heatwave were the maximum temperatures observed, which peaked between the 25<sup>th</sup> and 27<sup>th</sup> for the majority of the affected area. Focusing on this relatively short time-period ensures that the synoptic situation driving the heat is coherent throughout the event definition window. For the spatial extent of the event, we use the eight European sub-areas described in ref. (30). The use of regions previously defined in the literature aims to avoid selection bias. Our resulting event definition is as follows: the hottest temperature observed between 2019-02-25 and 2019-02-27, then averaged over the land points within each region (the temporal maximum is calculated before the spatial averaging). Although we carry out our calculations for all sub-areas, several regions were characteristically very similar in terms of both the event itself, and the forecasts of the event. We therefore focus on three of the eight regions: the British Isles (BI), which experienced exceptional heat and was well predicted; France (FR), which experienced exceptional heat but where the magnitude of the heat was less well forecast; and the Mediterranean (MD), which experienced well-predicted but climatologically average heat.

#### 4.4. The 2019 February heatwave in Europe



**Figure 4.1: The 2019 February heatwave in Europe: synoptic characteristics & historical context.**

#### 4.4.1 Forecasts of the heatwave

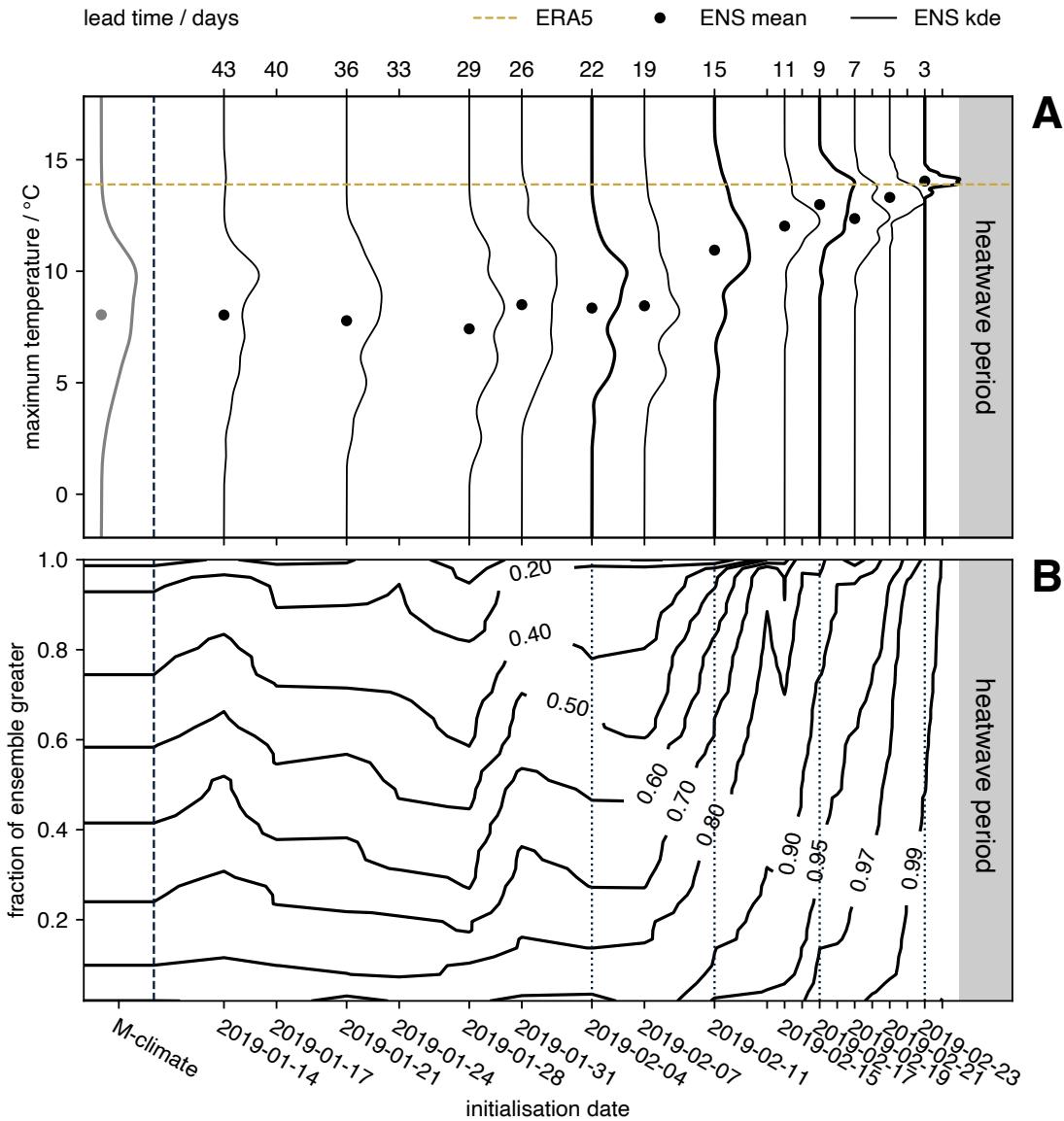
This heatwave was well-predicted by the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system. Their forecasts indicated “extreme” heat was possible at a lead time of around two weeks, and probable at a lead time of around ten days (Fig. 2A), despite the exceptional nature of the heatwave in both the model climatology and real world. As expected, the forecast’s performance in predicting the extreme heat at the surface is reflected in variables more closely linked to the dynamic drivers of the heat, such as 500 hPa geopotential height (Fig. 2B).

This successful forecast is a crucial part of our study as it means that we are not only confident that the model used is able to simulate the event in question; but that we are unequivocally performing an attribution analysis of the specific winter heatwave that occurred in Europe during February 2019. This is an important distinction to the framework used in “conventional” or “risk-based” (131) attribution studies (19, 79, 143, 145), which in general reduce the event to some impact-relevant quantitative index , then estimate the increase in likelihood of events that exceed the magnitude of the event in question. For example, a heatwave attribution study may choose to define the event as the hottest observed temperature during the heatwave, and then compute the attributable change in likelihood of temperatures hotter than this recorded maximum (eg. using models or historical records). While this does provide useful information, it does not answer the question of how much more likely anthropogenic activities have made the *specific* heatwave that occurred, rather the question of how much more likely anthropogenic activities have made a mixture of events that share one or more characteristics. Studies have attempted to provide a more satisfactory answer to this first question by including a level of conditioning on the set of events considered by using circulation analogues (107), or by nudging model simulations towards the specific dynamical situation that occurred during the event in question (132, 133). Here we are evidently performing an attribution study of the specific record-breaking heatwave that occurred in February 2019 due to the use of these successful forecasts, that not only captured the heat experienced at the surface, but also the dynamical drivers behind the heat.

As well as enabling us to answer the attribution question for a single specific heatwave, the use of a numerical weather prediction model provides additional benefits. Since large model ensembles are required to properly capture the statistics of extreme events, many previous attribution studies, especially in the context of heatwaves, have used relatively coarse, atmosphere-only climate models (33, 34, 40), which may not fully capture all the physical processes required to credibly simulate the extreme in question (18). In particular, the use of atmosphere-only simulations may result in the full space of climate variability being under-sampled due to the lack of atmosphere-ocean interaction (48). This can lead to studies overestimating the impact of anthropogenic activity on weather extremes (129, 143). More generally, Bellprat et al., and Palmer and Weisheimer (47, 130) have shown the importance of initial-value reliability in model ensembles underlying robust attribution statements. Model evaluation is therefore a key part of any robust model-based attribution study. Here, the demonstrably successful forecast enables us to be confident that the model used is providing credible realisations of the event.

A clear distinction between the typical climate model simulations used for attribution (33, 40) and the forecasts used here is that the climate model simulations are usually allowed to spin out for a sufficient length of time such that they have no memory of their initial conditions; an ensemble constructed in this way will therefore be representative of the climatology of the model. If such simulations use prescribed-SST boundary conditions,

then the ensemble will be representative of the climatology conditioned on the prescribed SST pattern (34). Unlike climatological simulations, a successful forecast is conditioned upon the component of the weather that is predictable at initialisation. In general, the level of conditioning imposed upon the ensemble by the initial conditions reduces as the model integrates forwards from the initialisation date. Hence a forecast ensemble initialised only a few days before an event will be much more heavily conditioned (and therefore much less spread) than one initialised weeks before. As the lead time increases, a forecast ensemble will tend towards the model climatology, analogous to the climate model simulations discussed above. We can relate these situations to the two broad attribution frameworks discussed in (131): very long lead times, where the forecast simulates model climatology, are analogous to “conventional” attribution; while short lead times, in which the forecast ensemble is heavily dependent on the initial conditions and therefore conditional on the actual dynamical drivers that lead to the extreme event, are analogous to the “storyline” approach in (133, 146). In order to synthesize between these two frameworks, here we have chosen 4 initialisation times (3-, 9-, 15-, and 22-day leads) for our experiments that span the range from a near-unconditioned climatological forecast to a short-term forecast that is tightly conditioned on the actual dynamical drivers of the heatwave.



**Figure 4.2: Medium- to extended-range forecasts of the heatwave.**

## 4.5 Perturbed CO<sub>2</sub> forecasts

In this study we choose to only change one feature of the operational forecast in our experiments: the CO<sub>2</sub> concentration. This means that the analysis we carry out is limited to attributing the impact of diabatic heating due to increased CO<sub>2</sub> concentrations above pre-industrial levels just over the days between the model initialisation date and the event. Although this results in a counterfactual that does not correspond to any “real” world (since it is one with approximately present-day temperatures but pre-industrial CO<sub>2</sub> concentrations), and thus reduces the relevance of our analysis to stakeholders or policymakers; it does significantly increase the interpretability of our results, and remove a major source of uncertainty associated with a “complete” attribution to human influence: the estimation of the pre-industrial ocean and sea-ice state vector used to initialise the model (147). Here we define a complete attribution as an estimate of the total impact of human influence on the climate arising from anthropogenic emissions of greenhouse gases and aerosols since the pre-industrial period. For each lead time chosen, in addition to the operational forecast (ENS) we run two experiments using operational initial conditions and identical to the operational forecast in every way except the experiments have specified fixed CO<sub>2</sub> concentrations. One experiment has CO<sub>2</sub> concentrations fixed at pre-industrial levels of 285 ppm (PI-CO<sub>2</sub>), while in the other they are increased to 600 ppm (INCR-CO<sub>2</sub>). These represent approximately equal and opposite perturbations on global radiative forcing (148). We carry out these two experiments for each lead time, perturbing the CO<sub>2</sub> concentration in opposite directions, to ensure that any changes to the likelihood of the event can be confidently attributed to the changed CO<sub>2</sub> concentrations. It is possible that, due to the chaotic nature of the weather, the operational conditions were ideal for generating the observed extreme, and any perturbation to the dynamical system would reduce the likelihood of its occurrence (131). If this were the case we would see a reduction in event probability regardless of whether we increased or reduced the CO<sub>2</sub> concentration.

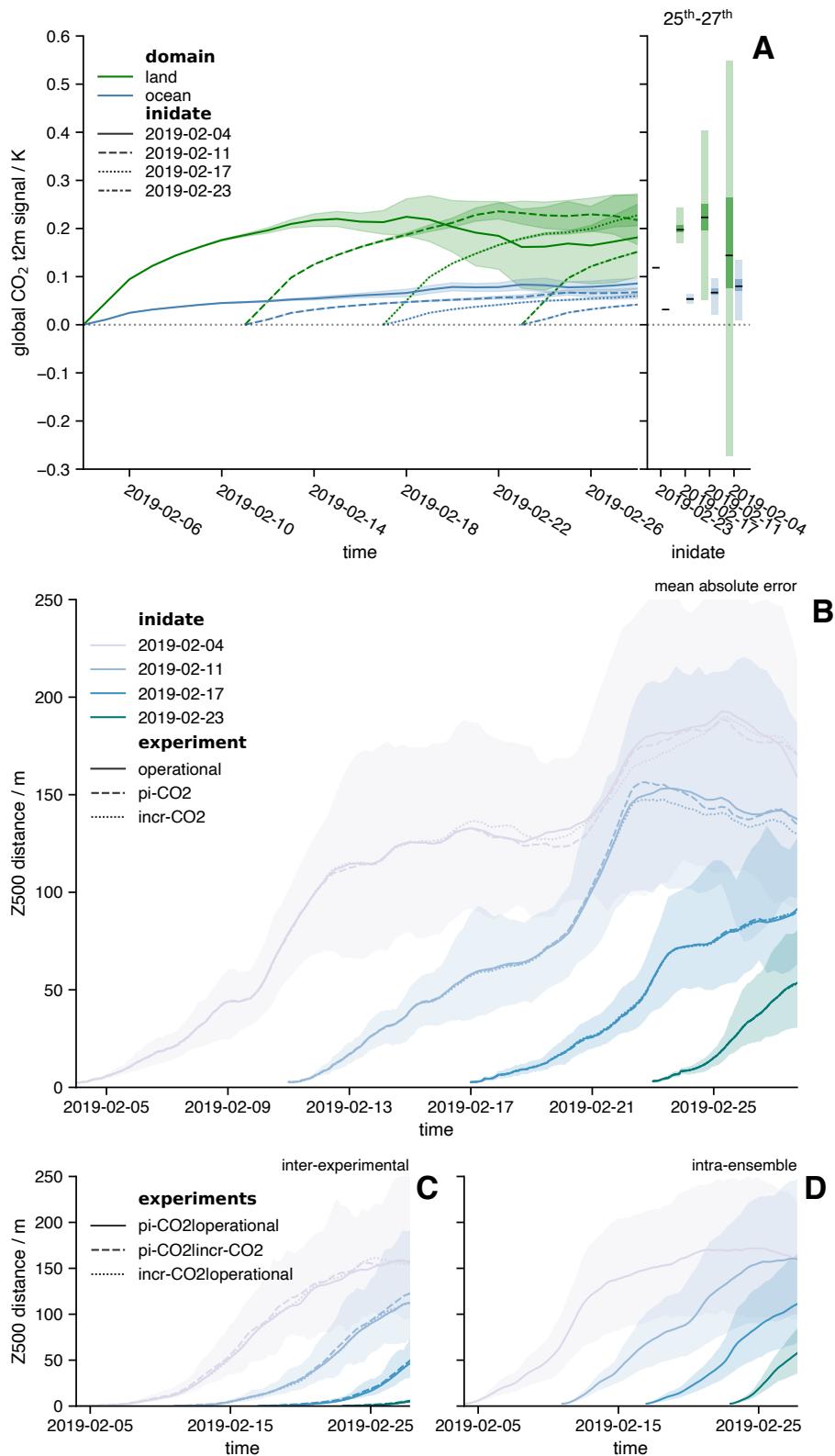
Some previous work has been done on the impact of reduced CO<sub>2</sub> concentrations in the absence of changes to global sea surface temperatures. Baker et al. (149) explored how temperature and precipitation extremes were affected by the direct effect of CO<sub>2</sub> concentrations (defined there as all the effects of CO<sub>2</sub> on climate beside those occurring through ocean warming), finding the direct effect of CO<sub>2</sub> increases risk of temperature extremes, especially within the Northern hemisphere summer. Our experimental design is also reminiscent of some of the earliest work done on investigating the impact of CO<sub>2</sub> on climate in global circulation models (150, 151). This work found that, in the absence of changes to sea surface temperatures or sea ice concentrations, a doubling of CO<sub>2</sub> concentrations would change global mean surface temperatures over land by  $\sim 0.4$ . These early studies indicate that changes in global land temperatures are approximately linear with the logarithm of CO<sub>2</sub> concentration.

We find that the best-estimate global mean change in land surface temperatures attributable to the additional diabatic heating due to CO<sub>2</sub> over pre-industrial levels (henceforth the “CO<sub>2</sub> signal”, calculated as half the difference between the two experiments for a particular variable) at a lead time of two weeks (over the final 5 days of the forecasts initialised on 2019-02-11) is 0.22 [0.20, 0.25] (square brackets indicate a 90 % confidence interval throughout). In general, the further away from the initialisation date, the slower the rate of change of the globally-averaged ensemble mean CO<sub>2</sub> signal, and the larger the ensemble spread (Fig. 3A). While in experiments with prescribed SSTs, we might

expect the CO<sub>2</sub> signal in surface temperatures to approach a maximum value within timescales on the order of months, in our experiments the CO<sub>2</sub> signal will likely continue to increase in magnitude for centuries due to the ocean-coupling, as is the case in the abrupt-4xCO<sub>2</sub> experiment carried out in CMIP (123, 152, 153). The zonal-mean patterns of surface temperature CO<sub>2</sub> signal are qualitatively similar to those exhibited by CMIP5 and CMIP6 models during the abrupt-4xCO<sub>2</sub> experiment (154, 155), despite the considerably shorter timescales involved: small and very confident changes in the tropics become larger but much less confident changes at the poles. This heterogeneity in the zonal distribution of warming appears to originate in the zonal distribution of the lapse-rate feedback; the weekly timescales of these experiments is insufficient for the surface-albedo feedbacks to have any significant impact (156).

We also examine the impact on the specific event dynamics over our region of interest; since these were crucial in developing the extremes observed. Fig. 3B shows the growth in Z500 errors (measured as the mean absolute distance from ERA5 over the European domain) for each of the experiments. This figure illustrates that there are no clear differences in the ability of each experimental ensemble to predict the dynamical characteristics of the event. In other words, we have not made the synoptic event any more or less likely as a result of our perturbations. This is crucial as it means that we can consider any changes to the magnitude of the temperatures observed to be entirely due to the thermodynamic effect of changed diabatic CO<sub>2</sub> heating, and not due to the attractor of the dynamical system having changed as a result of the perturbations we have made.

Figs. 3C and 3D show analogous plots to 3B, but for inter-experimental and intra-ensemble errors respectively. These indicate a couple of important features. Firstly, no two experiments are more similar than any other two; the magnitude of Z500 distances in Fig. 3C are near identical for all lead and validation times. Secondly, the error growth due to the CO<sub>2</sub> perturbation is slower than due to the initial condition perturbations; the errors in Fig. 3C increase slower than in 3D. However, by the end of the longest lead forecast, we can see that the intra-ensemble errors have saturated, and the inter-experimental errors have grown to be the same magnitude. The saturation of intra-ensemble errors by the end of this lead time reinforces our assertion that at this lead the forecast is a good approximation of a climatological simulation.



**Figure 4.3: Global temperature and synoptic-scale dynamical response to  $CO_2$  perturbations.**

## 4.6 Attributing the heatwave to diabatic CO<sub>2</sub> heating

First, we examine the geographical pattern of the CO<sub>2</sub> signal in the heatwave in Fig. 4A – D. These indicate several key features of the attributable direct CO<sub>2</sub> effect on the heatwave. The CO<sub>2</sub> effect tends to grow with lead time, consistent with its impact on global mean temperatures. It is generally stronger over land than ocean, also consistent with global mean temperatures. Finally, the ensemble tends to become less confident in its effect as the lead time increases and the ensemble members diverge. The CO<sub>2</sub> signal magnitude in the heatwave generally exceeds the signal in the global mean surface temperature (Fig. 3A), in particular in Central Europe; possibly due to the high contribution of diabatic heating to the heatwave arising from ideal dynamical conditions. Fig. 4E shows boxplots of the heatwave CO<sub>2</sub> signal for the three regions of interest plus the full European land area. Although there is some region-specific variability, these reinforce the main messages illustrated by the maps: the CO<sub>2</sub> signal grows and decreases in confidence as the lead time increases.

In addition to the absolute impact of the direct CO<sub>2</sub> effect on the heatwave, we also carry out a probabilistic assessment of its impact, consistent with conventional “risk-based” attribution studies (131, 157). Due to the novel approach we are taking within this study, it is worth clarifying exactly what question we are answering with this probabilistic analysis. The specific question is: “given the forecast initial conditions, how did the direct impact of increased CO<sub>2</sub> concentrations compared to pre-industrial levels just over the days between initialisation and the heatwave itself change the probability of temperatures at least as hot as were observed?”. Using conventional attribution terminology, we call the operational forecast ensemble of the event as our “factual” ensemble, and the pre-industrial CO<sub>2</sub> experiment as our “counter-factual” ensemble. We calculate the probability of simulating an event at least as extreme as observed in the factual ensemble,  $P_1$ , and in the counterfactual ensemble,  $P_0$ . These probabilities are estimated by fitting a generalised extreme value distribution to the 51-member ensemble in each case. We then express the change in event probability as a risk-ratio,  $RR = P_1/P_0$ , which represents the fractional increase in the likelihood of an event at least as extreme as observed in the factual ensemble over the counterfactual ensemble (19, 112). Uncertainties are estimated with a 100,000 member bootstrap with replacement, rejecting samples for which the probability of the event in the factual ensemble is zero. The resulting risk-ratios are shown in Fig. 4F. There are several key factors that contribute to the best-estimate and confidence in the risk ratios: the CO<sub>2</sub> signal growth with lead time; the ensemble spread growth with lead time; how extreme the event was; and how well-forecast the event was. The larger the CO<sub>2</sub> signal, the greater the increase in risk; the larger the ensemble spread, the lesser the increase in risk and the lower the confidence; the more extreme the event, the greater the increase in risk; and the better the forecast (ie. the closer the event to the ensemble centre), the greater the confidence.

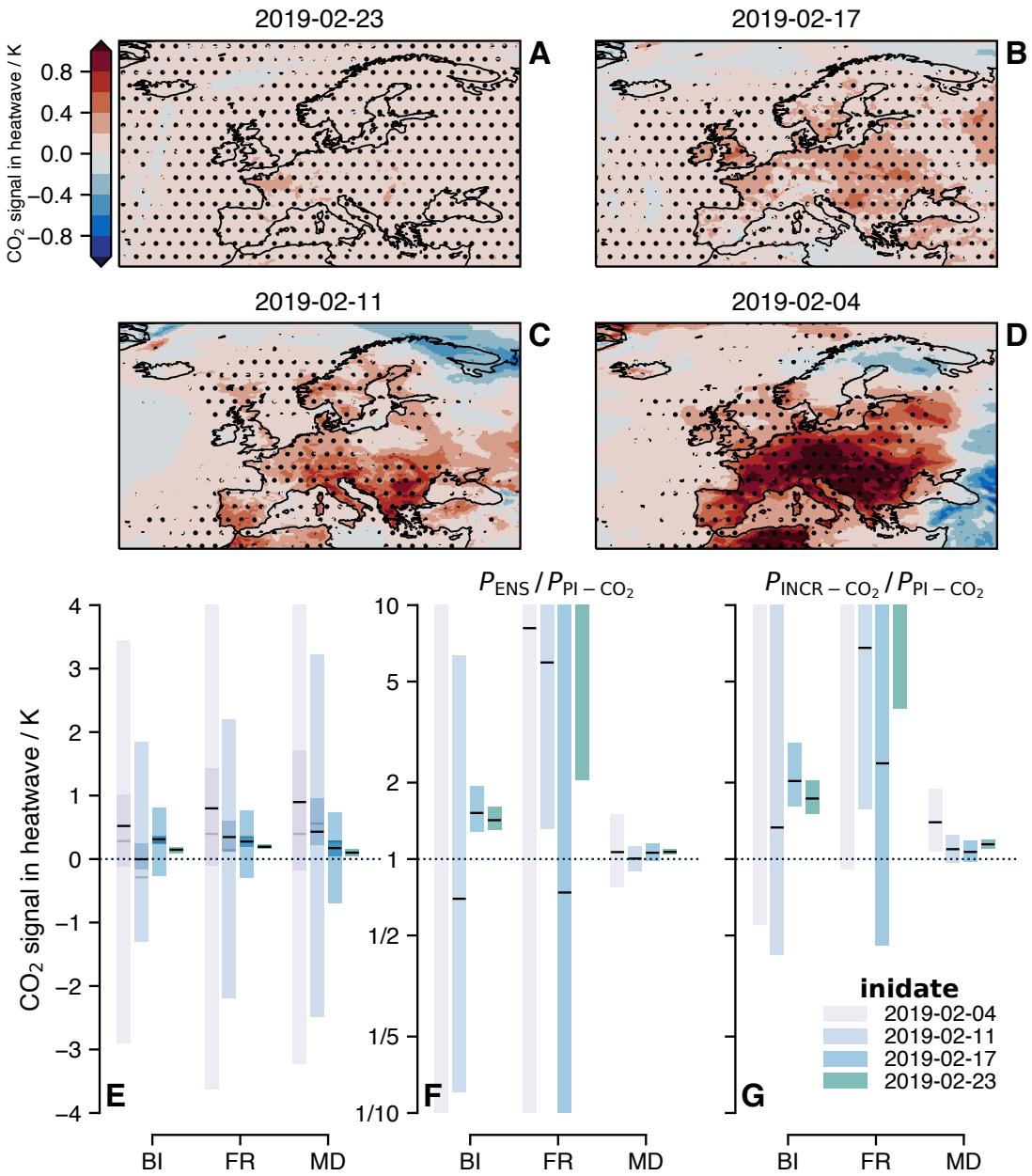
We find that on the shortest lead time, the direct CO<sub>2</sub> effect increases the probability of the event over all European regions (significant at the 5 % level based on a one-sided test). For the well-forecast event experienced over the British Isles, the direct CO<sub>2</sub> effect increases the probability of the extreme heat by 42 [30 , 60] %. For the France heatwave, which was well-forecast given its exceptional nature, but for which the ensemble did not quite reach the total magnitude of the heat experienced, the event probability increased by at least 100 % (5<sup>th</sup> percentile), but with a very wide uncertainty range. Finally, for

the least remarkable but relatively well-forecast event over the Mediterranean, the direct impact of  $CO_2$  increased the event probability by 6.7 [4.6 , 9.7] %. These results from the very short lead experiments represent very highly conditioned statements: in both ensembles the dynamical evolution of the event was near-identical (pattern correlation of > 0.99 for all ensemble members, Fig. 2B).

Moving out to the longer lead times, we find that the confidence in the change in event probability decreases almost ubiquitously. This is as expected, since the further we move away from the event, the less highly conditioned our ensemble is, and the more dynamical noise we are adding to the system (131). However, for the 9-day lead forecast, the uncertainty is low enough to have confidence in the results for the majority of study regions. In particular, the British Isles heatwave, for which the 9-day lead forecast was better than several of the regional 3-day lead forecasts (as measured by the Continuous Ranked Probability Skill Score), increases in probability by 52 [29 , 94] % due to the direct  $CO_2$  effect. However, for France the uncertainty range is so large that based on these results alone we would have no confidence in the direction of the  $CO_2$  effect. Moving further out to the 15- and 22-day lead forecasts, this loss in confidence becomes more pronounced, especially for the British Isles region. For this region, we can get virtually no useful information out of these probabilistic results for the two longest lead experiments. This drop-off in confidence arises due to the increasing ensemble spread from dynamical noise, and large reduction in the number of factual ensemble members able to simulate an event as hot as occurred in reality between the 9- and 15-day leads. A similar, though generally less pronounced drop-off in confidence is found in all other regions.

We can make use of our INCR- $CO_2$  experiment to increase our confidence that the positive results we obtained in the probabilistic analysis above are in fact due to the direct  $CO_2$  effect, and not just random variability. If  $CO_2$  were driving the changes in event probability between the PI- $CO_2$  and operational forecasts, then we would expect to see an even more dramatic increase in event probability between the PI- $CO_2$  and INCR- $CO_2$  forecasts. This is indeed what we find. For all regions and lead times, our best-estimate change in event probability is above zero when  $CO_2$  concentration is increased from pre-industrial levels of 285 ppm to 600 ppm. This therefore increases our confidence further that the positive attribution to  $CO_2$  under high conditioning is genuinely significant. From these results, it also appears that there is a general trend of change in event probability increasing as the forecast lead increases, similar to the absolute impact of the direct  $CO_2$  effect trend; though it is still somewhat masked by uncertainty.

An important caveat on all of these results, probabilistic and absolute, is that they represent a lower bound on the estimate of the direct  $CO_2$  effect. As is clear from the development of the  $CO_2$  signal estimates with lead time, the model is still adjusting to the sudden change in  $CO_2$  concentration (and would continue to do so for centuries due to the very long deep ocean equilibration timescales). Hence we would expect the “full” effect of  $CO_2$  to be greater than the estimates we present here. This is consistent with a recent study that used unconditioned climate model simulations to carry out an attribution of the complete anthropogenic contribution to the same event, which produced much higher estimates of the risk-ratio (142).



**Figure 4.4: Attribution of the direct  $\text{CO}_2$  influence on the heatwave.**

## 4.7 Discussion

Here we have presented a partial, forecast-based attribution of the European 2019 winter heatwave. Taking advantage of successful medium-range forecasts from ECMWF, we used a state-of-the-art numerical weather prediction model that was demonstrably able to predict the event to attribute the direct impact of CO<sub>2</sub> through diabatic heating over pre-industrial levels and just over the days immediately preceding the event on the high temperatures experienced in several regions of Europe. We explored how the level of dynamical conditioning imposed can be specified by changing the lead time of the forecasts. Finally, we presented our quantitative results using two different approaches: measuring the attributable absolute and probabilistic impacts of CO<sub>2</sub>; inspired by the “storyline” and “risk-based” attribution frameworks (19, 131, 157, 158).

There are several advantages associated with this novel forecast-based attribution methodology, compared to conventional climate model based attribution. One simple advantage is that forecast models generally represent the technological peak within the spectrum of General Circulation Models. They tend to have a higher resolution than the models used for climate simulation. In addition, the forecast model used here is coupled, while the large climate model ensembles used for attribution tend to use prescribed sea surface temperatures (34). The use of prescribed SSTs can lead to model biases that project strongly onto attribution results (48). A final advantage arising from the use of an operational forecast model is the wealth of literature and model analysis that will already be available before an attribution study is initiated. As well as these advantages associated with the type of model there is the crucial advantage associated with using successful forecasts: the specific and intrinsic model verification. Due to the difficulty in fully quantifying how well climate models can represent an individual specific event (in particular, the very large ensembles required to have a large enough sample of characteristically similar events), climate model based attribution studies tend to perform statistical model evaluations; or/and account for this uncertainty through multi-model ensembles (159). On the other hand, if a forecast model that demonstrably predicted the event as it occurred is used, no further model verification or evaluation is required to test whether the model is capable of producing a faithful representation of the specific event.

Related to this intrinsic verification is an important point on the framing of forecast-based attribution studies. Climate model based attribution studies tend to characterise an event in terms of some quantitative index closely related to the impact of the event (such as the maximum temperature observed during a heatwave). They then use climate model simulations to determine how climate change has affected the probability of observing an event at least as extreme as the actual event. This is often done without imposing any dynamical conditioning on the simulations, though this is an area of active research (107, 160). This unconditional approach means that the specific question being answered is not “how has anthropogenic climate change affected the probability of event X?”, but “how has anthropogenic climate change affected the probability of all events that are at least as extreme as event X in terms of the index used to define X?”. This second question does not fully answer the question of how climate change has affected the actual event that the study is concerned with. In contrast, the use of a forecast model that predicted the event ensures that any attribution analysis is unequivocally an attribution of that specific event (136).

In addition to its advantages, this novel forecast-based attribution methodology also has associated issues that must be overcome. Firstly, the forecast model must have

produced a “good” forecast of the event. If the model is unable to represent the event as it happened, then we cannot have confidence in any estimates of the impact of climate change on that event. Issues can arise even in qualitatively “good” forecasts, such as the forecast of the heatwave over France in this study. As very few ensemble members, if any, exceeded the observed magnitude of the event for this region, the confidence in our estimates of the probabilistic impact of CO<sub>2</sub> on the event is extremely low (since we are extrapolating the distribution shape outside of the range of our data). Although the estimates of the absolute impact of CO<sub>2</sub> do not share this lack of confidence, this is still a problem. It is possible that applying some bias correction procedure (e.g. 44, 161, 162) based on the model climatology to the model output before analysis might alleviate these issues to some extent, but not if the model is simply unable to predict the event in question (ie. a forecast bust). Secondly, the short timescales involved in these medium-range forecasts mean that the interpretation of any results becomes more difficult as the model is still adjusting to the perturbations imposed (134), at least in the case of the CO<sub>2</sub> perturbations applied here. This adjustment is clear on a global scale in Fig. 3A. Due to this incomplete adjustment, any quantitative statements of attribution represent a lower bound on the “true” value.

We have shown that the direct effect of CO<sub>2</sub> concentrations over pre-industrial levels on the February heatwave is significant, even on timescales as short as a few days. Based on the very good 9-day lead forecast of the heatwave over the British Isles, the region that saw the most climatologically exceptional event, the direct effect of CO<sub>2</sub> was to increase the magnitude of the heatwave by 0.31 [0.24 , 0.37] K, and the conditional probability of the heatwave by 52 [29 , 94] %. It is very important to bear in mind that this statement of risk is highly dynamically conditioned (Fig. 2B). These estimates of the impact of CO<sub>2</sub> on the heatwave follow the storyline attribution framework, since we have effectively removed the dynamical uncertainty from our simulations with this strong conditioning imposed by the short lead time (131, 158, 163). Our longer, 22-day lead experiments can contrast this storyline analysis with relatively unconditioned results much closer to the climatological simulations typically used in the conventional “Risk-based” attribution framework (19, 159). At this lead, we find that although over all regions the best-estimate impact of the direct CO<sub>2</sub> effect is to enhance the heatwave by approximately 0.5 K, in none of the regions is this impact significantly positive at the 90 % level (based on the bootstrapped confidence in the median value). Corresponding estimates of the risk ratio have so low confidence that they provide virtually no useful information. Increasing the forecast ensemble size, which is small compared to the climate model ensembles used in most attribution studies, would increase the confidence, potentially resulting in useful quantitative estimates of the risk ratio even at these longer lead times. Our results illustrate some of the concerns voiced recently over the conventional risk-based approach to attribution (131, 157). Due to the dynamical noise present in unconditioned ensembles, it is possible to obtain an inconclusive attribution within a conventional risk-based framework, and at the same time obtain a confident positive attribution if the dynamical uncertainty is removed through conditioning (in our case achieved by reducing the forecast lead).

While this study provides a demonstration of the potential use for forecast models within attribution science, it remains a partial attribution to the direct CO<sub>2</sub> effect only. For forecast-based attribution to provide results that are fully comparable to conventional climate model-based attribution, we will need to demonstrate how the complete anthropogenic contribution to an extreme event could be estimated with successful forecasts. The next step to progress forecast-based attribution further will be to remove

an estimate of the anthropogenic contribution to ocean temperatures from the model initial conditions (e.g. 147). If performed in addition to reducing other greenhouse gas concentrations and aerosol climatology down to their pre-industrial levels, this should allow us to run pre-industrial forecasts of an event. This has been done previously for a seasonal forecast model by Hope et al. (134–136). They removed the anthropogenic signal from 1960 onwards from the initial conditions, but we could in principle remove the signal from pre-industrial times onwards in order to estimate the complete anthropogenic contribution to an event. Although it is highly likely that there will be methodology specific issues that arise in this direction, we suggest that being able to estimate the complete anthropogenic contribution to an extreme event using a forecast model that was able to predict the event in question would be extremely valuable. Developing a methodology to allow us to do so might also provide a pathway to operational attribution being able to be carried out by weather prediction centres, due to the routine frequency at which they produce forecasts. In addition to attempting a “complete” forecast-based attribution of an extreme event, we would like to explore how increasing the ensemble size may allow us to provide confident forecast-based attribution analyses within the unconditioned risk-based framework (ie. at long forecast lead times). One potential avenue to allow us to do this efficiently might be to reduce the resolution of the forecasts, though this would not be appropriate if it reduced the ability of the model to represent the event in question. On a similar note, we would also like to extend our experiments out to seasonal timescales. This would reduce the issues with the interpretation of our medium-range results that occurred due to the model adjustment to the sudden changes to the CO<sub>2</sub> concentration. It is possible that seasonal forecasts have the greatest potential to target for an operational forecast-based attribution methodology.

## 4.8 Chapter close

*Quote*

— author

# 5

## Forecast-based attribution

Chapter description.

**Author contributions:** This chapter is based on the the following publication \*

Surname, I1. I2., Surname, I1. I2. (year). **Title.** *Journal*, vol(issue), pages. DOI

### Contents

5.1 Section .....	62
-------------------	----

---

\*with the author contributing as follows.

## 5.1 Section

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

*Quote*

— author

# 6

## Discussion

Chapter description.

### Contents

6.1	An overview of this thesis . . . . .	64
6.2	This thesis in the context of previous work . . . . .	64
6.3	Section . . . . .	64
6.4	Limitations . . . . .	64
6.5	Future research directions . . . . .	65
6.5.1	Assimilating the perturbations . . . . .	65
6.5.2	Alternative extremes . . . . .	65
6.5.3	Alternative forecast models . . . . .	65
6.5.4	Incorporating perturbation uncertainty . . . . .	65
6.5.5	Projections of future extremes . . . . .	65
6.5.6	Impact assessment . . . . .	66
6.6	Concluding remarks . . . . .	66

## 6.1 An overview of this thesis

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 6.2 This thesis in the context of previous work

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 6.3 Section

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 6.4 Limitations

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 6.5 Future research directions

### 6.5.1 Assimilating the perturbations

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### 6.5.2 Alternative extremes

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### 6.5.3 Alternative forecast models

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### 6.5.4 Incorporating perturbation uncertainty

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### 6.5.5 Projections of future extremes

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest

gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

### **6.5.6 Impact assessment**

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## **6.6 Concluding remarks**

# **Appendices**



# A

## Resources

Examples of output from the thesis besides research:

- mystatsfunctions plus other packages (eg. FairR)
- Code accompanying papers
- CEDA archive data
- MARS datasets



# B

## Appendix 2



*Quote*

— author

## References

- <sup>1</sup>M. Allen, “Liability for climate change”, *Nature* **421**, 891–892 (2003) (pp. 2, 45).  
**A seminal paper that introduces the idea of extreme event attribution, including the litigative motivation behind such attribution, and the proposed methodology, grounded in probabilistic concepts from epidemiology.**
- <sup>2</sup>F. E. Otto, “Attribution of Weather and Climate Events”, *Annual Review of Environment and Resources* **42**, 627–646 (2017) (p. 2).
- <sup>3</sup>M. Allen, P. Pall, D. Stone, P. Stott, D. Frame, S.-K. Min, T. Nozawa, and S. Yukimoto, “Scientific Challenges in the Attribution of Harm to Human Influence on Climate”, *University of Pennsylvania Law Review* **155**, 1353–1400 (2007) (p. 2).  
**A useful paper aimed at legal professionals that discusses how extreme event attribution could be used in climate change litigation suits. Includes some references to operational attribution and analogy to weather forecasting. Ends with questions for the legal community that may be of interest.**
- <sup>4</sup>E. A. Lloyd and N. Oreskes, “Climate Change Attribution: When Is It Appropriate to Accept New Methods?”, *Earth’s Future* **6**, 311–325 (2018) (p. 2).
- <sup>5</sup>E. A. Lloyd and T. G. Shepherd, “Environmental catastrophes, climate change, and attribution”, *Annals of the New York Academy of Sciences* **1469**, 105–124 (2020) (p. 2).
- <sup>6</sup>E. A. Lloyd, N. Oreskes, S. I. Seneviratne, and E. J. Larson, “Climate scientists set the bar of proof too high”, *Climatic Change* **165**, 55 (2021) (p. 2).
- <sup>7</sup>E. A. Lloyd and T. G. Shepherd, “Climate change attribution and legal contexts: evidence and the role of storylines”, *Climatic Change* **167**, 28 (2021) (p. 2).
- <sup>8</sup>R. F. Stuart-Smith, F. E. L. Otto, A. I. Saad, G. Lisi, P. Minnerop, K. C. Lauta, K. van Zwieten, and T. Wetzer, “Filling the evidentiary gap in climate litigation”, *Nature Climate Change* **11**, 1–5 (2021) (p. 2).
- <sup>9</sup>R. F. Stuart-Smith, G. H. Roe, S. Li, and M. R. Allen, “Increased outburst flood hazard from Lake Palcacocha due to human-induced glacier retreat”, *Nature Geoscience* **14**, 85–90 (2021) (p. 2).
- <sup>10</sup>S. Marjanac, L. Patton, and J. Thornton, “Acts of God, human influence and litigation”, *Nature Geoscience* **10**, 616–619 (2017) (p. 2).
- <sup>11</sup>S. Marjanac and L. Patton, “Extreme weather event attribution science and climate change litigation: an essential step in the causal chain?”, *Journal of Energy & Natural Resources Law* **36**, 265–298 (2018) (p. 2).
- <sup>12</sup>M. Burger, J. Wentz, and R. Horton, “The Law and Science of Climate Change Attribution”, *Columbia Journal of Environmental Law* **45** (2020) [10.7916/cjel.v45i1.4730](https://doi.org/10.7916/cjel.v45i1.4730) (p. 2).

**Extensive review of both the scientific understanding of climate change attribution as it relates to legal contexts, and the legal and policy applications of such and future research with case examples.**

- <sup>13</sup>B. J. Clarke, F. E. L. Otto, and R. G. Jones, “Inventories of extreme weather events and impacts: Implications for loss and damage from and adaptation to climate extremes”, *Climate Risk Management* **32**, 100285 (2021) (p. 2).
- <sup>14</sup>M. F. Wehner and K. A. Reed, “Operational extreme weather event attribution can quantify climate change loss and damages”, *PLOS Climate* **1**, e0000013 (2022) (p. 2).
- <sup>15</sup>J. Ettinger, P. Walton, J. Painter, S. Osaka, and F. E. L. Otto, ““What’s Up with the Weather?” Public Engagement with Extreme Event Attribution in the United Kingdom”, *Weather, Climate, and Society* **13**, 341–352 (2021) (p. 2).
- <sup>16</sup>S. Osaka and R. Bellamy, “Natural variability or climate change? Stakeholder and citizen perceptions of extreme event attribution”, *Global Environmental Change* **62**, 102070 (2020) (p. 2).
- <sup>17</sup>D. L. Swain, D. Singh, D. Touma, and N. S. Diffenbaugh, “Attributing Extreme Events to Climate Change: a New Frontier in a Warming World”, *One Earth* **2**, 522–527 (2020) (p. 2).
- <sup>18</sup>J. Sillmann, T. Thorarinsdottir, N. Keenlyside, N. Schaller, L. V. Alexander, G. Hegerl, S. I. Seneviratne, R. Vautard, X. Zhang, and F. W. Zwiers, “Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities”, *Weather and Climate Extremes* **18**, 65–74 (2017) (pp. 2, 49).
- <sup>19</sup>P. A. Stott, D. A. Stone, and M. R. Allen, “Human contribution to the European heatwave of 2003”, *Nature* **432**, 610–614 (2004) (pp. 2, 10, 29, 49, 55, 58, 59).  
**The first example of a probabilistic extreme event attribution study. Applies optimal fingerprinting to all- and natural-only-forcing simulations to produce a conservative estimate of the change in probability of the 2003 heatwave due to human influence on the climate.**
- <sup>20</sup>M. Hoerling, A. Kumar, R. Dole, J. W. Nielsen-Gammon, J. Eischeid, J. Perlitz, X. W. Quan, T. Zhang, P. Pegion, and M. Chen, “Anatomy of an extreme event”, *Journal of Climate* **26**, 2811–2832 (2013) (pp. 2, 46).  
**The first example of a storyline attribution study. The authors investigate the natural and anthropogenic contributions to the 2011 Texas heatwave and associated drought. Of additional interest in the context of this thesis is their exploration of the predictability of the extreme using two forecasting systems at different CO<sub>2</sub> concentrations.**
- <sup>21</sup>C. Johnston, *Heatwave temperatures may top 45C in southern Europe*, 2018 (p. 7).
- <sup>22</sup>NESDIS, *Record Summer Heat Bakes Europe*, 2018 (p. 7).
- <sup>23</sup>F. Krikken, F. Lehner, K. Haustein, I. Drobyshev, and G. J. van Oldenborgh, “Attribution of the role of climate change in the forest fires in Sweden 2018”, *Natural Hazards and Earth System Sciences* **21**, 2169–2179 (2021) (p. 7).
- <sup>24</sup>J. Watts, *Wildfires rage in Arctic Circle as Sweden calls for help*, 2018 (p. 7).
- <sup>25</sup>Publico, *Nueve fallecidos por la ola de calor en España*, 2018 (p. 7).
- <sup>26</sup>C. Harris, *Heat, hardship and horrible harvests: Europe’s drought explained*, 2018 (p. 7).

- <sup>27</sup>World Weather Attribution, *Heatwave in northern Europe, summer 2018*, 2018 (p. 7).
- <sup>28</sup>Press Office, *Chance of summer heatwaves now thirty times more likely*, 2018 (p. 7).
- <sup>29</sup>M. McCarthy, N. Christidis, N. Dunstone, D. Fereday, G. Kay, A. Klein-Tank, J. Lowe, J. Petch, A. Scaife, and P. Stott, “Drivers of the UK summer heatwave of 2018”, *Weather, wea.3628 (2019)* (p. 7).
- <sup>30</sup>J. H. Christensen and O. B. Christensen, “A summary of the PRUDENCE model projections of changes in European climate by the end of this century”, *Climatic Change 81*, 7–30 (2007) (pp. 7, 47).
- <sup>31</sup>D. D’Ippoliti, P. Michelozzi, C. Marino, F. De’Donato, B. Menne, K. Katsouyanni, U. Kirchmayer, A. Analitis, M. Medina-Ramón, A. Paldy, R. Atkinson, S. Kovats, L. Bisanti, A. Schneider, A. Lefranc, C. Iñiguez, and C. A. Perucci, “The impact of heat waves on mortality in 9 European cities: Results from the EuroHEAT project”, *Environmental Health: A Global Access Science Source 9* (2010) 10.1186/1476-069X-9-37 (p. 7).
- <sup>32</sup>J. Cattiaux, A. Ribes, J. Cattiaux, and A. Ribes, “Defining Single Extreme Weather Events in a Climate Perspective”, *Bulletin of the American Meteorological Society 99*, 1557–1568 (2018) (p. 8).
- <sup>33</sup>N. Christidis, P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J. Tennant, “A New HadGEM3-a-Based System for Attribution of Weather- and Climate-Related Extreme Events”, *Journal of Climate 26*, 2756–2783 (2013) (pp. 8, 49).
- <sup>34</sup>A. Ciavarella, N. Christidis, M. Andrews, M. Groenendijk, J. Rostron, M. Elkington, C. Burke, F. C. Lott, and P. A. Stott, “Upgrade of the HadGEM3-a based attribution system to high resolution and a new validation framework for probabilistic event attribution”, *Weather and Climate Extremes 20*, 9–32 (2018) (pp. 8, 49, 50, 58).
- <sup>35</sup>R. Vautard, A. Gobiet, D. Jacob, M. Belda, A. Colette, M. Déqué, J. Fernández, M. García-Díez, K. Goergen, I. Güttler, T. Halenka, T. Karacostas, E. Katragkou, K. Keuler, S. Kotlarski, S. Mayer, E. van Meijgaard, G. Nikulin, M. Patarčić, J. Scinocca, S. Sobolowski, M. Suklitsch, C. Teichmann, K. Warrach-Sagi, V. Wulfmeyer, and P. Yiou, “The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project”, *Climate Dynamics 41*, 2555–2575 (2013) (p. 8).
- <sup>36</sup>D. Jacob, J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, E. Georgopoulou, A. Gobiet, L. Menut, G. Nikulin, A. Haensler, N. Hempelmann, C. Jones, K. Keuler, S. Kovats, N. Kröner, S. Kotlarski, A. Kriegsmann, E. Martin, E. van Meijgaard, C. Moseley, S. Pfeifer, S. Preuschmann, C. Radermacher, K. Radtke, D. Rechid, M. Rounsevell, P. Samuelsson, S. Somot, J.-F. Soussana, C. Teichmann, R. Valentini, R. Vautard, B. Weber, and P. Yiou, “EURO-CORDEX: new high-resolution climate change projections for European impact research”, *Regional Environmental Change 14*, 563–578 (2014) (p. 8).
- <sup>37</sup>M. Vrac, P. Vaittinada Ayar, M. Vrac, and P. V. Ayar, “Influence of Bias Correcting Predictors on Statistical Downscaling Models”, *Journal of Applied Meteorology and Climatology 56*, 5–26 (2017) (p. 8).

- <sup>38</sup>E. E. Aalbers, G. Lenderink, E. van Meijgaard, and B. J. J. M. van den Hurk, “Local-scale changes in mean and heavy precipitation in Western Europe, climate change or internal variability?”, *Climate Dynamics* **50**, 4745–4766 (2018) (p. 8).
- <sup>39</sup>G. Lenderink, B. J. J. M. van den Hurk, A. M. G. Klein Tank, G. J. van Oldenborgh, E. van Meijgaard, H. de Vries, and J. J. Beersma, “Preparing local climate change scenarios for the Netherlands using resampling of climate model output”, *Environmental Research Letters* **9**, 115008 (2014) (p. 8).
- <sup>40</sup>N. Massey, R. Jones, F. E. L. Otto, T. Aina, S. Wilson, J. M. Murphy, D. Hassell, Y. H. Yamazaki, and M. R. Allen, “Weather@home-development and validation of a very large ensemble modelling system for probabilistic event attribution”, *Quarterly Journal of the Royal Meteorological Society* **141**, 1528–1545 (2015) (pp. 8, 49).
- <sup>41</sup>K. Haustein, M. R. Allen, P. M. Forster, F. E. L. Otto, D. M. Mitchell, H. D. Matthews, and D. J. Frame, “A real-time Global Warming Index”, *Scientific Reports* **7**, 15417 (2017) (p. 9).
- <sup>42</sup>C. P. Morice, J. J. Kennedy, N. A. Rayner, J. P. Winn, E. Hogan, R. E. Killick, R. J. Dunn, T. J. Osborn, P. D. Jones, and I. R. Simpson, “An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set”, *Journal of Geophysical Research: Atmospheres* **126** (2021) 10.1029/2019JD032361 (p. 9).
- <sup>43</sup>N. S. Diffenbaugh, D. Singh, J. S. Mankin, D. E. Horton, D. L. Swain, D. Touma, A. Charland, Y. Liu, M. Haugen, M. Tsiang, and B. Rajaratnam, “Quantifying the influence of global warming on unprecedented extreme climate events”, *Proceedings of the National Academy of Sciences* **114**, 4881–4886 (2017) (pp. 9, 19).
- <sup>44</sup>S. Jeon, C. J. Paciorek, and M. F. Wehner, “Quantile-based bias correction and uncertainty quantification of extreme event attribution statements”, *Weather and Climate Extremes* **12**, 24–32 (2016) (pp. 9, 59).
- <sup>45</sup>G. J. Van Oldenborgh, A. Van Urk, and M. Allen, “The absence of a role of climate change in the 2011 Thailand floods”, *Bulletin of the American Meteorological Society* **93**, 1047–1049 (2012) (p. 10).
- <sup>46</sup>J. R. M. Hosking, “L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics”, *Journal of the Royal Statistical Society: Series B (Methodological)* **52**, 105–124 (1990) (pp. 10, 27).
- <sup>47</sup>O. Bellprat, V. Guemas, F. Doblas-Reyes, and M. G. Donat, “Towards reliable extreme weather and climate event attribution”, *Nature Communications* **10**, 1732 (2019) (pp. 13, 15, 45, 49).
- <sup>48</sup>E. M. Fischer, U. Beyerle, C. F. Schleussner, A. D. King, and R. Knutti, “Biased Estimates of Changes in Climate Extremes From Prescribed SST Simulations”, *Geophysical Research Letters* **45**, 8500–8509 (2018) (pp. 13, 40, 49, 58).
- <sup>49</sup>J. He and B. J. Soden, “Does the Lack of Coupling in SST-Forced Atmosphere-Only Models Limit Their Usefulness for Climate Change Studies?”, *Journal of Climate* **29**, 4317–4325 (2016) (pp. 13, 40).
- <sup>50</sup>M. C. Kirchmeier-Young, H. Wan, X. Zhang, and S. I. Seneviratne, “Importance of Framing for Extreme Event Attribution: The Role of Spatial and Temporal Scales”, *Earth’s Future* **7**, 1192–1204 (2019) (pp. 14, 47).

- <sup>51</sup>WEF, *The Global Risks Report 2021*, tech. rep. (World Economic Forum, 2021) (p. 19).
- <sup>52</sup>S. I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S. M. Vincente-Serrano, M. Wehner, and B. Zhou, “Weather and Climate Extreme Events in a Changing Climate”, in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, 2021) (p. 19).
- <sup>53</sup>M. R. Allen, D. J. Frame, C. Huntingford, C. D. Jones, J. A. Lowe, M. Meinshausen, and N. Meinshausen, “Warming caused by cumulative carbon emissions towards the trillionth tonne”, *Nature* **458**, 1163–1166 (2009) (p. 19).
- <sup>54</sup>S. Rahmstorf and D. Coumou, “Increase of extreme events in a warming world”, *Proceedings of the National Academy of Sciences of the United States of America* **108**, 17905–17909 (2011) (p. 19).
- <sup>55</sup>J. A. Lowe, D. Bernie, P. Bett, L. Bricheno, S. Brown, D. Calvert, R. Clark, K. Eagle, T. Edwards, G. Fosser, F. Fung, L. Gohar, P. Good, J. Gregory, G. Harris, T. Howard, N. Kaye, E. Kendon, J. Krijnen, P. Maisey, R. McDonald, R. McInnes, C. McSweeney, J. F. B. Mitchell, J. Murphy, M. Palmer, C. Roberts, J. Rostron, D. Sexton, H. Thornton, J. Tinker, S. Tucker, K. Yamazaki, and S. Belcher, *UKCP18 Science Overview Report*, tech. rep. (Met Office Hadley Centre, 2018) (p. 19).
- <sup>56</sup>J. Murphy, G. Harris, D. Sexton, E. Kendon, P. Bett, R. Clark, K. Eagle, G. Fosser, F. Fung, J. Lowe, R. McDonald, R. McInnes, C. McSweeney, J. Mitchell, J. Rostron, H. Thornton, S. Tucker, and K. Yamazaki, *UKCP18 Land Projections: Science Report*, tech. rep. (Met Office, Exeter, 2018) (pp. 19, 21, 22, 28, 41).
- <sup>57</sup>T. Kelder, M. Müller, L. J. Slater, T. I. Marjoribanks, R. L. Wilby, C. Prudhomme, P. Bohlinger, L. Ferranti, and T. Nipen, “Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes”, *npj Climate and Atmospheric Science* **3**, 47 (2020) (p. 20).
- <sup>58</sup>V. Thompson, N. J. Dunstone, A. A. Scaife, D. M. Smith, J. M. Slingo, S. Brown, and S. E. Belcher, “High risk of unprecedented UK rainfall in the current climate”, *Nature Communications* **8**, 1–6 (2017) (p. 20).
- <sup>59</sup>E. M. Fischer, S. Sippel, and R. Knutti, “Increasing probability of record-shattering climate extremes”, *Nature Climate Change*, 1–7 (2021) (p. 21).
- <sup>60</sup>C. Gessner, E. M. Fischer, U. Beyerle, and R. Knutti, “Very Rare Heat Extremes: Quantifying and Understanding Using Ensemble Reinitialization”, *Journal of Climate* **34**, 6619–6634 (2021) (pp. 21, 37).
- <sup>61</sup>K. Riahi, S. Rao, V. Krey, C. Cho, V. Chirkov, G. Fischer, G. Kindermann, N. Nakicenovic, and P. Rafaj, “RCP 8.5—a scenario of comparatively high greenhouse gas emissions”, *Climatic Change* **109**, 33–57 (2011) (p. 21).

- <sup>62</sup>K. D. Williams, D. Copsey, E. W. Blockley, A. Bodas-Salcedo, D. Calvert, R. Comer, P. Davis, T. Graham, H. T. Hewitt, R. Hill, P. Hyder, S. Ineson, T. C. Johns, A. B. Keen, R. W. Lee, A. Megann, S. F. Milton, J. G. L. Rae, M. J. Roberts, A. A. Scaife, R. Schiemann, D. Storkey, L. Thorpe, I. G. Watterson, D. N. Walters, A. West, R. A. Wood, T. Woollings, and P. K. Xavier, “The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations”, *Journal of Advances in Modeling Earth Systems* **10**, 357–380 (2018) (p. 21).
- <sup>63</sup>A. V. Karmalkar, D. M. H. Sexton, J. M. Murphy, B. B. B. Booth, J. W. Rostron, and D. J. McNeall, “Finding plausible and diverse variants of a climate model. Part II: development and validation of methodology”, *Climate Dynamics* **53**, 847–877 (2019) (p. 21).
- <sup>64</sup>D. M. H. Sexton, A. V. Karmalkar, J. M. Murphy, K. D. Williams, I. A. Boutle, C. J. Morcrette, A. J. Stirling, and S. B. Vosper, “Finding plausible and diverse variants of a climate model. Part 1: establishing the relationship between errors at weather and climate time scales”, *Climate Dynamics* **53**, 989–1022 (2019) (p. 21).
- <sup>65</sup>D. M. H. Sexton, C. F. McSweeney, J. W. Rostron, K. Yamazaki, B. B. B. Booth, J. M. Murphy, L. Regayre, J. S. Johnson, and A. V. Karmalkar, “A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations”, *Climate Dynamics* **56**, 3395–3436 (2021) (pp. 21, 24).
- <sup>66</sup>K. Yamazaki, D. M. H. Sexton, J. W. Rostron, C. F. McSweeney, J. M. Murphy, and G. R. Harris, “A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 2: global performance and future changes”, *Climate Dynamics* **56**, 3437–3471 (2021) (p. 21).
- <sup>67</sup>D. Sexton, K. Yamazaki, J. Murphy, and J. Rostron, *Assessment of drifts and internal variability in UKCP projections*, tech. rep. (Met Office, 2020), p. 20 (p. 21).
- <sup>68</sup>M. Webb, C. Senior, S. Bony, and J.-J. Morcrette, “Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models”, *Climate Dynamics* **17**, 905–922 (2001) (p. 21).
- <sup>69</sup>K. D. Williams, M. A. Ringer, and C. A. Senior, “Evaluating the cloud response to climate change and current climate variability”, *Climate Dynamics* **20**, 705–721 (2003) (p. 21).
- <sup>70</sup>V. D. Pope, M. L. Gallani, P. R. Rountree, and R. A. Stratton, “The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3”, *Climate Dynamics* **16**, 123–146 (2000) (p. 21).
- <sup>71</sup>E. Bevacqua, T. G. Shepherd, P. A. G. Watson, S. Sparrow, D. Wallom, and D. Mitchell, “Larger Spatial Footprint of Wintertime Total Precipitation Extremes in a Warmer Climate”, *Geophysical Research Letters* **48**, e2020GL091990 (2021) (pp. 21, 28).
- <sup>72</sup>P. Watson, S. Sparrow, W. Ingram, S. Wilson, D. Marie, G. Zappa, R. Jones, D. Mitchell, T. Woollings, and M. Allen, *Multi-thousand member ensemble atmospheric simulations with global 60km resolution using climateprediction.net*, tech. rep. EGU2020-10895 (Copernicus Meetings, 2020) (pp. 21, 28, 41).
- <sup>73</sup>M. Allen, “Do-it-yourself climate prediction”, *Nature* **401**, 642–642 (1999) (p. 22).
- <sup>74</sup>D. Anderson, “BOINC: a system for public-resource computing and storage”, in *Fifth IEEE/ACM International Workshop on Grid Computing* (2004), pp. 4–10 (p. 22).

- <sup>75</sup>D. Stainforth, J. Kettleborough, M. Allen, M. Collins, A. Heaps, and J. Murphy, “Distributed computing for public-interest climate modeling research”, *Computing in Science Engineering* **4**, 82–89 (2002) (p. 22).
- <sup>76</sup>A. Brown, S. Milton, M. Cullen, B. Golding, J. Mitchell, and A. Shelly, “Unified Modeling and Prediction of Weather and Climate: a 25-Year Journey”, *Bulletin of the American Meteorological Society* **93**, 1865–1877 (2012) (p. 22).
- <sup>77</sup>D. A. Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, and M. R. Allen, “Uncertainty in predictions of the climate response to rising levels of greenhouse gases”, *Nature* **433**, 403–406 (2005) (p. 22).
- <sup>78</sup>D. Frame, T. Aina, C. Christensen, N. Faull, S. Knight, C. Piani, S. Rosier, K. Yamazaki, Y. Yamazaki, and M. Allen, “The climateprediction.net BBC climate change experiment: design of the coupled model ensemble”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**, 855–870 (2009) (p. 22).
- <sup>79</sup>P. Pall, T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, and M. R. Allen, “Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000”, *Nature* **470**, 382–385 (2011) (pp. 22, 49).
- <sup>80</sup>E. Kendon, G. Fosser, J. Murphy, S. Chan, R. Clark, G. Harris, A. Lock, J. Lowe, G. Martin, J. Pirret, N. Roberts, M. Sanderson, and S. Tucker, *UKCP Convection-permitting model projections: Science report*, tech. rep. (Met Office Hadley Centre, Exeter, 2019) (p. 22).
- <sup>81</sup>S. Sparrow, D. Sexton, N. J. Leach, P. A. G. Watson, and D. C. H. Wallom, “ExSamples Simulation Dataset”, *Scientific Data* **in prep.** (2021) (p. 22).
- <sup>82</sup>D. M. H. Sexton, D. P. Rowell, C. K. Folland, and D. J. Karoly, “Detection of anthropogenic climate change using an atmospheric GCM”, *Climate Dynamics* **17**, 669–685 (2001) (p. 23).
- <sup>83</sup>R. Neal, D. Fereday, R. Crocker, and R. E. Comer, “A flexible approach to defining weather patterns and their application in weather forecasting over Europe”, *Meteorological Applications* **23**, 389–400 (2016) (pp. 25, 27).
- <sup>84</sup>M. Kendon, D. Sexton, and M. McCarthy, “A temperature of 20°C in the UK winter: a sign of the future?”, *Weather* **75**, 318–324 (2020) (pp. 25, 37, 47).
- <sup>85</sup>C. Deser, M. A. Alexander, S.-P. Xie, and A. S. Phillips, “Sea Surface Temperature Variability: Patterns and Mechanisms”, *Annual Review of Marine Science* **2**, 115–143 (2010) (p. 25).
- <sup>86</sup>W. T. K. Huang, A. Charlton-Perez, R. W. Lee, R. Neal, C. Sarran, and T. Sun, “Weather regimes and patterns associated with temperature-related excess mortality in the UK: a pathway to sub-seasonal risk forecasting”, *Environmental Research Letters* **15**, 124052 (2020) (p. 25).
- <sup>87</sup>D. Richardson, R. Neal, R. Dankers, K. Mylne, R. Cowling, H. Clements, and J. Millard, “Linking weather patterns to regional extreme precipitation for highlighting potential flood events in medium- to long-range forecasts”, *Meteorological Applications* **27**, e1931 (2020) (p. 25).

- <sup>88</sup>D. Richardson, H. J. Fowler, C. G. Kilsby, and R. Neal, “A new precipitation and drought climatology based on weather patterns”, *International Journal of Climatology* **38**, 630–648 (2018) (p. 25).
- <sup>89</sup>C. Deser, I. R. Simpson, K. A. McKinnon, and A. S. Phillips, “The Northern Hemisphere Extratropical Atmospheric Circulation Response to ENSO: How Well Do We Know It and How Do We Evaluate Models Accordingly?”, *Journal of Climate* **30**, 5059–5082 (2017) (pp. 25, 33).
- <sup>90</sup>M. P. King, E. Yu, and J. Sillmann, “Impact of strong and extreme El Niños on European hydroclimate”, *Tellus A: Dynamic Meteorology and Oceanography* **72**, 1–10 (2020) (p. 25).
- <sup>91</sup>M. P. King, I. Herceg-Bulić, I. Bladé, J. García-Serrano, N. Keenlyside, F. Kucharski, C. Li, and S. Sobolowski, “Importance of Late Fall ENSO Teleconnection in the Euro-Atlantic Sector”, *Bulletin of the American Meteorological Society* **99**, 1337–1343 (2018) (p. 25).
- <sup>92</sup>J. López-Parages, B. Rodríguez-Fonseca, D. Dommenget, and C. Frauen, “ENSO influence on the North Atlantic European climate: a non-linear and non-stationary approach”, *Climate Dynamics* **47**, 2071–2084 (2016) (p. 25).
- <sup>93</sup>J. A. Francis and S. J. Vavrus, “Evidence linking Arctic amplification to extreme weather in mid-latitudes”, *Geophysical Research Letters* **39** (2012) [10.1029/2012GL051000](https://doi.org/10.1029/2012GL051000) (p. 25).
- <sup>94</sup>R. A. Pedersen, I. Cvijanovic, P. L. Langen, and B. M. Vinther, “The Impact of Regional Arctic Sea Ice Loss on Atmospheric Circulation and the NAO”, *Journal of Climate* **29**, 889–902 (2016) (p. 25).
- <sup>95</sup>M. Kretschmer, G. Zappa, and T. G. Shepherd, “The role of Barents–Kara sea ice loss in projected polar vortex changes”, *Weather and Climate Dynamics* **1**, 715–730 (2020) (p. 25).
- <sup>96</sup>J. A. Screen and I. Simmonds, “Exploring links between Arctic amplification and mid-latitude weather”, *Geophysical Research Letters* **40**, 959–964 (2013) (p. 25).
- <sup>97</sup>J. A. Screen, “The missing Northern European winter cooling response to Arctic sea ice loss”, *Nature Communications* **8**, 14603 (2017) (p. 25).
- <sup>98</sup>J. R. Hosking, J. R. Wallis, and E. F. Wood, “Estimation of the generalized extreme-value distribution by the method of probability-weighted moments”, *Technometrics* **27**, 251–261 (1985) (p. 27).
- <sup>99</sup>J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis* (Cambridge University Press, Cambridge, 1997) (p. 27).
- <sup>100</sup>S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics (Springer-Verlag, London, 2001) (pp. 27, 37).
- <sup>101</sup>J. R. M. Hosking and J. R. Wallis, “Parameter and Quantile Estimation for the Generalized Pareto Distribution”, *Technometrics* **29**, 339–349 (1987) (p. 27).
- <sup>102</sup>J. Kysely and J. Picek, “Probability estimates of heavy precipitation events in a flood-prone central-European region with enhanced influence of Mediterranean cyclones”, in *Advances in Geosciences*, Vol. 12 (2007), pp. 43–50 (p. 27).

- <sup>103</sup>W. Z. Wan Zin, A. A. Jemain, and K. Ibrahim, “The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of I-moment and LQ-moment”, *Theoretical and Applied Climatology* **96**, 337–344 (2009) (p. 27).
- <sup>104</sup>M. Marani and M. Ignaccolo, “A metastatistical approach to rainfall extremes”, *Advances in Water Resources* **79**, 121–126 (2015) (p. 27).
- <sup>105</sup>J. Cattiaux, R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron, “Winter 2010 in Europe: a cold extreme in a warming climate”, *Geophysical Research Letters* **37** (2010) [10.1029/2010GL044613](https://doi.org/10.1029/2010GL044613) (p. 27).
- <sup>106</sup>R. Vautard, P. Yiou, F. Otto, P. Stott, N. Christidis, G. J. van Oldenborgh, and N. Schaller, “Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events”, *Environmental Research Letters* **11**, 114009 (2016) (p. 27).
- <sup>107</sup>P. Yiou, A. Jézéquel, P. Naveau, F. E. L. Otto, R. Vautard, and M. Vrac, “A statistical framework for conditional extreme event attribution”, *Advances in Statistical Climatology, Meteorology and Oceanography* **3**, 17–31 (2017) (pp. 27, 49, 58).
- <sup>108</sup>J. L. Hodges, “The significance probability of the smirnov two-sample test”, *Arkiv för Matematik* **3**, 469–486 (1958) (p. 28).
- <sup>109</sup>A. N. Kolmogorov, “Sulla Determinazione Empirica di Una Legge di Distribuzione”, *Giornale dell’Istituto Italiano degli Attuari* **4**, 83–91 (1933) (p. 28).
- <sup>110</sup>N. Smirnoff, “On the estimation of the discrepancy between empirical curves of distribution for two independent samples”, *Bulletin Mathématique de L’Université de Moscow* **2**, 3–11 (1939) (p. 28).
- <sup>111</sup>N. Smirnoff, “Sur les écarts de la courbe de distribution empirique”, *Matematicheskii Sbornik* **6(48)**, 3–26 (1939) (p. 28).
- <sup>112</sup>D. A. Stone and M. R. Allen, “The end-to-end attribution problem: From emissions to impacts”, *Climatic Change* **71**, 303–318 (2005) (pp. 29, 55).
- <sup>113</sup>S. Brönnimann, “Impact of El Niño–Southern Oscillation on European climate”, *Reviews of Geophysics* **45** (2007) [10.1029/2006RG000199](https://doi.org/10.1029/2006RG000199) (p. 33).
- <sup>114</sup>D. King, D. Schrag, Z. Dadi, Q. Ye, and A. Ghosh, *Climate change: a risk assessment*, tech. rep. (UK Foreign & Commonwealth Office, London, 2015) (p. 37).
- <sup>115</sup>J. A. Lowe, T. Howard, A. Pardaens, J. Tinker, J. Holt, S. Wakelin, G. Milne, J. Leake, J. Wolf, K. Horsburgh, T. Reeder, G. Jenkins, J. Ridley, S. Dye, and S. Bradley, *UK Climate Projections science report: Marine and coastal projections*, tech. rep. (Met Office Hadley Centre, Exeter, UK, 2009) (p. 37).
- <sup>116</sup>S. Wade, M. Sanderson, N. Golding, J. A. Lowe, R. A. Betts, N. Reynard, A. L. Kay, L. Stewart, C. Prudhomme, L. Shaffrey, B. Lloyd-Hughes, and B. Harvey, *Developing h++ climate change scenarios for heat waves, droughts, floods, windstorms and cold snaps*, tech. rep. (Committee on Climate Change, London, 2015), p. 144 (pp. 37, 40).
- <sup>117</sup>S. J. Brown, J. M. Murphy, D. M. H. Sexton, and G. R. Harris, “Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases”, *Climate Dynamics* **43**, 2681–2705 (2014) (p. 37).

- <sup>118</sup>S. Sippel, D. Mitchell, M. T. Black, A. J. Dittus, L. Harrington, N. Schaller, and F. E. Otto, “Combining large model ensembles with extreme value statistics to improve attribution statements of rare events”, *Weather and Climate Extremes* **9**, 25–35 (2015) (p. 37).
- <sup>119</sup>M. Young and J. Galvin, “The record-breaking warm spell of February 2019 in Britain, the Channel Islands, France and the Netherlands”, *Weather* **75**, 36–45 (2020) (pp. 37, 47).
- <sup>120</sup>J. J. Barsugli and D. S. Battisti, “The Basic Effects of Atmosphere–Ocean Thermal Coupling on Midlatitude Variability”, *Journal of the Atmospheric Sciences* **55**, 477–493 (1998) (p. 40).
- <sup>121</sup>B. Dong, R. T. Sutton, L. Shaffrey, and N. P. Klingaman, “Attribution of Forced Decadal Climate Change in Coupled and Uncoupled Ocean–Atmosphere Model Experiments”, *Journal of Climate* **30**, 6203–6223 (2017) (p. 40).
- <sup>122</sup>C. source id values, *CMIP6 source id values*, 2022 (p. 40).
- <sup>123</sup>V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization”, *Geoscientific Model Development* **9**, 1937–1958 (2016) (pp. 40, 53).
- <sup>124</sup>R. Charlton, R. Fealy, S. Moore, J. Sweeney, and C. Murphy, “Assessing the Impact of Climate Change on Water Supply and Flood Hazard in Ireland Using Statistical Downscaling and Hydrological Modelling Techniques”, *Climatic Change* **74**, 475–491 (2006) (p. 40).
- <sup>125</sup>T. C. Peterson, P. A. Stott, and S. Herring, “Explaining Extreme Events of 2011 from a Climate Perspective”, *Bulletin of the American Meteorological Society* **93**, 1041–1067 (2012) (p. 45).
- <sup>126</sup>M. Hulme, “Attributing weather extremes to ‘climate change’: a review”, *Progress in Physical Geography* **38**, 499–511 (2014) (p. 45).
- <sup>127</sup>S. J. Hassol, S. Torok, S. Lewis, and P. Luganda, “(Un)Natural Disasters: Communicating Linkages Between Extreme Events and Climate Change”, *WMO Bulletin* **65**, 2–9 (2016) (p. 45).
- <sup>128</sup>S. C. Herring, N. Christidis, A. Hoell, M. P. Hoerling, and P. A. Stott, “Explaining Extreme Events of 2019 from a Climate Perspective”, *Bulletin of the American Meteorological Society* **102**, S1–S116 (2021) (p. 45).
- <sup>129</sup>O. Bellprat and F. Doblas-Reyes, “Attribution of extreme weather and climate events overestimated by unreliable climate simulations”, *Geophysical Research Letters* **43**, 2158–2164 (2016) (pp. 45, 49).
- <sup>130</sup>T. N. Palmer and A. Weisheimer, “A SIMPLE PEDAGOGICAL MODEL LINKING INITIAL-VALUE RELIABILITY WITH TRUSTWORTHINESS IN THE FORCED CLIMATE RESPONSE”, *Bulletin of the American Meteorological Society* **99**, 605–614 (2018) (pp. 45, 49).  
**An important paper that lucidly explains how initial-value reliability relates to the trustworthiness of the forced response in climate models, with specific reference to extreme event attribution.**

- <sup>131</sup>T. G. Shepherd, "A Common Framework for Approaches to Extreme Event Attribution", *Curr Clim Change Rep* **2**, 28–38 (2016) (pp. 45, 49, 50, 52, 55, 56, 58, 59).
- <sup>132</sup>E. P. Meredith, V. A. Semenov, D. Maraun, W. Park, and A. V. Chernokulsky, "Crucial role of Black Sea warming in amplifying the 2012 Krymsk precipitation extreme", *Nature Geoscience* **8**, 615–619 (2015) (pp. 46, 49).
- <sup>133</sup>L. Van Garderen, F. Feser, and T. G. Shepherd, "A methodology for attributing the role of climate change in extreme events: a global spectrally nudged storyline", *Natural Hazards and Earth System Sciences* **21**, 171–186 (2021) (pp. 46, 49, 50).
- <sup>134</sup>P. Hope, E.-P. Lim, G. Wang, H. H. Hendon, and J. M. Arblaster, "Contributors to the Record High Temperatures Across Australia in Late Spring 2014", *Bulletin of the American Meteorological Society* **96**, S149–S153 (2015) (pp. 46, 59, 60).
- <sup>135</sup>P. Hope, G. Wang, E.-P. Lim, H. H. Hendon, and J. M. Arblaster, "What Caused the Record-Breaking Heat Across Australia in October 2015?", *Bulletin of the American Meteorological Society* **97**, S122–S126 (2016) (pp. 46, 60).
- <sup>136</sup>P. Hope, M. T. Black, E.-P. Lim, A. Dowdy, G. Wang, R. J. B. Fawcett, and A. S. Pepler, "On Determining the Impact of Increasing Atmospheric CO<sub>2</sub> on the Record Fire Weather in Eastern Australia in February 2017", *Bulletin of the American Meteorological Society* **100**, S111–S117 (2019) (pp. 46, 58, 60).
- <sup>137</sup>K. A. Reed, A. M. Stansfield, M. F. Wehner, and C. M. Zarzycki, "Forecasted attribution of the human influence on Hurricane Florence", *Science Advances* **6**, eaaw9253 (2020) (p. 46).
- <sup>138</sup>G. M. Lackmann, "Hurricane Sandy before 1900 and after 2100", *Bulletin of the American Meteorological Society* **96**, 547–560 (2015) (p. 46).
- <sup>139</sup>A. Hannart, A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz, and P. Tandeo, "DADA: data assimilation for the detection and attribution of weather and climate-related events", *Climatic Change* **136**, 155–174 (2016) (p. 46).
- <sup>140</sup>R. C. Cornes, G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones, "An Ensemble Version of the e-OBS Temperature and Precipitation Data Sets", *Journal of Geophysical Research: Atmospheres* **123**, 9391–9409 (2018) (p. 47).
- <sup>141</sup>P. M. Sousa, R. M. Trigo, D. Barriopedro, P. M. M. Soares, and J. A. Santos, "European temperature responses to blocking and ridge regional patterns", *Climate Dynamics* **50**, 457–477 (2018) (p. 47).
- <sup>142</sup>N. Christidis and P. A. Stott, "Extremely Warm Days in the United Kingdom in Winter 2018/19", *Bulletin of the American Meteorological Society* **102**, S39–S44 (2021) (pp. 47, 56).
- <sup>143</sup>N. J. Leach, S. Li, S. Sparrow, G. J. van Oldenborgh, F. C. Lott, A. Weisheimer, and M. R. Allen, "Anthropogenic Influence on the 2018 Summer Warm Spell in Europe: The Impact of Different Spatio-Temporal Scales", *Bulletin of the American Meteorological Society* **101**, S41–S46 (2020) (pp. 47, 49).
- <sup>144</sup>P. Uhe, F. E. L. Otto, K. Haustein, G. J. van Oldenborgh, A. D. King, D. C. H. Wallom, M. R. Allen, and H. Cullen, "Comparison of methods: Attributing the 2014 record European temperatures to human influences", *Geophysical Research Letters* **43**, 8685–8693 (2016) (p. 47).

- <sup>145</sup>S. Sparrow, Q. Su, F. Tian, S. Li, Y. Chen, W. Chen, F. Luo, N. Freychet, F. C. Lott, B. Dong, S. F. B. Tett, and D. Wallom, “Attributing human influence on the July 2017 Chinese heatwave: the influence of sea-surface temperatures”, *Environ. Res. Lett* **13**, 114004 (2018) (p. 49).
- <sup>146</sup>W. Hazeleger, B. J. J. M. Van Den Hurk, E. Min, G. J. Van Oldenborgh, A. C. Petersen, D. A. Stainforth, E. Vasileiadou, and L. A. Smith, “Tales of future weather”, *NATURE CLIMATE CHANGE* **5** (2015) 10 . 1038/NCLIMATE2450 (p. 50).
- <sup>147</sup>D. A. Stone and P. Pall, “Benchmark estimate of the effect of anthropogenic emissions on the ocean surface”, *International Journal of Climatology* **41**, 3010–3026 (2021) (pp. 52, 60).
- <sup>148</sup>M. Etminan, G. Myhre, E. J. Highwood, and K. P. Shine, “Radiative forcing of carbon dioxide, methane, and nitrous oxide: a significant revision of the methane radiative forcing”, *Geophysical Research Letters* **43**, 12, 614–12, 623 (2016) (p. 52).
- <sup>149</sup>H. S. Baker, R. J. Millar, D. J. Karoly, U. Beyerle, B. P. Guillod, D. Mitchell, H. Shiogama, S. Sparrow, T. Woollings, and M. R. Allen, “Higher CO<sub>2</sub> concentrations increase extreme event risk in a 1.5 °c world”, *Nature Climate Change* **8**, 604–608 (2018) (p. 52).
- <sup>150</sup>W. L. Gates, K. H. Cook, and M. E. Schlesinger, “Preliminary analysis of experiments on the climatic effects of increased CO<sub>2</sub> with an atmospheric general circulation model and a climatological ocean”, *Journal of Geophysical Research* **86**, 6385 (1981) (p. 52).
- <sup>151</sup>J. F. Mitchell, “The seasonal response of a general circulation model to changes in CO<sub>2</sub> and sea temperatures”, *Quarterly Journal of the Royal Meteorological Society* **109**, 113–152 (1983) (p. 52).
- <sup>152</sup>K. E. Taylor, R. J. Stouffer, and G. A. Meehl, “An Overview of CMIP5 and the Experiment Design”, *Bulletin of the American Meteorological Society* **93**, 485–498 (2012) (p. 53).
- <sup>153</sup>M. Rugenstein, J. Bloch-Johnson, J. Gregory, T. Andrews, T. Mauritsen, C. Li, T. L. Frölicher, D. Paynter, G. Danabasoglu, S. Yang, J. Dufresne, L. Cao, G. A. Schmidt, A. Abe-Ouchi, O. Geoffroy, and R. Knutti, “Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Models”, *Geophysical Research Letters* **47** (2020) 10 . 1029/2019GL083898 (p. 53).
- <sup>154</sup>C. M. Flynn and T. Mauritsen, “On the climate sensitivity and historical warming evolution in recent coupled model ensembles”, *Atmospheric Chemistry and Physics* **20**, 7829–7842 (2020) (p. 53).
- <sup>155</sup>T. Andrews, J. M. Gregory, and M. J. Webb, “The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models”, *Journal of Climate* **28**, 1630–1648 (2015) (p. 53).
- <sup>156</sup>D. M. Smith, J. A. Screen, C. Deser, J. Cohen, J. C. Fyfe, J. García-Serrano, T. Jung, V. Kattsov, D. Matei, R. Msadek, Y. Peings, M. Sigmond, J. Ukita, J.-H. Yoon, and X. Zhang, “The Polar Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: investigating the causes and consequences of polar amplification”, *Geoscientific Model Development* **12**, 1139–1164 (2019) (p. 53).

- <sup>157</sup>E. Winsberg, N. Oreskes, and E. Lloyd, “Severe weather event attribution: Why values won’t go away”, *Studies in History and Philosophy of Science Part A* **84**, 142–149 (2020) (pp. 55, 58, 59).
- <sup>158</sup>A. Jézéquel, V. Dépoues, H. Guillemot, M. Trolliet, J.-P. Vanderlinden, and P. Yiou, “Behind the veil of extreme event attribution”, *Climatic Change* **149**, 367–383 (2018) (pp. 58, 59).
- <sup>159</sup>S. Philip, S. Kew, G. J. van Oldenborgh, F. Otto, R. Vautard, K. van der Wiel, A. King, F. Lott, J. Arrighi, R. Singh, and M. van Aalst, “A protocol for probabilistic extreme event attribution analyses”, *Advances in Statistical Climatology, Meteorology and Oceanography* **6**, 177–203 (2020) (pp. 58, 59).
- <sup>160</sup>P. Pall, C. M. Patricola, M. F. Wehner, D. A. Stone, C. J. Paciorek, and W. D. Collins, “Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013”, *Weather and Climate Extremes* **17**, 1–6 (2017) (p. 58).
- <sup>161</sup>S. Sippel, F. E. Otto, M. Forkel, M. R. Allen, B. P. Guillod, M. Heimann, M. Reichstein, S. I. Seneviratne, K. Thonicke, and M. D. Mahecha, “A novel bias correction methodology for climate impact simulations”, *Earth System Dynamics* **7**, 71–88 (2016) (p. 59).
- <sup>162</sup>S. Li, D. E. Rupp, L. Hawkins, P. W. Mote, D. McNeall, S. N. Sparrow, D. C. H. Wallom, R. A. Betts, and J. J. Wettstein, “Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation”, *Geoscientific Model Development* **12**, 3017–3043 (2019) (p. 59).
- <sup>163</sup>T. G. Shepherd, E. Boyd, R. A. Calel, S. C. Chapman, S. Dessai, I. M. Dima-West, H. J. Fowler, R. James, D. Maraun, O. Martius, C. A. Senior, A. H. Sobel, D. A. Stainforth, S. F. Tett, K. E. Trenberth, B. J. van den Hurk, N. W. Watkins, R. L. Wilby, and D. A. Zenghelis, “Storylines: an alternative approach to representing uncertainty in physical aspects of climate change”, *Climatic Change* **151**, 555–571 (2018) (p. 59).