

Reliable 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

Anthropogenic greenhouse gas emissions are now well-understood to be causing widespread and often damaging changes to the climate. One of the many ways in which the climate is changing is through the characteristics of extreme weather. Given the severe consequences that these devastating events can have, understanding how human influence on the climate is affecting them is of considerable importance. This is the aim of the relatively new field of ‘extreme event attribution’. Although there now exist a number of established methods for attributing individual weather events to climate change, ranging from probabilistic approaches with large climate model ensembles to very conditioned storyline approaches, questions still remain over the reliability of these approaches, especially when considering the most unprecedented events. We need to have confidence in this understanding in order to plan adaptation measures effectively to mitigate changes in risk into the future. In this thesis, I show how the models that we use to forecast the weather could provide us with such reliable information about how humanity is affecting extreme weather — and in particular extreme heat. These models are state-of-the-art and can be shown to be unequivocally able to simulate the detailed physics of specific extreme weather through a successful prediction. Here I develop a perturbed initial and boundary condition approach within an operational ensemble prediction system that aims to produce forecasts of individual events as if they had occurred in a world without human influence on the climate. These ‘counterfactual’ forecasts can then be used to assess how not only the intensity, but also the probability of these events has changed. Although extreme weather attribution typically focuses on the past climate, the same approach can in theory be used to produce forecasts in possible warmer future worlds — thus providing vital information about how the most damaging weather in the present may be expected to change in the future. In the final chapter, I depart from attribution but continue with this theme of extreme weather projection, examining a novel approach to producing large model ensembles that explore the range of uncertainty in future extreme weather. This work contrasts the very specific nature of the attribution of single weather events, but both approaches to extreme weather projection together could provide complementary and useful information for adaptation planning.

Reliable 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

Anthropogenic greenhouse gas emissions are now well-understood to be causing widespread and often damaging changes to the climate. One of the many ways in which the climate is changing is through the characteristics of extreme weather. Given the severe consequences that these devastating events can have, understanding how human influence on the climate is affecting them is of considerable importance. This is the aim of the relatively new field of ‘extreme event attribution’. Although there now exist a number of established methods for attributing individual weather events to climate change, ranging from probabilistic approaches with large climate model ensembles to very conditioned storyline approaches, questions still remain over the reliability of these approaches, especially when considering the most unprecedented events. We need to have confidence in this understanding in order to plan adaptation measures effectively to mitigate changes in risk into the future. In this thesis, I show how the models that we use to forecast the weather could provide us with such reliable information about how humanity is affecting extreme weather — and in particular extreme heat. These models are state-of-the-art and can be shown to be unequivocally able to simulate the detailed physics of specific extreme weather through a successful prediction. Here I develop a perturbed initial and boundary condition approach within an operational ensemble prediction system that aims to produce forecasts of individual events as if they had occurred in a world without human influence on the climate. These ‘counterfactual’ forecasts can then be used to assess how not only the intensity, but also the probability of these events has changed. Although extreme weather attribution typically focuses on the past climate, the same approach can in theory be used to produce forecasts in possible warmer future worlds — thus providing vital information about how the most damaging weather in the present may be expected to change in the future. In the final chapter, I depart from attribution but continue with this theme of extreme weather projection, examining a novel approach to producing large model ensembles that explore the range of uncertainty in future extreme weather. This work contrasts the very specific nature of the attribution of single weather events, but both approaches to extreme weather projection together could provide complementary and useful information for adaptation planning.

Contents

List of Figures	xi
List of Abbreviations	xiii
1 Introduction	1
1.1 The problem of extreme event attribution	3
1.2 Motivating the question	4
1.3 Answering the question	7
1.3.1 Probabilistic attribution	7
1.3.2 Attribution through storylines	12
1.3.3 A forecast-based approach	15
1.4 Conceptualising different approaches	19
1.5 The meteorology of heatwaves	19
1.6 What to expect in this thesis	22
2 Conventional probabilistic attribution	25
2.1 Chapter open	27
2.2 Abstract	27
2.3 The 2018 heatwave in Europe	27
2.3.1 Defining the event	28
2.4 Materials & methods	29
2.5 Results	31
2.6 Discussion	36
2.7 Chapter close	37
3 Forecast-based attribution: perturbing the boundary conditions	39
3.1 Chapter open	41
3.2 Abstract	41
3.3 Introduction	42
3.4 The 2019 February heatwave in Europe	44
3.5 Materials & methods	46
3.6 Forecasts of the heatwave	46
3.7 Perturbed CO ₂ forecasts	50

3.8 Attributing the heatwave to diabatic CO ₂ heating	53
3.9 Discussion	57
3.10 Chapter close	60
4 Forecast-based attribution: perturbing the initial and boundary conditions	61
4.1 Chapter open	63
4.2 Abstract	63
4.3 Introduction	64
4.4 The Pacific Northwest heatwave	66
4.5 Methods	68
4.6 Forecast-based attribution	74
4.6.1 Results	78
4.7 Discussion	80
4.8 Chapter close	81
5 Attribution and projection	83
5.1 Chapter open	85
5.2 Abstract	85
5.3 Introduction	86
5.4 Study design & methods	88
5.4.1 Models	88
5.4.2 ExSamples experiment design	89
5.4.3 Statistical methods	95
5.5 Results	97
5.5.1 Comparison of HadAM4 and HadGEM3-GC3.05 baseline ensembles	97
5.5.2 Projections of future extremes	100
5.5.3 Sampling record-shattering subseasonal events	107
5.6 Discussion	109
5.7 Concluding remarks	111
5.8 Chapter close	112
6 Discussion	113
6.1 An overview of this thesis	114
6.2 This thesis in the context of previous work	116
6.3 Limitations	121
6.4 Future research directions	124
6.4.1 Addressing the rapid atmospheric adjustment	124
6.4.2 Expanding the scope of this work	127
6.4.3 Alternative applications	128
6.5 Concluding remarks	131

Appendices

A Resources	135
B Appendix 2	137
References	139

x

List of Figures

1.1 Comparing typical horizontal resolutions of weather and climate models	16
2.1 The 2018 heatwave in Europe: observed mean temperature anomalies over a range of timescales	28
2.2 Calculating the heatwave threshold in HadGEM3-A from observations	30
2.3 Factual and counterfactual PDFs of the 2018 heatwave defined over three temporal scales	32
2.4 Maps of the estimated change in probability of the 2018 heatwave due to anthropogenic influence on the climate	33
2.5 Estimated changes in probability of the 2018 heatwave defined using regional mean temperatures	34
2.6 Estimated changes in probability of the 2018 British Isles heatwave across a range of observations and model simulations	35
2.7 Historical power spectrum of summer daily mean temperatures over the British Isles across a range of observations and model simulations	36
3.1 The 2019 February heatwave in Europe: synoptic characteristics & historical context.	45
3.2 Medium- to extended-range forecasts of the heatwave.	49
3.3 Global temperature and synoptic-scale dynamical response to CO ₂ perturbations.	52
3.4 Attribution of the direct CO ₂ influence on the heatwave.	56
4.1 Features and forecasts of the Pacific Northwest heatwave.	67
4.2 The initial ocean state perturbation applied.	71
4.3 Validation of the bias correction applied to the SEAS5 seasonal forecast simulations.	73
4.4 Drivers of the PNW heatwave and their predictability in the 11-day lead forecast.	77
4.5 Return-time diagram of the PNW heatwave in the operational and counterfactual forecast ensembles.	79

5.1	UKCP PPE 2061–2080 deviations from forced response.	92
5.2	Synoptic characteristics of the study winters within the UKCP simulations.	94
5.3	PDFs of Euclidean MSLP Distance between ExSamples ensemble members and corresponding UKCP18 extreme winters.	96
5.4	Rainfall variability-explaining PC histograms and EOF patterns in ExSamples and UKCP baselines.	99
5.5	Comparing statistics of DJF mean of daily maximum temperatures averaged over the UK region for the HOT1 winter.	102
5.6	Comparing statistics of DJF mean of daily maximum temperatures averaged over the UK region for the HOT2 winter.	104
5.7	Comparing statistics of DJF mean precipitation averaged over the UK region for the WET winter.	106
5.8	Examining statistics of subseasonal extreme weather events. . .	108

List of Abbreviations

CDF	Cumulative distribution function
CI	Confidence interval
CO₂	Carbon dioxide
DJF	December–January–February (meteorological winter)
ECMWF	European Centre for Medium-range Weather Forecasts
EEA	Extreme event attribution
FAR	Fraction of attributable risk
GCM	General circulation model
GEV	Generalised extreme value
GMST	Global mean surface temperature
GSAT	Global surface air temperature
IFS	Integrated forecasting system
JJA	June–July–August (meteorological summer)
MSLP	Mean sea level pressure
PDF	Probability density function
PPE	Perturbed parameter ensemble
SIC	Sea ice concentration
SST	Sea surface temperature
UK	United Kingdom
UKCP	UK Climate Projections
UKMO	UK Met Office
WWA	World Weather Attribution

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.

Contents

1.1	The problem of extreme event attribution	3
1.2	Motivating the question	4
1.3	Answering the question	7
1.3.1	Probabilistic attribution	7
1.3.2	Attribution through storylines	12
1.3.3	A forecast-based approach	15
1.4	Conceptualising different approaches	19
1.5	The meteorology of heatwaves	19
1.6	What to expect in this thesis	22

1.1 The problem of extreme event attribution

The link between greenhouse gas emissions and anthropogenic climate change has been known about for over a century (1). In recent decades, significant advances have been made in understanding the human influence (or ‘fingerprint’) in multi-decadal trends in global and regional climate (2–4). A number of statistical advances developed around the millennium (5–9) have allowed so-called ‘detection and attribution’ of human influence on a very wide range of climate variables in the decades since, including global temperature (10–12), regional temperature (13), global precipitation (14), global soil moisture (15), ocean salinity (16), tropospheric thickness (17), and many others (3). However, such large-scale and long-term changes in mean climate are far detached from individual human perception (18). On the other hand, individuals are able to relate to extreme weather events, due to the immediate and severe impacts they can have. This interest in such damaging events led scientists in the early 2000s to consider whether it would be possible to detect and attribute human influence on an individual extreme weather event (19).

What do I mean by detection and attribution of human influence in the context of a single extreme event? Possibly the simplest way to pose this question in the aftermath of such an event would be to ask: ‘was this weather *caused* by climate change?’. However, this framing sets the standard of proof at an exceptionally high bar — if the event were at all possible without human influence in any sense, regardless of how unlikely it might be, then the answer would be ‘no’. Despite the prevalence of this framing in both scientific and non-scientific circles (20, 21), and some recent studies that claim to have achieved this standard of proof (22), it is not a particularly useful or informative question to ask, and may have contributed to the belief that attribution of individual weather events is not possible (23). A more relevant and answerable question might be: ‘has anthropogenic climate change made this event more likely?’. This was the question posed in 2003 by Myles Allen in the seminal commentary “Liability for climate change” (19), widely acknowledged as the first time the idea that individual extreme events could be attributed to external drivers such as human influence was proposed (24). This frequency-based question has been tackled in many studies since (25). An alternative question could be: ‘has anthropogenic climate change made this event more intense?’. Although Allen et al. (20) suggested that it makes little sense in the context of a nonlinear and chaotic weather system, many studies have now examined this magnitude-based question (26).

These two different framings of the same ultimate question have led to some apparent discrepancies at times (27). By their nature, extreme events nearly always have a significant contribution from the natural internal variability of the weather. This natural contribution often far exceeds any estimated human contributions to the event magnitude. However, these ‘small’ human-induced changes in magnitude can lead significant changes in event probability. In this way, an event can be accurately described as both ‘mainly natural in origin’ (26) and having ‘an approximate 80 % probability that it would not have occurred without climate warming’ (28). I, and many others (29–34) believe that both of these framings are useful. Although I will develop these ideas further in due course, one context in which understanding changes in both probability and magnitude can be important is adaptation to climate change. Changes in probability are important for societal resilience: *how strong do I need to make a flood barrier given an increase in the number of floods of a certain severity?*; while changes in magnitude are vital for preparing for the potentially very non-linear impacts of extreme weather events: *how high do I need to make a flood barrier to ensure that it doesn't fail given the severity of a 1-in-100 year flood has now increased?*. Of course, the picture is more nuanced than this simplistic view, but my underlying point that understanding both changes in probability and magnitude are important is now widely accepted. Hence, the question that this thesis is largely concerned with answering is: **‘how has human influence on the climate affected both the probability and intensity of specific extreme weather events?’**.

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?*

To begin, I return to “Liability for climate change” (19), now focusing on the motivation for extreme event attribution. The motivation behind extreme event attribution as proposed in “Liability for climate change” is compensation for damage to individuals caused by climate change or, as Allen puts it,

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

Allen suggests that in the future those affected by particular extreme weather may, given sufficient scientific certainty, be able to claim compensation from greenhouse gas emitters for damages caused by the extreme weather. He

proposes a framework, grounded in concepts from epidemiology (35), in which emitters pay for the ‘fraction’ of an extreme weather event that they caused, even in the absence of absolute causation. This fraction is estimated probabilistically based on the change in likelihood of the event in a world in which the emissions never happened (i.e. if the event is half as likely to occur without the emissions, then the fraction of the event that is attributable to the emitters is 50%). This specific application is therefore using extreme event attribution as evidence in environmental tort law. Since Allen (19), much has been written in relation to this application. Allen et al. (20) presents an overview of the state of climate change detection and attribution aimed at legal professionals, concluding with a set of related questions for the legal community. More recently, Stuart-Smith et al. (36) provided a set of suggestions for potential plaintiffs on how to best make use of the climate science available (noting that evidence used in previous cases ‘lags substantially behind the state of the art’). Coming from the other side of the coin, Marjanac et al. (37) provide suggestions for climate change scientists, emphasising that ‘clear and confident expression of science in a manner that can be applied by non-scientists, including lawyers’ is key. Elisabeth A. Lloyd has authored a number of studies exploring various issues including the different standards of proof in scientific and legal contexts (38); how different approaches to extreme event attribution can complement one another to provide the most useful picture of climate change impacts for a broad range of contexts (39, 40); and finally, examines a specific tort law case that made use of extreme event attribution (41). For thorough review of both the science and legal context of extreme event attribution, written from a legal perspective, with reference to specific cases, I recommend Burger et al. (42) and Marjanac and Patton (43).

In addition to the original application in tort law suggested in Allen (19), more recently it has been suggested that extreme event attribution could ‘play a significant role in quantifying loss and damage’ (44). Loss and damage is generally understood as the unavoidable adverse impacts of climate change (45), and has become a key piece of international climate change policy since the inclusion of Article 8 of the Paris Agreement. Several recent extreme event attribution based studies may have considerable influence in the future in this space, including Clarke et al. (46), who set out a framework for recording key details of high-impact weather events as a new source of evidence for global stocktakes on loss and damage; and Otto et al. (47) and Lott et al. (48), who adapt conventional extreme event attribution approaches to estimate the contributions of *specific* emitters to individual extreme weather events (as opposed to the more

usual broad ‘human influence’ considered). Perhaps we won’t have to wait much longer before the question posed by Allen is answered?

The other key non-scientific motivating factor for extreme event attribution is the public engagement and interest in the research (32). The ‘headline’ number in climate science has for a long time been change in global mean temperature (49–51). While this is clearly a very important number as the primary metric of the impact that humanity is having on the climate, it is not a number that individuals can easily relate to due to the large scales involved and indirect nature of the associated impacts. On the other hand, extreme weather events are phenomena that are actually experienced in real time by individuals — and regularly reported on by the media. Since extreme weather events can cause severe and direct socioeconomic impacts (52), increases in their frequency would likely be a considerably more relatable and concerning consequence of climate change than the distant change in global mean quantities. Previous work has shown that extreme event attribution may be an exceptionally useful tool for climate change communication (53), though can prove unhelpful if results are not clear and comprehensible for a general audience (for example if different attribution studies regarding a single event appear to provide conflicting headline results 54). A specific study investigating the experience-perception link of climate disasters in the context of Floridians that had experienced hurricane Irma found that this experience increased both their belief and concern in global warming (55), though a more recent study looking at connections between local weather and climate change awareness in Germany did not find a link (56).

There are also a number of more scientific motivating factors behind studying extreme event attribution. The first of these is the improved understanding of how climate change extreme weather events. By studying a single weather event in detail, we can gain a significant amount of knowledge about not only the processes that drove that event, but also about how those processes may have changed due to human influence on the climate. The nature of extreme events means that they are often driven by either unusual processes — or combinations of processes — that rarely occur. As such, they provide an opportunity to explore these rare drivers in detail and obtain a better understanding of their underlying physics. One example of this improved understanding has been the role of soil moisture on heatwaves (57–59).

On a related note, this same granular look at a single event allows us to examine the performance of the dynamical models we use by ensuring that the processes and physics of the modelled climate matches those same processes

in the real world (60, 61). This is especially true for weather forecast models, which aim to be able to precisely reproduce the real world. Even in the case of climate models, extreme event attribution studies have identified key issues in the simulations produced by climate models such as poor representation of climate variability (62, 63).

The final reason I shall mention here — though I discuss it further in the discussion section of this thesis 6 — is that how extreme events are changing in the present and will continue to change in the future is a question of vital importance for adaptation to climate change. Although the focus of attribution is typically understanding how extreme events have changed relative to a world without human influence on the climate, another side of a very similar coin is understanding how such influence will continue to change them into the future (64). This projection of future extremes will be vital for effective and targeted adaptation planning given the severe damage such events can cause. In addition to policymakers, a large proportion of the industrial sector, and especially the financial sector, need to know how climate change is affecting the risk from extreme weather on a continual basis in the present day. For example, insurance and reinsurance companies base their risk capacity on historical loss estimates — but in the case of many types of extreme weather event the baseline is shifting with climate change and historical data may not necessarily be representative of the present. This shifting baseline is something that extreme event attribution is well-placed to inform on.

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 describe and explore the ways in which previous studies have approached this question, in particular focusing on the probabilistic (often “conventional”, 25) and storyline (65) 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 by Allen (19) and first applied by Stott

et al. (25) to the 2003 European heatwave. They applied an optimal fingerprinting technique (5, 6) to transient climate model simulations. They used five simulations, one set of four with all forcings included, plus one with natural forcings only; generating scaling factors of the correspondence between the modelled response and observed changes by regressing each set onto observed central European summer temperature. The scaling factors could then be used to determine the 1990s temperature anomalies attributable to all forcings combined and natural forcings alone. A third control run at a fixed, pre-industrial level of forcing was used to estimate internal variability corresponding likelihood functions for these temperature anomalies. They finally used a peak-over-threshold extreme value analysis of the same control run to determine the probability of exceeding the temperature of the pre-2003 hottest summer in worlds with and without climate change by adjusting the mean summer temperature to the estimated 1990s temperature anomalies both with and without anthropogenic forcing. These probabilities (and associated uncertainty) could be used to determine the likelihood function of the change in risk of the heatwave attributable to human influence. This fairly involved statistical approach (in particular, the necessity for the use of a control run to estimate uncertainty due to internal variability) has been largely replaced by the use of much larger single- or multiple- model ensembles.

The next advance in probabilistic extreme event attribution came with Van Oldenborgh (66), who developed a methodology for estimating the change in risk of an extreme event using observations alone. Van Oldenborgh took observed timeseries of autumn temperatures measured by the De Bilt weather station, and removed the climate change signal by regression onto low-pass filtered global mean surface temperature (a reasonable proxy for anthropogenic influence on the climate, given the small contributions from natural forcing). Using this detrended series and extreme value analysis, he then computed the return period of the exceptional warm 2006 autumn in Europe, and compared it to the return period estimated using the original series. This method was later extended to use the return period to compute the return period of the extreme for both the present-day and pre-industrial period by shifting the detrended series by the attributable trend computed in the regression (eg. 61, 63). In this way, the change in risk between pre-industrial and present climates can be estimated. It is worth noting that this methodology does not formally attribute and changes in risk to anthropogenic influence, since trends in local climate may be influenced by other factors, and no use is made of a counterfactual world without human influence on the climate (since no observations of such a world exist). This method can also be applied

to transient simulations from climate models.

The final advance that I shall highlight is from Pall et al. (67), and was the first instance where specific fixed forcing (as opposed to transient) factual and counterfactual simulations were used. Pall et al. generated very large (2000+ member) atmosphere-only climate model ensembles of autumn 2000. One ensemble was driven using observed sea surface temperatures and sea ice, and corresponding atmospheric conditions (greenhouse gas, aerosol and ozone concentrations) for that time. The other four were driven using atmospheric conditions for the year 1900, and subtracting four estimates of attributable twentieth-century SST warming from the observed sea surface temperatures; the four estimates of attributable warming were derived from four different coupled climate model simulations. River runoff for England and Wales in the factual ensemble was compared to runoff in the four ‘naturalised’ counterfactual ensembles to determine the difference in risk of exceeding the value actually observed in autumn 2000 in the different climates. In this case, the ensembles are sufficiently large that extreme value analysis was not required (the – SST conditional – return period could be calculated by simply counting the number of members that exceeded the observed threshold and dividing by the total ensemble size in each case). This methodology is attractive due to the clear and straightforward statistical analysis of the large ensembles (with no reliance on extreme value analysis or optimal fingerprinting techniques). However, it does generally require very large ensembles of the time period when the event took place (i.e. autumn 2000 in this case); and is conditional on the prescribed SST pattern (requiring either that anthropogenic influence isn’t affecting interannual modes of SST variability, or that the extreme in question is independent of such modes).

The ‘standard’ approach to probabilistic attribution in the present draws upon each of these previous advances (61). Here I am taking the World Weather Attribution project (WWA) methodology as standard, since they are the most prolific group both in terms of number of events analysed and media coverage of their results (68), though other methodologies exist (69). Their approach involves:

1. Defining the event. Extreme events are, by the nature of the weather, exceptionally high dimensional and could be described in a practically unlimited number of ways (i.e. what variables to use? What spatial scale? What temporal scale?). However, to be able to analyse changes to such events, we must be able to define them quantitatively. The WWA attempts to select a metric that most closely corresponds to the impacts of the extreme

event in question, taking in account what questions are being asked by the various stakeholders. For example, if the key impact of interest is heatwave-associated mortality, then peak 3-day moving average daily maximum temperatures may be selected due to their close connection to health impacts (70). Once a metric has been chosen by which to define the extreme event, they use a ‘class-based’ framing considering all events of a similar magnitude, often by choosing the annual maximum values of the metric. This class-based framing results in a largely unconditional analysis, which I will discuss further below.

2. Analysing trends in observed data following Van Oldenborgh (66). This is typically done by fitting an appropriate distribution whose parameters shift or scale with low-pass filtered global mean surface temperature (GMST). The shifting or scaling depends upon the chosen metric and its observed or expected response to climate change. For example, the temperature based heatwave metrics that are the primary concern of this thesis are typically assumed to simply shift with global warming. From this GMST-covarying distribution, the return period is then computed for present-day and pre-industrial values of GMST. From these returns periods the change in risk of the observed extreme can be calculated. As with Van Oldenborgh (66), these changes in risk are not strictly attributable to human influence on the climate due to the lack of a no-human counterfactual.
3. Analysing simulations from as many models as possible. Transient model simulations are analysed in an identical way to observations. Fixed forcing simulations are analysed following Pall et al. (67). Only models which are able to closely represent the observed climate are considered – evaluated on the basis of the trends and distribution parameters implied by the model data.
4. Synthesising these various strands of evidence. The aim behind using as many lines of evidence as possible, combining statistical and numerical models, is to try and determine the most robust conclusion possible within the context of the associated uncertainties.

This ‘standard’ approach has been refined over the past decade by the WWA team (68), learning through application to a wide range of locations and types of extreme. The widespread recognition and understanding of extreme event attribution by the general public is due, in no small part, to this approach and

how rapidly the WWA team have been able to apply it to extreme events in the past few years. Their rapid response has meant that they are able to answer the questions people ask in the aftermath of such events when they are actually asking them – rather than following a lengthy peer-review process. However, this standard probabilistic approach to attribution is not without issues of its own – hence this thesis – which I shall now discuss.

The first issue arises due to the unconditional use of climate models. By their nature, extreme events are typically produced by exceptional physical processes, or combinations of processes. The type of models used in extreme event attribution are typically coarse ($O(100 \text{ km})$), and may well not be able to physically represent all the processes involved in the production of specific extreme events even if they can accurately represent the average climate (60, 71, 72). Such models still have serious known biases relevant to the simulation of extremes, including poor representation of Euro-Atlantic blocking (73, 74), which is a key synoptic driver of heatwaves over the continent. These biases become even more important when considering not only the models' ability to simulate the present climate, but also their response to external forcings such as anthropogenic climate change (75, 76). The use of biased models can lead to potentially incorrect quantitative attribution statements (62, 77).

The second key issue derives from the treatment of individual extremes as one of an event class. For example, in the conventional probabilistic approach to attributing a particular heatwave, one might consider all the previous annual maximum temperatures (e.g. in order to apply extreme value analysis as discussed above). However, the particular heatwave in question might have arisen from a very different – possibly unique in the context of the relatively short historical record – set of meteorological circumstances and physical drivers compared to all the previous heatwaves. This not only renders such extreme value analysis as is often performed potentially inappropriate (as the heatwave in question is drawn from a different underlying distribution to the others), but also any estimated climate change responses. What if the combination of the particular processes involved in producing the heatwave in question responds fundamentally differently to the processes that have generated past heatwaves?

The final issue I shall discuss is the one that has been most often commented on in previous work: the risk of type II errors (71, 78). This risk arises because some aspects of the climate system response to external forcing are much more certain and well-understood than others. For example, the thermodynamic aspects of climate change are broadly very well understood and certain: rising

greenhouse gas concentrations lead to a thicker troposphere, thus raising surface temperatures and increasing the moisture capacity of the troposphere. On the other hand, the dynamic aspects of climate change are considerably less certain, with models often disagreeing over the direction of changes in atmospheric dynamics (79). At this point, I note that this is not an entirely independent issue to the first issue discussed since much of this uncertainty arises due to model biases and imperfections. Since extreme events are often driven by a combination of both thermodynamic and dynamic processes, the very certain thermodynamic climate change impacts can be masked to some extent by the much less certain dynamic climate change impacts. This uncertainty can lead to falsely asserting that there is no overall impact – a type II error. Trenberth et al. (71) argue that it is better to focus on the aspects of the event that are well understood, for example by conditioning analyses upon the large scale circulation of the event in question, thus removing the potential very uncertain dynamic aspects of climate change. This suggestion was extended and discussed at length by Shepherd (78), and has become the basis for the new kid on the block in extreme weather attribution: the storyline approach.

1.3.2 Attribution through storylines

The storyline approach (or ‘Boulder’ approach, 24) aims to determine the contribution of various causal factors that to the extreme event, and considers how anthropogenic climate change has affected those factors (and thus the extreme) in a deterministic manner. This approach was first applied by Hoerling et al. (65) to the 2011 summer combined Texas heatwave and drought. They used a variety of simulations, including atmosphere-only and coupled climate model runs, and seasonal forecasts. They examined the influence of rainfall deficit in the months preceding the heatwave, SST influence on the drought, and the overall predictability of the extreme at the start of May. As such, their intended goal was much broader than just assessing the anthropogenic contribution to the heatwave, aiming to advance the overall predictability of such events by examining and understanding many drivers, both human and natural. This approach of trying to disentangle and quantifying the contributions of many different drivers individually laid the basis for what most would understand as the storyline approach to extreme event attribution.

The next key text that I discuss is a perspective, “Attribution of climate extreme events” by Trenberth et al. (71). This highlighted the potential for conventional

probabilistic attribution to suffer from type II errors due to the large component of extreme weather events that arises from internal variability of the climate, thus masking any clear anthropogenic signals. The authors suggest that attribution studies should focus on the drivers of extreme weather and associated impacts whose response to climate change is determinable given the present state of knowledge. For instance, following a tropical cyclone event, rather than focussing on the probabilistic question of whether the unconditional probability of an event has changed, which would require assessing potential changes in large scale circulation patterns; they suggest focussing on aspects of the event whose response to climate change is well-understood and physically motivated, such as the increase in sea surface temperatures and available moisture leading to a deeper storm and more intense precipitation. They distil their approach down to answering the conditional question: ‘given the change in atmospheric circulation that brought about the event, how did climate change alter its impacts?’. In general, this is equivalent to determining how the known changes in the climate system’s thermodynamic state have affected the event in question. Of particular interest in relation to this thesis is their suggestion that ‘one needs to be able to simulate the event in question (perhaps with short-term forecasts...)’, since this is a strong motivating factor for the overarching approach taken here.

A key figure in the recent advancement of the storyline approach is Theodore Shepherd. Shepherd (78) discusses the differences in framing between the probabilistic and storyline approaches to extreme event attribution and demonstrates that they can be cast into a single framework based on conditional probabilities. He also discusses similar issues with probabilistic attribution to those I have outlined above, including the class-based framing and climate model deficiencies. In Shepherd et al. (80), the authors present an argument for the use of storylines more generally to understand and communicate the physical impacts of climate change in both present events and future projections. They summarise their argument in four reasons:

- improving risk awareness due to the episodic (experience-based) nature of storylines as opposed to the semantic (knowledge-based) nature of conventional probabilistic approaches, given humans are more likely to respond to episodic rather than semantic information;
- strengthening decision-making by allowing decision-makers to work backwards from a particular vulnerability incorporating climate change information with other factors, or to develop stress-tests based on (modified)

historical events;

- physically motivated partitioning of uncertainty - explicitly drawing the distinction between the more certain thermodynamic aspects and the less certain dynamic aspects of climate change. The basis for this argument is set out in Trenberth et al. (71) as discussed above;
- exploring the boundaries of plausible risks by combining information from climate model simulations with scientific understanding of the climate system rather than simply relying on quantitative results from climate model simulations. In this way, they suggest that the storyline approach could guard against surprise impacts from processes not well simulated or whose uncertainty is not well reflected by the current range of climate model projections, using local-scale precipitation events as a key example.

Now that I have discussed the formal basis for the storyline approach, here I present a selection of recent studies that apply it to various extreme weather events. Van Garderen et al. (81) introduces a methodology for extreme event attribution based on constraining the large-scale circulation in a climate model to that observed in reality. They employed global spectral nudging to push the mid-troposphere to upper-stratosphere (the ‘free’ atmosphere) towards reanalysis data, allowing the thermodynamic fields they were interested in (temperature and moisture) to respond to the imposed circulation. Their aim was to ‘constrain the model as little as possible... while still achieving an effective control of the large-scale weather situation’. The produced factual and counterfactual nudged simulations of the 2003 and 2010 heatwaves in Europe. The factual simulations were based on using present-day observed reanalysis, SST, and greenhouse gas concentration data; while in the counterfactual simulations, the SSTs were perturbed based on estimated changes since the pre-industrial, and GHG concentrations were changed to their pre-industrial levels. They compared their factual and counterfactual simulations to find that (domain-averaged) anthropogenic contributions to the 2003 and 2010 heatwaves were 0.6 °C and 2 °C respectively. This approach was further developed by Benítez et al. (82). Unlike Van Garderen et al. (81), who used prescribed SST patterns and an atmosphere-only model, Benítez et al. used a coupled climate model, in order to permit ocean atmosphere interactions that may be associated with the particular event of interest. To do this, they branched from free-running simulations of a coupled climate model at particular levels of global warming (pre-industrial, present-day, 2 °C and 4

°C). They allow for one year of spin-up time to allow slowly responding fields to develop prior to the event of interest. The event that they focussed on was the July 2019 heatwave in Germany, which they assessed had been made 2 °C warmer since pre-industrial times. Overall, their analysis could be considered an archetypal storyline approach to a heat extreme, and provides a useful contrast to the approach developed over the course of this thesis. The final study I discuss here, Schaller et al. (83), does not concern a heatwave, but a flood event. However, it is notable due to the authors' comprehensive analysis of a single event and focus on using operational systems as they are familiar to the stakeholders who would benefit most from the information provided by the study. They combined high-resolution climate model simulations and regional model downscaling of the same simulations with hydrological models to provide physically plausible storylines of present-day and future flooding associated with atmospheric river events in western Norway.

The storyline approach is able to address several of the issues with probabilistic event attribution raised above, and is clearly able to provide extremely valuable information about the changes in present extremes due to human influence on the climate, as well as potential future changes. However, it is unable to quantitatively estimate the changes in probability of such extremes — the aim of the probabilistic approach. I argue that provision of this probabilistic information remains important for public communication of climate change risks and potential future litigation. In this thesis I primarily explore an approach that leverages numerical weather prediction models in order to synthesise between the storyline and probabilistic frameworks, alleviating many of the issues of the latter I discussed above whilst maintaining the ability to provide meaningful probabilistic information.

1.3.3 A forecast-based approach

Although often considered distinct fields of research, weather prediction and climate projection are ultimately seeking to understand the same physical system — just on very different timescales. As such, the models used in each field share many similarities and components. For example, the Met Office Hadley Center uses a unified framework for weather and climate modelling based on their 'Unified Model' (84). Differences between the two begin to arise when considering the questions that each field has traditionally aimed to answer. Weather prediction tries to determine how the atmosphere will evolve over typical timescales of hours to months. Due to the chaotic nature of the atmosphere (85), a key part

of weather forecasting is estimating its initial state as accurately as possible — weather prediction is an initial value problem. Since many weather phenomena of interest are highly localized (for example, convective storms in the UK), providing information on a similarly local level is important, hence weather prediction models are typically run at very high resolutions. On the other hand, climate simulation and projection does not try to estimate what the precise state of the earth system will be in the decades to come, but to produce realisations that lie within the space of all possible states. By averaging such realisations over many years or different simulations (ensembles), you can build up a complete picture of this space (e.g. ‘Weather is what you get, climate is what you expect’). An important aspect of climate projection is that where this space lies depends strongly on an external forcing you apply to the system — climate projection is a boundary-value problem. Due to the very long simulations and traditionally larger spatio-temporal scales of interest, climate models are typically run at much lower resolutions. This difference in typical resolution provides the first justification for why forecast-based approaches to extreme event attribution: using these higher-resolution forecast models may resolve many of the known issues and biases in climate models, especially in the context of extreme weather ([60](#), [72](#), [73](#)).

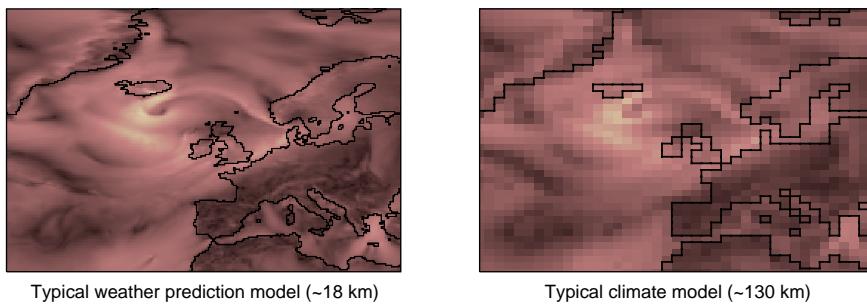


Figure 1.1: Comparing typical horizontal resolutions of weather and climate models. The left panel shows wind gusts at 10 m on 2021-10-01 00:00:00 over Europe, based on reanalysis ([86](#)), at the 18 km resolution (T_{co}639) of the ECMWF IFS ensemble prediction system. The right panel shows the same image, but at the 130 km resolution (N96) of the HadGEM3-GC3.1-LL climate model. This is arguably a generous comparison to the climate model, since this figure shows the same field at different resolutions, while it is conceivable that some features would not be able to be simulated at all in the lower resolution climate model.

One way in which weather prediction and climate simulation or projection fundamentally differ is in the extent to which they can be validated. Forecasts from weather models are able to be validated against reality (once reality has

come to pass), but this is not possible with climate models since they do not aim to precisely replicate reality (only produce a possible realisation). This intrinsic validation of weather forecast models has led to a significant amount of work in which they are used to examine the sources of biases and errors in climate models (87), or are used to calibrate climate projections (88, 89). This is the philosophy behind seamless prediction: that in a unified weather to climate prediction system, insights gained from the validatable initial-value problem can be transferred to improve the confidence in answers to the boundary-value problem. In the context of extreme event attribution, this idea has been discussed and examined by Bellprat et al. (62), Palmer and Weisheimer (76), Bellprat and Doblas-Reyes (77), Weisheimer et al. (90), and Lott and Stott (91). These studies have typically focussed on using the known ‘reliability’ of weather prediction systems to calibrate the outcome of probabilistic extreme event attribution. Here I use reliability in the statistical sense, meaning that the forecast probabilities of a particular event match up with the observed frequency at which that event occurs (92). Although such calibration is vital for improving the trustworthiness of existing probabilistic approaches to attribution based on climate models, I argue that even more benefit would be gained by using weather prediction systems for attribution directly. The key advantage of using a weather prediction system directly is that you can know with certainty if the model is capable of simulating the event you are interested in. If a weather forecast model is able to successfully predict an extreme event, then you can have considerable confidence that the model is able to represent all the important physical processes involved in the development of the event. You can also have confidence in the response of these — well represented in the model — processes to external forcing. This translates to increased confidence in any attribution statements that arise from analysis of the model. The same is not true of climate model based analyses. It would be incredibly difficult and involved to test whether a climate model is even able to simulate a specific extreme weather event and all the associated processes and drivers as they evolved in reality. In the conventional approach to probabilistic attribution, the validation of models is purely statistical (61).

This intrinsic validation of using weather forecast models directly is related to another motivating factor. Given a successful prediction, weather forecast ensembles produce realisations of the specific event of interest. This means that an attribution analysis based on such an ensemble can claim to be attributing the individual weather extreme, and therefore answering the question of how human influence has affected the probability of that precise event. This is in

contrast to conventional probabilistic approaches based on climate models, which frame attribution analyses in terms of the event as one occurrence of a particular class of events (e.g. the 2019 heatwave in Europe might be abstracted into the class of annual maximum temperature events). As discussed above, this means that the physical processes behind the event, that may be unique to that specific event, are not taken into account. The question answered by conventional probabilistic attribution is somewhat different and less specific. For the 2019 heatwave example, this question would be: how has human influence affected the probability of the peak annual temperature exceeding the value experienced during the 2019 heatwave (93)? The more specific nature of the question answered by a forecast-based approach brings it closer to the episodic nature of the storyline approach. It may also be useful when considering the legal contexts in which attribution studies are used (41).

Operationalising attribution

One of the traditional ‘problems’ with extreme event attribution has been that the timescales on which scientific research is published are very different from the timescales on which the public and other stakeholders are (increasingly) most interested in. When an extreme weather event occurs, interest in the media is usually highest in the days immediately following the event, while a scientific analysis would typically take months to be published. A second problem that is not unique to extreme event attribution is that studies tend to be focussed on the same regions that the research is conducted in. This leads to a bias in the coverage of such studies, with far more studies looking at events in the global North, especially Europe. The World Weather Attribution Project has sought to alleviate this bias by engaging with local researchers (68). A proposed solution to these issues is an operational attribution system, that could provide rapid results when extreme events occur using a well-established and consistent approach (34).

There are a number of practical reasons why a forecast-based approach is an attractive proposition as the basis for an operational attribution system (in addition to the science-focused arguments presented above). Firstly, weather models are run routinely on sub-daily cycles. If a methodology were created that allowed an operational weather forecast system to be easily switched to run in a counterfactual pre-industrial mode, then such simulations could be generated extremely rapidly. With sufficient interest and funding in the future, perhaps such counterfactual simulations could be run regularly alongside the routine operational

forecasts. Given the experience and expertise of weather forecast centres in the operational application of the science of weather prediction, it seems clear that they could add considerable value to the efforts towards operational attribution (44) — but especially if consistent models are used between the two. A final point to consider is that weather forecast model output is already used widely in hazard warning systems. The findings of extreme event attribution studies are of considerably relevance to such hazard warning systems, and other users in this space, since they can quantify how the risk of such hazards has changed due to climate change, and how the risk might change further in the future. Using tools that these users are already familiar with, in this case weather forecast models, will maximize the utility and impact of further studies (83).

1.4 Conceptualising different approaches

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 The meteorology of heatwaves

Thus far, I have discussed attribution of extreme weather in a general context — and the ideas and approaches above can be applied to any kind of extreme event. However, much of this thesis concerns heatwaves. There are several reasons for this. Firstly, heatwaves can have severe and wide-ranging socioeconomic and ecological impacts (94–97). Their broad response to climate change, and the underlying physical basis is well understood. Heatwaves have been the most common class of extreme weather event analysed in past attribution studies, which provides an extensive collection of literature for comparison over the course of this thesis (98). Due to the focus on heatwaves here, in this section I will review the typical meteorology of heatwaves, and the physical basis for their expected response to climate change. I will arrange this review in approximate order of

proximity (starting with planetary- and synoptic-scale drivers, and ending at the mesoscale), and focus on the meteorology of mid-latitude heatwaves, since this is where all the case studies covered in this thesis occurred.

I will begin by discussing features that would be classed as ‘dynamical’ drivers of heatwaves. One widely acknowledged feature is atmospheric blocking (99, 100). Blocking systems cover a wide range of atmospheric structures that are characterised by their persistence and stationarity. Although there is no unique accepted definition of a block, a rapid change from zonal to meridional tropospheric flow would be considered essential by many. In summer, over the regions where blocking generally occurs, zonal flow brings cooler oceanic air, and its interruption therefore leads to increased temperatures. Blocks often feature large anticyclonic anomalies, which can lead to subsidence under high pressure, adiabatically heating air as it falls in a so-called ‘heat dome’. Anticyclonic anomalies also reduce cloud cover and precipitation, increasing temperatures in summer even further. Their persistence enhances these effects by allowing them to build up over one or more weeks. These processes that lead to anomalously high summer temperatures are why blocking has been a key contributor to a number of noteworthy heatwaves (101). Although I will not discuss the precursors of blocking in depth, I note that some work has shown that blocks can arise as a result of tropical-origin wave-trains (102, 103), and can be preconditioned by specific regional sea surface temperature and sea ice patterns (104, 105). Blocking systems are projected to gradually decrease in frequency in the mid-latitude with global warming, though considerable uncertainty is associated with this change due to its sensitivity to the specific method used, lack of theoretical support, and poor representation in climate models (101). However, alongside this decrease in blocking frequency, some work has shown that block size might be expected to increase with global warming (106), which may change the spatio-temporal characteristics of weather extremes associated with these systems. This increase in block size has recently been suggested to be driven by stronger latent heating, which may also lead to an increased average block intensity (107).

Next, I shall review how synoptic-scale heat is actually generated from a physical, rather than a meteorological perspective. I shall examine three different processes by which heating occurs: advective, adiabatic and diabatic. Although these are clearly not entirely distinct in the atmosphere, I believe that treating each of these processes in turn will aid the clarity of this discussion. Advective heating occurs when regional temperatures increase as a result of warm air transport into the region. In the mid-latitudes, this warmer air is typically transported

from the tropics. Advective heating is dynamical in origin, and may arise as a result of a ‘ridged’ tropospheric flow, as was the case in the exceptionally warm February temperatures observed over the UK in February 2019 (108, 109). Adiabatic heating occurs when an air mass descends, compressing as the pressure increases, and thereby increasing its internal energy and temperature. This is the type of heating directly associated with anticyclonic blocking systems (or ‘heat domes’), in which high pressure causes air parcels to subside, increasing their temperature as they approach the surface. Diabatic heating is a broad class of many relevant processes, including condensational (latent) heating and radiative heating from absorption of infrared light, primarily by atmospheric water vapour, carbon dioxide and ozone. Rather than focussing on the broad atmospheric response to greenhouse gas emissions that arises as a result of these processes, which has been well-understood for a long time (1), here I shall briefly discuss feedbacks arising from these processes that may enhance heat extremes in particular due to human influence on the climate. Under normal meteorological conditions, the well-understood increase in atmospheric water vapour (110) results increased latent heating and cloud cover. However, under sufficiently anticyclonic conditions (for example in a ‘heat dome’) that inhibit condensation, an increase in water vapour will instead increase diabatic heating, thus increasing the moisture carry capacity of the local atmosphere, and potentially resulting in further radiative heating, feeding back into the intensity of the heat dome. Although Steinfeld et al. (107) showed that such moist processes only lead to modest changes in average block size and intensity, it is possible that such a feedback could enhance the most intense (and therefore potentially impactful) blocks by more than the average.

Finally, I consider meso- and local scale meteorological processes that control the intensity of heat extremes. As mentioned above, the anticyclonic systems often associated with heat extremes result in low cloud cover, due to the lack of condensation in subsiding air. This low cloud cover results in increased solar radiation at and near the surface. This increase in radiation results in diabatic heating, enhancing the impact of the (adiabatic) heating through subsidence. One way in which this radiative heating can be moderated is through the presence of soil moisture, which results in latent heating, thus reducing the energy transfer contributing directly to increased air temperatures. As a result, soil-moisture feedbacks are a key physical component of heat extremes that have been studied extensively (57–59, 111–115). Vogel et al. (113) partitioned the effect of soil moisture into soil-temperature, soil-precipitation and soil-radiation feedbacks

to allow for a clear explanation of the underlying processes involved. Soil-temperature feedbacks occur when dry soils lead to reduced latent heating, thus increasing temperatures and latent heat flux, further drying out the soil. Soil-precipitation feedbacks dampen the temperature response when an increase in soil moisture leads to increased latent heat flux, resulting in increased cloud cover and precipitation, thus further increasing soil moisture. Soil-radiation feedbacks are similar, in that reduced soil moisture results in reduced latent heat flux, thus reducing cloud cover and enhancing incoming solar radiation, thus further drying the soil. Fischer et al. (57) additionally demonstrated for the case of the 2003 heatwave over Central Europe that reduced soil moisture enhanced positive geopotential height anomalies, thus increasing the persistence and strength of the associated anticyclonic circulation, further reducing soil moisture. Human influence is understood to be reducing soil moisture on a global scale (15), though the confidence in this impact (and its sign) varies considerably across regions. Marvel et al. (116) used detection and attribution techniques to find an emerging greenhouse gas signal forcing the drying trends over North America and Eurasia. Confident detection and attribution of these signals is impacted by the influence of anthropogenic aerosol emissions on regional hydroclimate, particularly during the mid to late twentieth century. Aerosol influence on such trends has a distinct spatial pattern that masks the greenhouse gas pattern and associated signal (14). Overall, however, reductions in soil moisture due to anthropogenic greenhouse gas emission are expected to enhance human influence on extreme heatwaves through the mechanisms discussed here.

1.6 What to expect in this thesis

This chapter has provided a review of the attribution of extreme weather events to human influence on the climate. I have introduced the question that this field aims to answer, motivated why it is an important question, and discussed several approaches to answering it. I finally provided an overview of the physical processes involved in generating the kinds of heatwaves that will be the focus of much of this thesis. This final section will summarise the content within the remaining chapters. One consistent feature of the science chapters (2–5) is that they begin with a ‘Chapter open’ and end with a ‘Chapter close’ section, written after I had completed them all. In the ‘Chapter open’ sections I discuss the aims of the chapter within the wider context of my whole PhD project, and

what the intended outcome of the chapter is. In the ‘Chapter close’ sections I consider how the outcome of the chapter fits within the whole project, what questions are still outstanding following the chapter, and how these questions influence the rest of the thesis.

In Chapter 2, I carry out a conventional probabilistic attribution study, following the methods I have described above. This analysis uses the hot summer over Europe in 2018 as its case study. The original motivation for this chapter was that following the event, groups from the UK Met Office and WWA released attribution statements, coming to apparently conflicting results that differed by an order of magnitude. This study aimed to resolve the discrepancies between these two estimates by examining the influence of how the event is defined, in terms of its temporal and spatial scale, upon the quantitative attribution results. Within the context of this thesis, this chapter provides a more in-depth introduction to the techniques and presentation of results within the conventional probabilistic attribution framework.

Chapter 3 introduces the approach that is the primary outcome of this thesis: extreme weather attribution using counterfactual weather forecasts of specific events. In this chapter, I look at the case study of the exceptional warmth experienced over much of Europe at the end of February 2019, and confine the scope of the attribution to examining the direct radiative influence of increased CO₂ levels over pre-industrial levels. I start from the successful operational ECMWF forecast ensemble system, re-run the ensemble with pre-industrial CO₂ levels, and look at the impact this boundary condition perturbation has on the extreme heat. With this partial attribution, I show how a forecast-based approach might be carried out in practice. I discuss how our forecast-based approach differs in terms of the question it is answering to conventional probabilistic approaches, alongside some of the arguments for and against using such an approach.

In chapter 4, I develop the partial attribution study in chapter 3 by additionally taking anthropogenic influence on ocean heat content into account. I do this by perturbing the initial ocean (and sea-ice) state of the weather model in such a way as to remove the human fingerprint from the forecast initial conditions. By perturbing both the initial and boundary conditions of the weather model, I can more completely estimate the human contribution to an individual weather event. The event that I study in this chapter is the 2021 Pacific Northwest Heatwave, an unprecedented extreme that generated significant attention from the media when it occurred. One reason why this event is of particular interest is because the conventional probabilistic approaches appear to break down when faced with

such an exceptional outlier in the context of the historical record. I show that the forecast-based approach taken here represents a robust and appropriate alternative to these conventional approaches.

The final science chapter 5 focuses on projections of future climate, a field which has strong links to attribution. However, unlike the other chapters in which I attempt to increase the specificity of attribution frameworks, in this chapter I explore questions over how to estimate and sample the full space of uncertainty surrounding future weather extremes. I use a novel approach to generate very large ensembles of extreme winters in an atmosphere-only model, comparing these ensembles to relatively smaller ensembles from a coupled model. I show that this novel approach represents a very efficient methodology for the provision of very extreme winters over the UK, which could be of considerable value for climate change adaptation planning. I consider the question of how the forecast-based approaches to attribution could be appropriated to provide physically-consistent samples of specific high-impact extreme weather events as if they occurred in a future climate. This forecast-based approach to climate projections would provide very specific information, in contrast to the general exploration of uncertainty allowed by the work in this chapter. Both approaches could, however, provide exceptionally useful information surrounding the risk from future extreme weather events.

Chapter 6 provides a discussion of this thesis. Beginning with a summary of the science I have presented, I move onto placing it within the body of previous work on the subject and assessing its novelty and utility. I consider some of the limitations within the studies and approaches taken here, and suggest ways in which they might be overcome with future work. There are a number of ways in which this work could progress further in the future, or could be used outside of attribution. I consider these possible directions and applications before giving my concluding thoughts.

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 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>

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

Contents

2.1	Chapter open	27
2.2	Abstract	27
2.3	The 2018 heatwave in Europe	27
2.3.1	Defining the event	28
2.4	Materials & methods	29
2.5	Results	31
2.6	Discussion	36
2.7	Chapter close	37

2.1 Chapter open

The aim of this chapter, both in this thesis and at the time I started the work, is to provide a practical example of how probabilistic extreme event attribution is typically carried out. At the time, the 2018 European heatwave was a clear case study to pick — it had not only been responsible for very significant damages, but had already been analysed by two separate attribution teams, who came to seemingly quite different conclusions. One group at the UK Met Office estimated that the heatwave had been made 30 times more likely, while the other group, the World Weather Attribution (WWA) project, stated that it had been made 2–5 times more likely due to human influence on the climate. These two numbers are an order of magnitude apart, and I aimed to determine why they were so different. Whilst resolving this discrepancy was the main research question, the other clear goal was to provide me with some experience in actually carrying out an attribution study, and therefore identify the many gaps in my knowledge that existed at the time. I collaborated with several researchers from the WWA to ensure that the methods I was using were consistent with those that they had used themselves.

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 (117, 118). Impacts were felt across Europe, with wildfires burning in Sweden (119, 120), heatstroke deaths in Spain (121), and widespread drought (122). During the summer, the World Weather Attribution (WWA) initiative released an analysis of the heat spell (123) 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 (124, 125). 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 under-represent the variability of temperature extremes, especially near the coast, with implications for any derived attribution results.

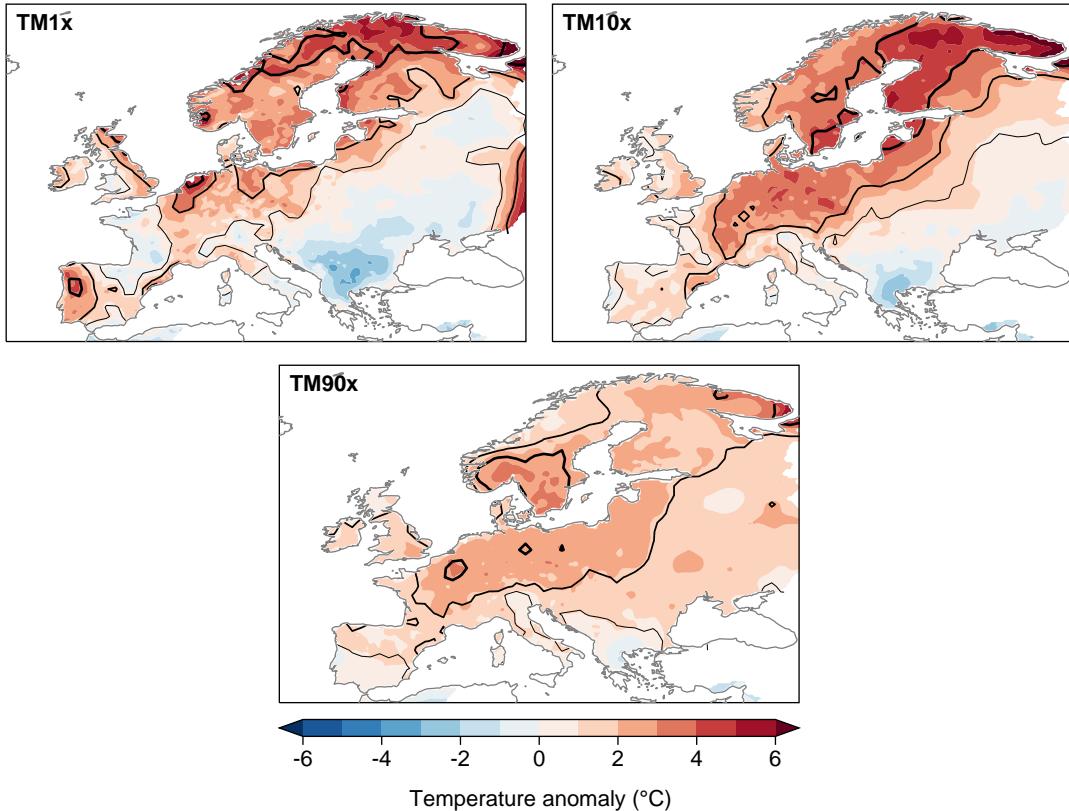


Figure 2.1: The 2018 heatwave in Europe: observed mean temperature anomalies over a range of timescales. Shading indicates mean temperature anomalies for the different temporal-scale heatwave metrics used. Black contours indicate z-scores of the 2018 heatwave for the three metrics based on detrended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

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 analyse 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 ([126](#)). 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 (70), 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 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 (127).

2.4 Materials & methods

Model simulations & validation

We primarily use three sets of simulations from the UKMO Hadley Centre HadGEM3-A global atmospheric model (128, 129). These are a 15-member historical ensemble (1960–2013; Historical), and 525-member factual (ACT, referred to as HistoricalExt by Ciavarella et al.) and counterfactual (a “natural” world without anthropogenic forcings; NAT, referred to as HistoricalNatExt by Ciavarella et al.) ensembles of 2018. Historical and ACT are forced by observed SSTs and sea ice concentrations from HadISST (130). NAT is forced by naturalised SST and sea ice concentrations estimated by subtracting the multi-model mean difference between the CMIP5 ‘historical’ and ‘historicalNat’ experiments (131), and run with pre-industrial greenhouse gas concentrations. For a complete description of these ensembles, see Ciavarella et al. (129). We compare results from these factual-counterfactual simulations with those from a trend-based analysis of the 15-member HadGEM3-A Historical ensemble, a 10-member ensemble from EURO-CORDEX (1971–2018; 132–134) and a 16-member ensemble from the RACMO regional downscaling of EC-EARTH 2.3 (1950–2018; 135, 136). The EURO-CORDEX ensemble used here is bias-corrected using the cumulative distribution function transform. Both the EURO-CORDEX and RACMO ensembles follow the CMIP5 ‘historical’ scenario to 2005 and the RCP4.5 scenario thereafter (131, 137). Observations are taken from E-OBS (1950–2018; 138) throughout. Initially, we performed our analysis with the weather@home HadRM3P European-25 km setup (139) but found that this model overestimates the variability over all Europe for daily through seasonal-scale event statistics, and so it was omitted.

Attribution methodology

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; 140–142). 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) 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 for the British Isles region in Figure 2.2. Using estimated event probabilities rather than observed magnitudes to define the event constitutes a quantile bias correction (143), accounting for model biases in the mean and variability of the temperatures simulated.

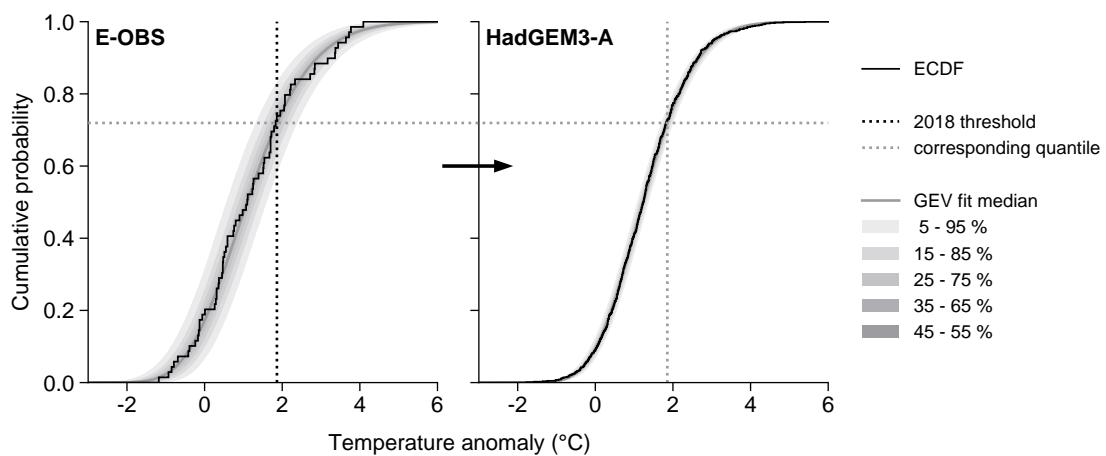


Figure 2.2: Calculating the heatwave threshold in HadGEM3-A from observations. The heatwave metric used here for illustration is TM1x for the BI region. Solid black lines indicate empirical CDFs. Solid grey lines indicate GEV distributions fit to the data, with grey shading indicating confidence intervals 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.

Counterfactual attribution

For the main analysis, which we term “counterfactual” attribution (25), 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_{ACT} and P_{NAT} . The estimated ACT and NAT distributions are shown for each metric in Figure 2.3. We express our results as the probability ratio, $PR = P_{ACT}/P_{NAT}$, representing the fractional change in probability of the 2018 event in the factual compared to the counterfactual world.

Trend-based attribution

We support the counterfactual attribution with a trend-based analysis (144) 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.

Statistical methods

We fit EV parameters using the method of L-Moments (145), 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 non-parametric bootstrap throughout.

2.5 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][†] 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

[†]Numbers in brackets [] represent a 90% confidence interval throughout this chapter.

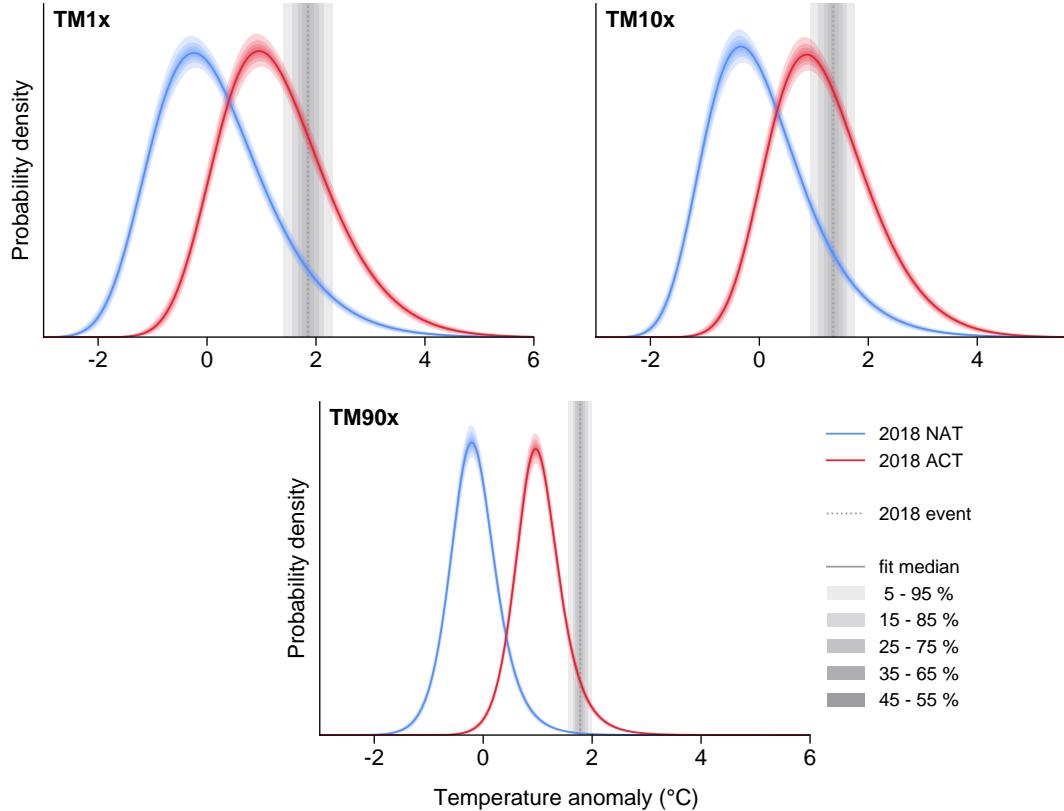


Figure 2.3: Factual and counterfactual PDFs of the 2018 heatwave defined over three temporal scales. The heatwave metric used is given in the title of each panel. Solid red and blue lines indicate best-estimate GEV distributions fit to HadGEM3-A 2018 ACT and NAT ensembles respectively. Dotted grey line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

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 (Figure 2.1). 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.

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. The corresponding best-estimate PRs are between 10 and 100 for all regions (Figure 2.5), including those with lower return periods.

A trend-based analysis yields similar results, with PRs for the British Isles region

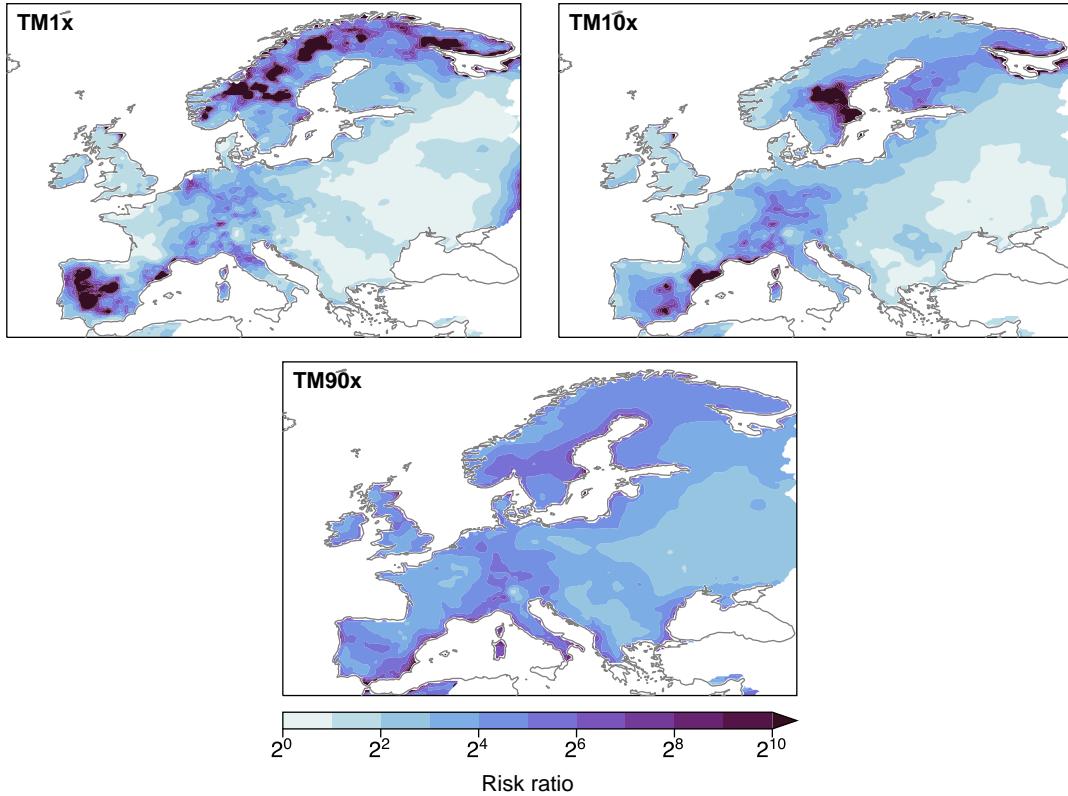


Figure 2.4: Maps of the estimated change in probability of the 2018 heatwave due to anthropogenic influence on the climate. The heatwave metric used is given in the title of each panel. Shading illustrates risk ratio of 2018 event at each gridpoint computed using HadGEM3-A ACT and NAT ensembles.

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 is 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 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.

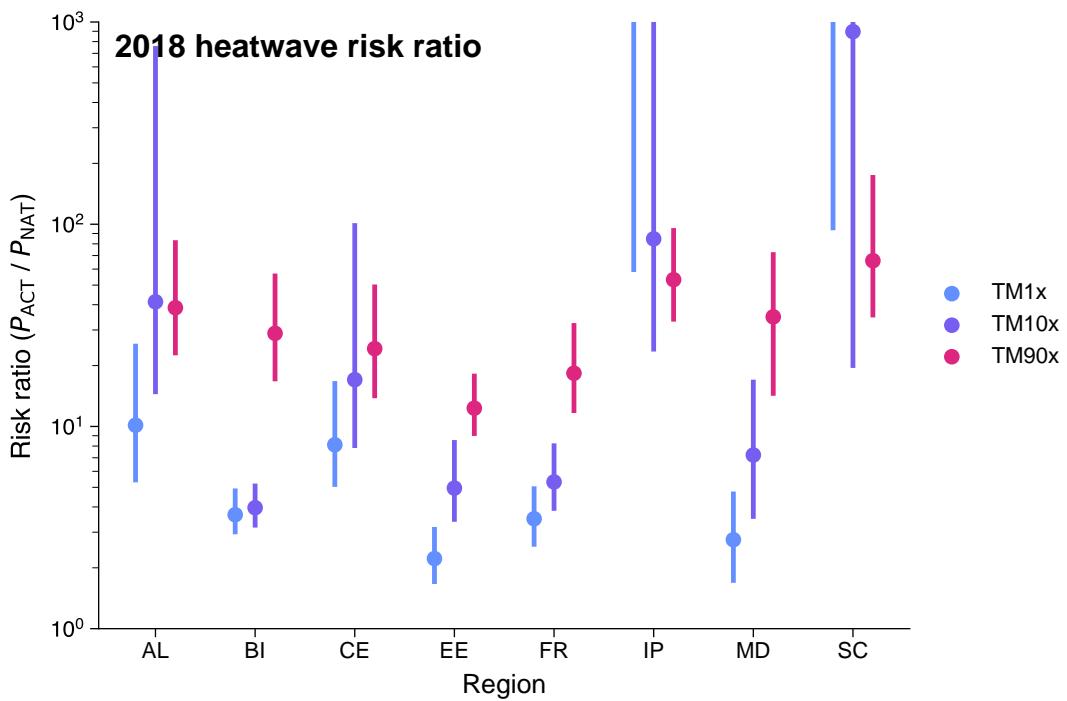


Figure 2.5: Estimated changes in probability of the 2018 heatwave defined using regional mean temperatures. Colour indicates heatwave metric. Dots indicate central risk-ratio estimate and bars indicate 90 % confidence interval.

The PR increases with the event statistic timescale 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 timescale 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. 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.

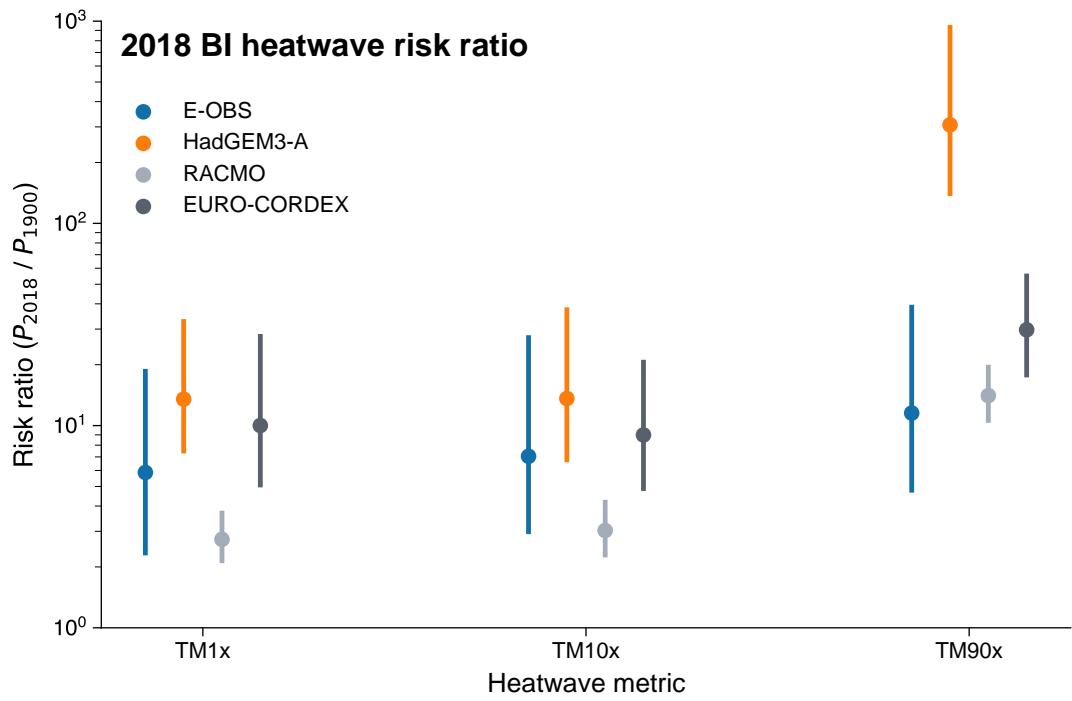


Figure 2.6: Estimated changes in probability of the 2018 British Isles heatwave across a range of observations and model simulations. Here, risk-ratios are estimated using a trend-based analysis. Colour indicates data source. Dots indicate central risk-ratio estimate and bars indicate 90 % confidence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

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 2.6). 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 (62) and means that

the trend-based PRs for this model and over the British Isles region presented here, especially for TM90x, are likely to be overestimated. Underrepresented variability often occurs in prescribed-SST models (146, 147) 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 under-represents 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.

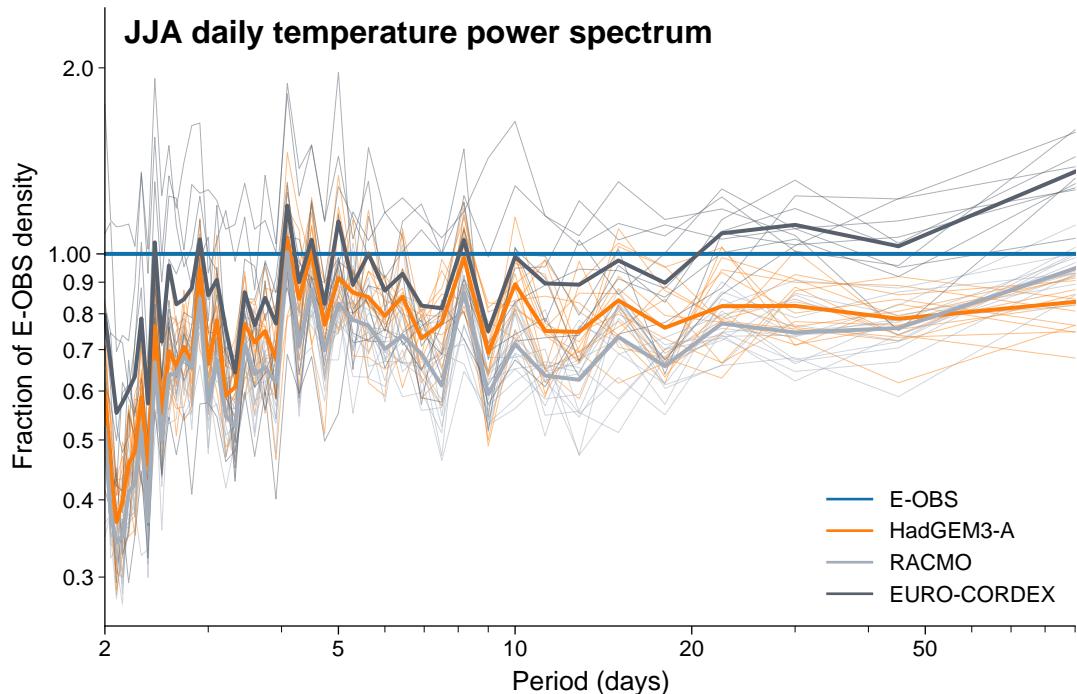


Figure 2.7: Historical power spectrum of summer daily mean temperatures over the British Isles across a range of observations and model simulations. Power spectra are estimated as periodograms averaged over all historical years available for each data source, expressed as a fraction of the E-OBS power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

2.6 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. (148), 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. 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 (62). 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.

2.7 Chapter close

At the start of this chapter, I set out that I had two main goals for the study: to understand how two seemingly contradictory attribution results arose; and to gain practical experience in carrying out probabilistic attribution analyses myself. I believe that the contents of this chapter demonstrate that I achieved the first, and (hopefully) that the contents of this thesis demonstrate that I achieved the second to the degree required by my work. However, although both of these outcomes were realised, this study left me with a number of outstanding questions. Can we claim to be attributing a specific event if we simply consider all weather events that happened to be the hottest during a particular year, without considering anything else about the meteorology? Is a purely statistical model validation sufficient? How important really is ocean-atmosphere coupling? How do you decide how to define your event given that this definition can have huge implications for the quantitative result? Should we include more information about the event in the event definition (such as the atmospheric flow, or other variables like soil moisture), in order to provide a more event-specific analysis?

Several of these questions are highly related (especially those concerning event specificity and model validation). Following this work, I spent some time looking into the biases in variability in the models I had used. Although my work into these biases did not

lead to any concrete and publishable results, I have attempted to address a number of the other questions raised here in the remainder of the thesis. In particular, the questions surrounding event specificity might be answered by using conditioned model simulations. This conditioning could be applied in many different ways, but one particularly attractive one would be to use successful weather forecasts. In this way you could reasonably claim to be attributing the *specific* event in question, but also have a much higher degree of confidence that the model being used was actually able to simulate such an event in the way that it had unfolded in reality. As a result, much of the rest of this thesis, and its contribution to the field, is concerned with the idea of forecast-based attribution.

Quote

— author

3

Forecast-based attribution: perturbing the boundary conditions

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 CO₂ concentration at initialisation. These perturbed forecasts allow me to estimate the direct contribution of diabatic heating due to CO₂ 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 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>

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

Contents

3.1	Chapter open	41
3.2	Abstract	41
3.3	Introduction	42
3.4	The 2019 February heatwave in Europe	44
3.5	Materials & methods	46
3.6	Forecasts of the heatwave	46
3.7	Perturbed CO ₂ forecasts	50
3.8	Attributing the heatwave to diabatic CO ₂ heating	53
3.9	Discussion	57
3.10	Chapter close	60

3.1 Chapter open

Just after the exceptionally warm period in February 2019 that is the subject of this chapter, my co-authors and I discussed this extreme event. Two features were particularly noteworthy: it appeared to be a particularly radiatively driven heat event; and it had been forecast exceptionally well at least a week in advance. Although at this point we had already discussed performing counterfactual forecasts using perturbed initial conditions, we hadn't yet worked out how we would actually achieve this in practice, and it seemed a long way off. Because of this, and the apparent radiative nature of the event, we wondered if we could start off by simply changing the CO₂ concentrations in the model — and leaving everything else the same. Although this would only represent a very partial attribution, to the direct radiative effect of increased CO₂ concentrations over pre-industrial levels, Baker et al. (149) had recently published a study that suggested we might still find that this direct effect would be sufficient to notice a difference. It turned out that changing the CO₂ concentrations in the model was relatively straightforward, and after waiting some time for computing resource to be granted, we were able to perform these perturbed-CO₂ counterfactual forecasts. This partial attribution would allow us to begin exploring a number of questions we had on the approach: how would the predictability of the heatwave change when we changed the CO₂ levels?; how would the attribution statements depend on lead time?; would the direct effect of CO₂ be large enough for us to even detect it? All of these questions are relevant to not only the partial attribution presented in this chapter, but also the more complete attribution based on perturbed boundary *and* initial condition forecasts that is the primary aim of this project.

3.2 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 methodology for attribution of extreme weather events using the operational European Centre for Medium-Range Weather Forecasts (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 synthesize 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 CO₂ concentrations on the exceptional European winter heatwave of February 2019.

3.3 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 (19) over the past twenty years; it now has an annual special issue in *The Bulletin of the American Meteorological Society* (150). Extreme event attribution is of particular importance for communicating the impacts of climate change to the public (151, 152), 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 exist numerous methodologies for carrying out an event attribution (153). Many of these rely on large ensembles of climate model simulations, the credibility of which has been questioned by recent studies (62, 76, 77). A particular issue is the dynamical response of the atmosphere to external forcing, which is highly uncertain within these models (78). 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 national and research centres. 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. (65) used two seasonal forecast ensembles to examine the predictability of the 2011 Texas drought/heatwave

within a comprehensive attribution analysis involving several types of climate simulation. Meredith et al. (154) used a triply nested convection-permitting regional forecast model to investigate the role of historical sea surface temperature (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. (81) 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 (155–157). They used a seasonal forecast model to assess anthropogenic CO₂ contributions to record-breaking heat and fire weather in Australia. Two more similar studies carried out forecast-based hurricane attribution studies (158, 159). 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. (160), 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 centres (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 ocean-atmosphere 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 CO₂ 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 dynamic and thermodynamic features of the event. In [Perturbed CO₂ forecasts](#), we outline the experiments we have performed in order to quantitatively determine the direct CO₂ 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.

3.4 The 2019 February heatwave in Europe

Between the 21st and 27th February 2019, climatologically exceptional warm temperature anomalies of 10-15 °C were experienced throughout Northern and Western Europe ([108](#)), as shown in Figure 3.1A. 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 Figure 3.1C ([138](#)). Figure 3.1D, 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 ([161](#)), as shown in the geopotential height field in Figure 3.1A. This dynamical driver was accompanied by another synoptic feature that further enhanced the warming: widespread clear skies between the 25th – 27th, shown in Figure 3.1B. These clear skies resulted in a widespread and persistent strong diurnal cycle, reaching 20 °C in some locations. Further details of the meteorological mechanisms and historical context of the heatwave are provided in Young and Galvin ([108](#)), Kendon et al. ([162](#)), and Christidis and Stott ([163](#)).

To quantify the direct impact of CO₂ 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 ([63](#), [148](#), [164](#)). The most remarkable feature of the February 2019 heatwave was the maximum temperatures observed, which peaked between the 25th and 27th 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 by Christensen and Christensen ([126](#)). 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.

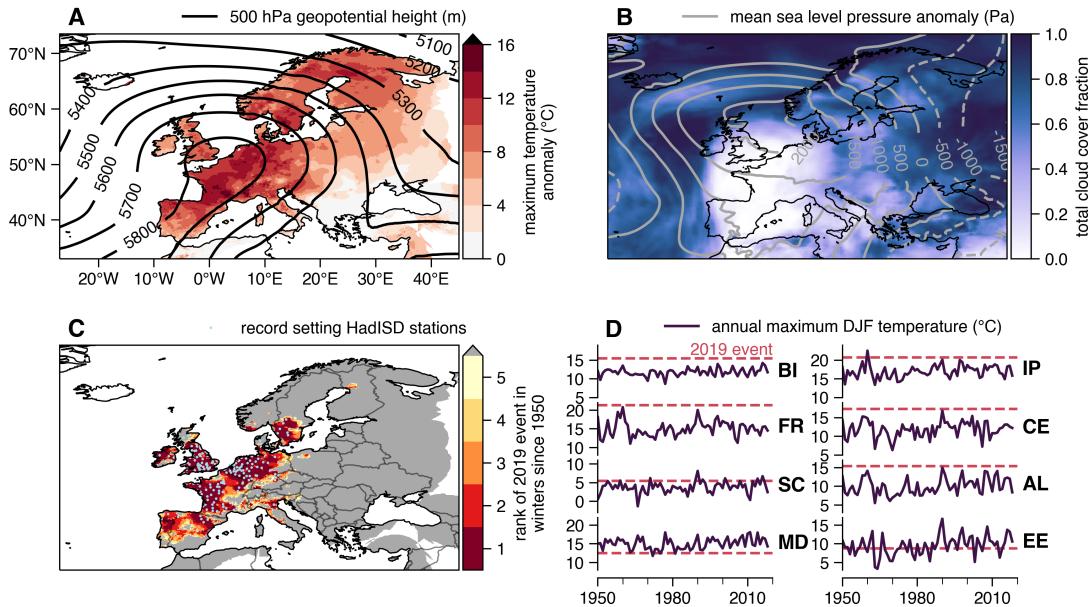


Figure 3.1: The 2019 February heatwave in Europe: synoptic characteristics & historical context. **A**, maximum temperature anomaly in E-OBS with overlying contours of mean Z500 anomaly from ERA5 (86) over 25-27 February 2019. **B**, mean total cloud cover with overlying contours of mean sea level pressure anomaly averaged over 25-27 February 2019. **C**, rank of the maximum temperature in E-OBS over 25-27 February 2019 out of all winter temperature maxima since 1950 and light-blue scatterplot of 216 HadISD stations (with > 30 winters of measurements) which recorded their highest observed value over the same three days (165–168). **D**, historical winter maximum regional mean daily maximum temperatures in E-OBS. Solid purple line shows timeseries of winter maxima for 1950–2018; dashed pink line indicates maximum value observed over the 25-27 February 2019. Regions are as Christensen and Christensen (126).

3.5 Materials & methods

The ECMWF Integrated Forecasting System

In this study, we use the Integrated Forecasting System (IFS) model cycle CY45R1, the operational cycle at the time of the event. The 51-member ensemble prediction system comprises a 91-layer, TCo639 resolution atmospheric model coupled to the 75-level, 0.25° resolution Nucleus for European Modelling of the Ocean v3.4 ([169](#)). Once the model integration reaches the extended range (day 15 onward), the atmospheric model resolution is reduced to TCo319.

The IFS model climatology We define the IFS model climatology, used to compute model anomalies and the continuous ranked probability skill score, in an identical manner as is done operationally (for example, to calculate the Extreme Forecast Index product [170](#)). This climatology is defined using nine consecutive reforecast sets, spanning 5 weeks centred on the forecast initialisation date (reforecast sets are run twice a week, every Monday and Thursday), of 11 members per reforecast. These sets are created by initialising the reforecast ensemble on the same calendar date over the previous 20 y. This procedure results in a model climatology of $9 \times 11 \times 20 = 1980$ members covering the 1999 to 2018 period. Throughout this article, we use the model climatology defined for the forecast initialised on 11 February 2019. Climatologies defined for other initialisation dates are virtually identical.

Statistical methods

Significance testing For the significance stippling displayed on the maps, we use a non-parametric (binomial) pairwise sign test at a 90% confidence level.

Distribution fitting When fitting statistical distributions to the ensembles during the risk-ratio calculation, we employ the method of L-moments ([145](#)), due to its numerical stability under small sample sizes.

3.6 Forecasts of the heatwave

This heatwave was well-predicted by the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system. Their coupled ocean-atmosphere forecasts indicated ‘extreme’ heat was possible at a lead time of around two weeks, and probable at a lead time of around ten days (Figure 3.2A), despite the exceptional nature of the heatwave in both the model climatology and real world. As expected,

the forecasts' 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 (Figure 3.2B).

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' (78) attribution studies (25, 63, 67, 171), 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 (e.g. 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 (172), or by nudging model simulations towards the specific dynamical situation that occurred during the event in question (81, 154). 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 (128, 129, 139), which may not fully capture all the physical processes required to credibly simulate the extreme in question (60). 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 (146). This can lead to studies overestimating the impact of anthropogenic activity on weather extremes (63, 77). More generally, Bellprat et al., and Palmer and Weisheimer (62, 76) 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

([128](#), [139](#)) 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 ([129](#)). 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 ([78](#)): 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 ([81](#), [173](#)). 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.

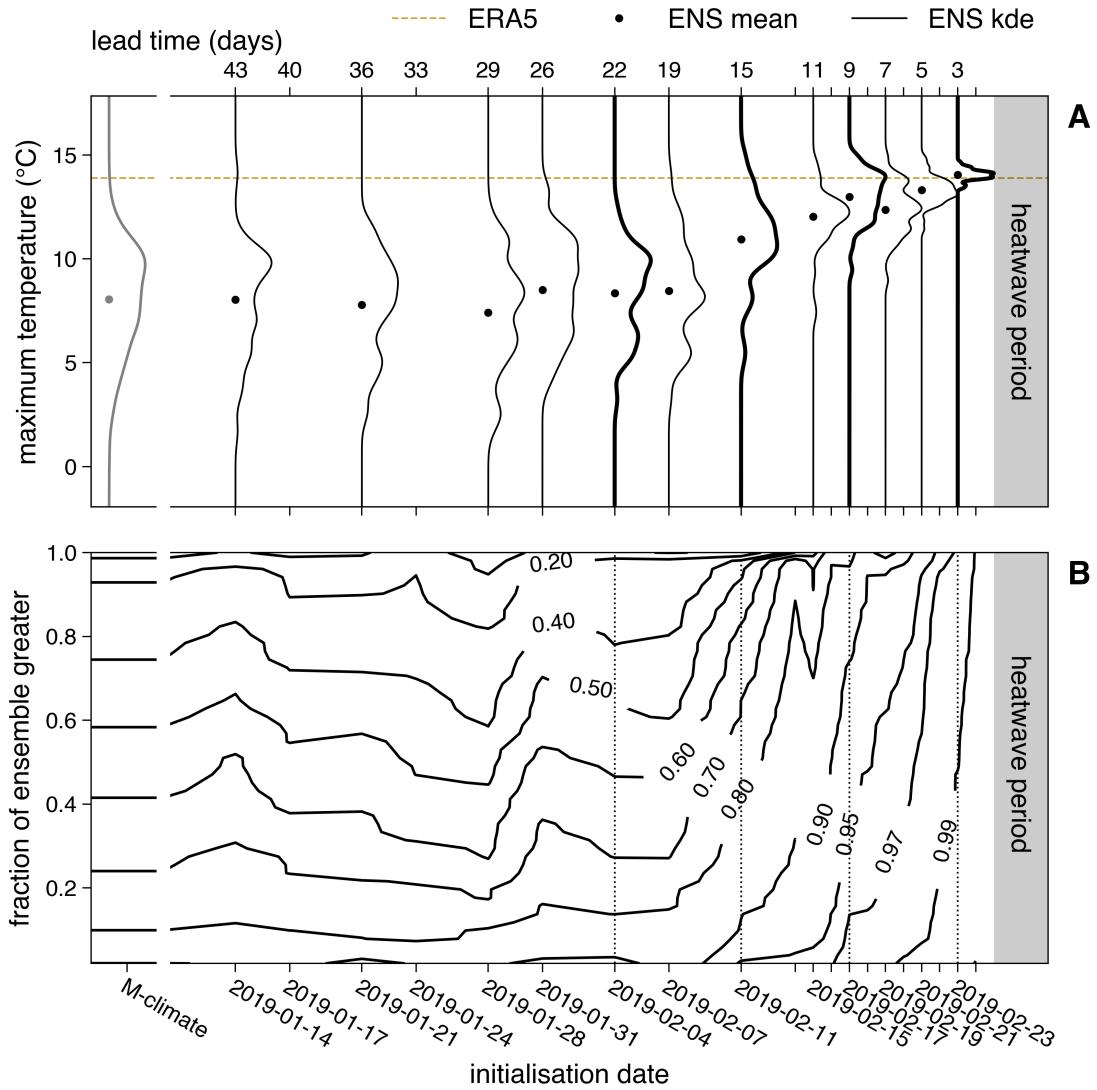


Figure 3.2: Medium- to extended-range forecasts of the heatwave. **A**, ensemble distribution of heatwave as event definition against forecast initialisation date for the British Isles region. Gray PDF on far left shows model climatology, thick black lines show lead times selected for the perturbed CO₂ experiments, dashed gold line shows heatwave magnitude in ERA5. Dots show ensemble mean. **B**, forecasts of Z500 over Europe during the heatwave period compared to ERA5. y-axis shows the fraction of the forecast ensemble with a pattern correlation at least as great as the levels indicated by the contour lines, against forecast initialisation date. Thin dotted lines show lead times selected for the perturbed CO₂ experiments.

3.7 Perturbed CO₂ forecasts

In this study we choose to only change one feature of the operational forecast in our experiments: the CO₂ concentration. This means that the analysis we carry out is limited to attributing the impact of diabatic heating due to increased CO₂ 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 CO₂ 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 (174). 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 (indicated by ‘ENS’ in the figures) we run two experiments using operational initial conditions and identical to the operational forecast in every way except the experiments have specified fixed CO₂ concentrations. One experiment has CO₂ concentrations fixed at pre-industrial levels of 285 ppm (PI-CO₂), while in the other they are increased to 600 ppm (INCR-CO₂). These represent approximately equal and opposite perturbations on global radiative forcing (175). We carry out these two experiments for each lead time, perturbing the CO₂ concentration in opposite directions, to ensure that any changes to the likelihood of the event can be confidently attributed to the changed CO₂ 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 (78). If this were the case we would see a reduction in event probability regardless of whether we increased or reduced the CO₂ concentration.

Some previous work has been done on the impact of reduced CO₂ 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 CO₂ concentrations (defined there as all the effects of CO₂ on climate beside those occurring through ocean warming), finding the direct effect of CO₂ 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 CO₂ on climate in global circulation models (176, 177). This work found that, in the absence of changes to sea surface temperatures or sea ice concentrations, a doubling of CO₂ concentrations would change global mean surface temperatures over land by ~ 0.4 °C. These early studies indicate that changes in global land temperatures are approximately

linear with the logarithm of CO₂ concentration.

We find that the best-estimate global mean change in land surface temperatures attributable to the additional diabatic heating due to CO₂ over pre-industrial levels (henceforth the ‘CO₂ 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] [†]. In general, the further away from the initialisation date, the slower the rate of change of the globally-averaged ensemble mean CO₂ signal, and the larger the ensemble spread (Figure 3.3A). While in experiments with prescribed SSTs, we might expect the CO₂ signal in surface temperatures to approach a maximum value within timescales on the order of months, in our experiments the CO₂ signal will likely continue to increase in magnitude for centuries due to the ocean-coupling, as is the case in the abrupt-4xCO₂ experiment carried out in the Coupled Model Intercomparison Project (CMIP) (131, 178, 179). The zonal-mean patterns of surface temperature CO₂ signal are qualitatively similar to those exhibited by CMIP5 and CMIP6 models during the abrupt-4xCO₂ experiment (180, 181), 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 (182).

We also examine the impact on the specific event dynamics over our region of interest; since these were crucial in developing the extremes observed. Figure 3.3B shows the growth in 500 hPa geopotential height (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 CO₂ heating, and not due to the attractor of the dynamical system having changed as a result of the perturbations we have made.

Figures 3.3C and 3.3D show analogous plots to 3.3B, 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 Figure 3.3C are near identical for all lead and validation times. Secondly, the error growth due to the CO₂ perturbation is slower than due to the initial condition perturbations; the errors in Figure 3.3C increase slower than in 3.3D. However, by the end of the longest lead forecast, we can see that the intra-ensemble errors have saturated,

[†]Numbers in square brackets [] represent a 90 % confidence interval throughout this chapter.

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.

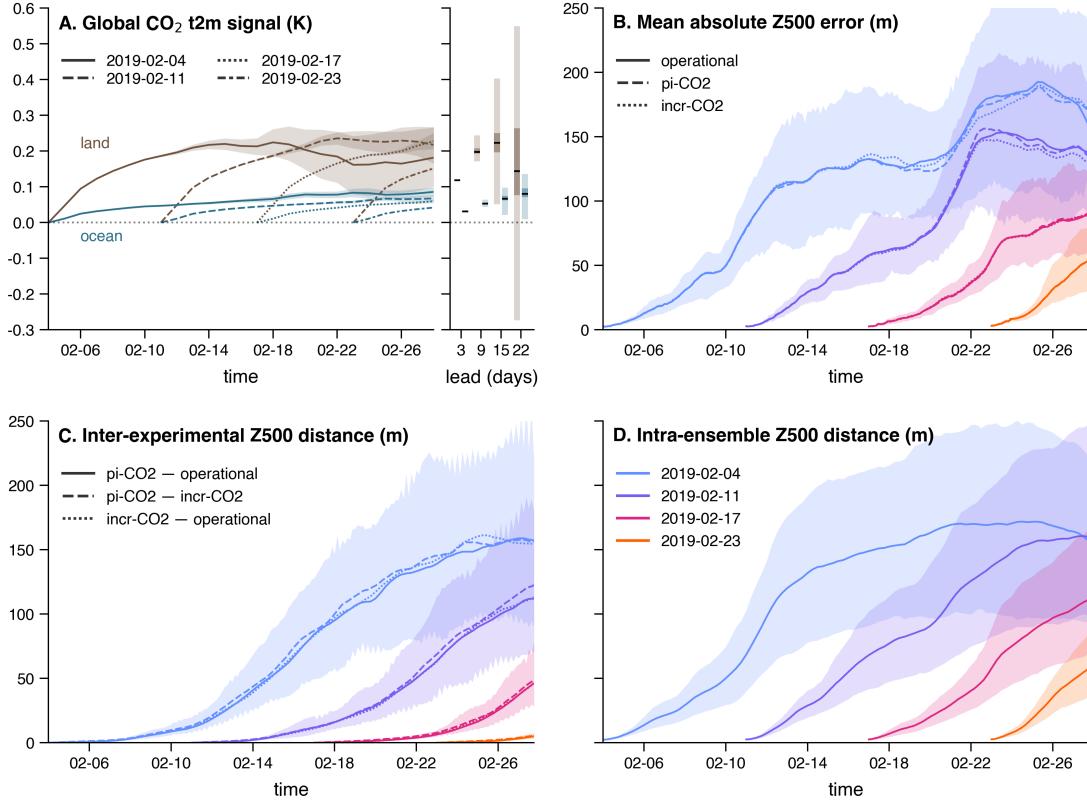


Figure 3.3: Global temperature and synoptic-scale dynamical response to CO_2 perturbations. **A**, CO_2 signal in global mean surface temperatures. Brown and blue features show quantities over land and ocean respectively. Line styles indicate initialisation date of the experiments. Boxplots show average over 25 to 27 February 2019, with the black line indicating the ensemble mean, dark shading the 90% confidence interval around the mean, and light shading the 90% range of the ensemble. **B**, mean absolute error in Z500 between experiments and ERA5. Colour indicates initialisation date and line style indicates experiment. Solid lines indicates ensemble mean. The shading shows the 5 to 95% range of the operational ensemble (ENS). **C**, as in **B** but for mean absolute distance between corresponding ensemble members of different experiments. Line style here indicates the experiments being differenced. The shading shows the 5 to 95% range of the differences between the PI- CO_2 and INCR- CO_2 experiments. **D**, as in **B** but for intraensemble distances of the operational ensemble.

3.8 Attributing the heatwave to diabatic CO₂ heating

First, we examine the geographical pattern of the CO₂ signal in the heatwave in Figures 3.4A-D. These indicate several key features of the attributable direct CO₂ effect on the heatwave. The CO₂ 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 CO₂ signal magnitude in the heatwave generally exceeds the signal in the global mean surface temperature (Figure 3.3A), in particular in Central Europe; possibly due to the high contribution of diabatic heating to the heatwave arising from ideal dynamical conditions. Figure 3.4E shows boxplots of the heatwave CO₂ signal for the three regions of interest. Although there is some region-specific variability, these reinforce the main messages illustrated by the maps: the CO₂ signal grows and decreases in confidence as the lead time increases.

In addition to the absolute impact of the direct CO₂ effect on the heatwave, we also carry out a probabilistic assessment of its impact, consistent with conventional ‘risk-based’ attribution studies (78, 183). Due to the distinct 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 CO₂ 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 CO₂ experiment as our ‘counter-factual’ ensemble. We calculate the probability of simulating an event at least as extreme as observed in the factual ensemble, P₁, and in the counterfactual ensemble, P₀. 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 = P₁/P₀, 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 (25, 35). 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 Figure 3.4F. There are several key factors that contribute to the best-estimate and confidence in the risk ratios: the CO₂ 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 CO₂ 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 (i.e. the closer the event to the ensemble centre), the greater the confidence.

We find that on the shortest lead time, the direct CO₂ 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 CO₂ 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 % (5th 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 CO₂ 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, Figure 3.2B).

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 (78). 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 CO₂ 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 CO₂ 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-CO₂ experiment to increase our confidence that the positive results we obtained in the probabilistic analysis above are in fact due to the direct CO₂ effect, and not just random variability. If CO₂ were driving the changes in event probability between the PI-CO₂ and operational forecasts, then we would expect to see an even more dramatic increase in event probability between the PI-CO₂ and INCR-CO₂ forecasts. This is indeed what we find. For all regions and lead times, our best-estimate change in event probability is above zero when CO₂ concentration is increased from pre-industrial levels of 285 ppm to 600 ppm. This therefore increases our confidence

further that the positive attribution to CO₂ 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 CO₂ 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 CO₂ effect. As is clear from the development of the CO₂ signal estimates with lead time, the model is still adjusting to the sudden change in CO₂ 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 CO₂ 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 ([163](#)).

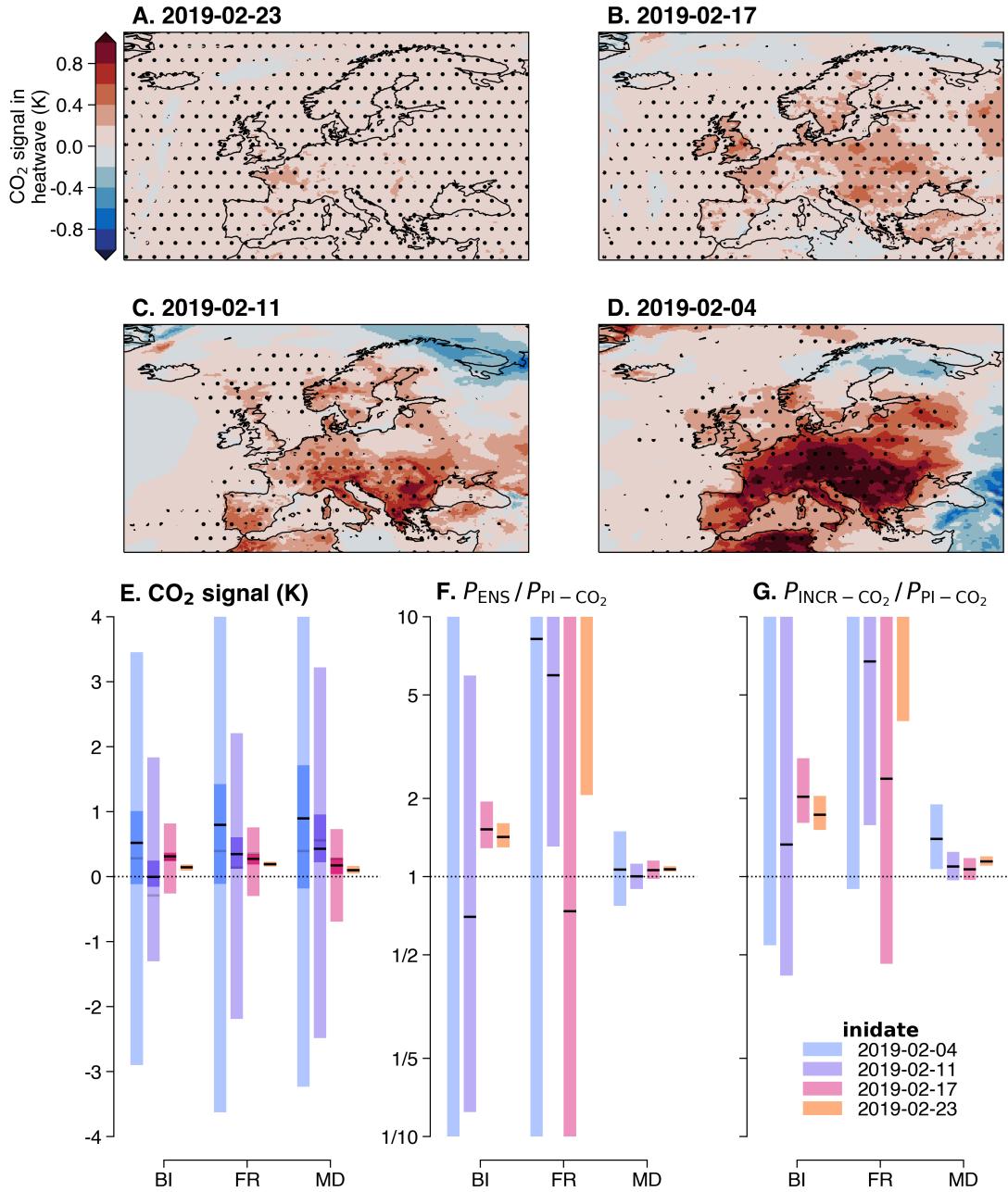


Figure 3.4: Attribution of the direct CO_2 influence on the heatwave. **A-D**, maps of the ensemble mean attributable CO_2 signal in the heatwave for the four forecast lead times, which are indicated by the subplot titles. Stippling indicates a significant positive signal at the 90% level. **E**, boxplot of the absolute CO_2 signal for the three regions of interest and over the four forecast lead dates. Black line indicates ensemble median. Dark shading indicates 90% confidence in the median, and light shading indicates 90% confidence in the ensemble. Gray line indicates median difference between the operational forecast and $\text{PI}-\text{CO}_2$ experiment. **F**, as in E, but showing risk ratios using the operational forecast as a factual and $\text{PI}-\text{CO}_2$ experiment as a counterfactual ensemble. **G**, as in F, but using the $\text{INCR}-\text{CO}_2$ experiment as a factual and $\text{PI}-\text{CO}_2$ experiment as a counterfactual ensemble.

3.9 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 CO₂ 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 CO₂; inspired by the ‘storyline’ and ‘risk-based’ attribution frameworks ([25](#), [78](#), [183](#), [184](#)).

There are several advantages associated with this 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 almost always have a higher resolution than the models used for climate simulation. In addition, the forecast model used here is coupled ocean-atmosphere, while the large climate model ensembles used for attribution often use prescribed sea surface temperatures ([129](#)). The use of prescribed SSTs can lead to model biases that project strongly onto attribution results ([146](#)). 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 ([61](#)). 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 ([172](#), [185](#)). 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?’. The latter 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 (157).

In addition to its advantages, this 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 CO₂ on the event is extremely low (since we are extrapolating the distribution shape outside the range of our data). Although the estimates of the absolute impact of CO₂ do not share this lack of confidence, this is still a problem. It is possible that applying some bias correction procedure (e.g. 143, 186, 187) 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 (i.e. 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 (155), at least in the case of the CO₂ perturbations applied here. This adjustment is clear on a global scale in Figure 3.3A. 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 CO₂ 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 CO₂ 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 (Figure 3.2B). These estimates of the impact of CO₂ 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 (78, 80, 184). 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 (25, 61). At this lead, we find that although over all regions the best-estimate impact of the direct CO₂ 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 (78, 183). 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 CO₂ 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. 174). 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. (155–157). 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 (i.e. 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 CO₂.

concentration. It is possible that seasonal forecasts have the greatest potential to target for an operational forecast-based attribution methodology.

3.10 Chapter close

There are a number of interesting outcomes from this chapter. The most important one was that this approach to attribution, using reinitialised state-of-the-art operational weather forecasts, was possible — even if here I used it to answer a limited and specific question. Another surprising finding was the rapidity on which the direct impact of CO₂ could be detected: my experiments suggest that if CO₂ were suddenly reduced to pre-industrial levels, it would detectably reduce the intensity of extreme heat events on timescales of days — i.e. well before the SSTs have had any time to respond. This finding translated into a significant increase in the likelihood of the exceptional temperatures observed over the UK in February 2019 attributable to increased CO₂ concentrations above pre-industrial levels alone. The final key takeaway from this study was that the predictability of the heatwave remained remarkably intact despite the hefty ‘kick’ we had given the model at initialisation in the perturbed CO₂ forecasts. This was vital for the reason described in more detail within the main sections of the chapter: it means we could ascribe our results to the impact on CO₂ on the heatwave, rather than to the chaotic nature of the weather system. If we had found that the predictability was less stable, we would have had to investigate further why this was the case, potentially returning to the drawing board for this approach. However, despite the promise that forecast-based approaches demonstrated within this study, the attribution I carried out was very limited in scope: to a single component of anthropogenic influence on the climate. In order to make results from the approach truly relevant to and of use for stakeholders I needed to determine how the methodology could be altered to allow more complete estimates of human influence on individual weather events to be made. This attempt at more complete attribution, through perturbing both the forecast boundary *and* initial conditions, is the focus of the next chapter of this thesis.

Quote

— author

4

Forecast-based attribution: perturbing the initial and boundary conditions

Chapter description.

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

Leach, N. J., Roberts, C. D., Heathcote, D., Mitchell, D. M., Thompson, V., Palmer, T. N., Weisheimer, A., & Allen, M. R. (2022). **Reliable heatwave attribution based on successful operational weather forecasts.** *Nature*, submitted.

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

Contents

4.1	Chapter open	63
4.2	Abstract	63
4.3	Introduction	64
4.4	The Pacific Northwest heatwave	66
4.5	Methods	68
4.6	Forecast-based attribution	74
4.6.1	Results	78
4.7	Discussion	80
4.8	Chapter close	81

4.1 Chapter open

Following on from the previous chapter, the primary aim was to try and use a forecast-based approach to provide a more complete estimate of human influence on an individual extreme. Based on previous work (67), the key component we needed to include was anthropogenic influence on the ocean — in addition to the changes to atmospheric CO₂ concentration. Although there is a considerable body of literature concerning how to perturb sea surface temperatures in climate models to include this effect (174), the model we used previously, and continued to use, is coupled, and therefore we needed a way to perturb not just the SSTs, but also the ocean subsurface. This technical challenge posed by the coupling also presented an opportunity, as it means that the resulting counterfactual forecasts are consistent with observed changes in ocean heat content (the primary energy sink for thermal energy accumulated in the earth system as a result of global warming), and do not contain any infinite sources or sinks of heat, as is the case in the prescribed-SST simulations used widely in attribution. We considered a number of different methods for imposing the ocean perturbations, including a nudged ocean spin-up simulation, but settled on arguably the simplest approach, which has been used previously by Chris Roberts at ECMWF for testing the impact of different ocean resolutions. At the time when we were planning these simulations, the 2021 Pacific Northwest Heatwave had just occurred, presenting a very natural case study for us to apply the forecast-based approach to, as conventional approaches had clearly been pushed to their limit with this event. The overall aim of our study was to demonstrate that we could provide a near-complete estimate of the human influence on this unprecedented event using a weather forecast model that was demonstrably able to simulate it.

4.2 Abstract

Extreme weather attribution, quantifying the role of human influence in specific weather events, is of interest to scientists, adaptation planners and the general public (34). However, the devastating 2021 Pacific Northwest heatwave challenged conventional statistical approaches to attribution due to the absence of similar events in the historical record, and model-based approaches due to poor representation of key causal processes in current climate models (188). Here we use state-of-the-art operational medium-range and seasonal weather prediction systems, applied for the first time to this kind of climate question and unequivocally able to simulate the detailed physics of the heatwave in question, to show that human influence on the climate made this event at least 8 [2–30] times more likely to occur. Quantifying the absolute probability of such an unprecedented event is more challenging, but the length of the observational record suggests at least a multi-decade return-time in the current climate, with the likelihood doubling every

17 [10–50] years at the current rate of global warming. Our forecast-based approach synthesises the storyline approach, which examines human influence on the physical drivers of an event in a deterministic manner (78), and the probabilistic approach, which assesses how the frequency of a class of events has been affected by human influence (25). If developed as a routine service in a number of forecasting centres, it could provide reliable estimates of the changing probabilities of all extreme events that can be represented in forecast models, which is critical to supporting effective adaptation planning (64, 189).

4.3 Introduction

Although considerable progress has been made over the past decade in quantifying the impact of climate change on individual extreme weather events (25, 34), challenges remain over the assessment of the most extreme events. Such events are particularly difficult to draw confident conclusions about due to the lack of historical analogues, and their often poor representation in the climate models normally used for event attribution. Two contrasting mainstream frameworks to event attribution have been developed: the storyline approach, which examines anthropogenic influence on the causal drivers of the extreme in question and is therefore highly conditioned on its specific characteristics (65, 71, 78); and the probabilistic approach, which aims to determine how anthropogenic influence has affected the likelihood of events at least as extreme as the one in question (61, 67).

A key challenge for extreme event attribution is that we cannot make direct observations of a world without human influence on the climate, so all approaches must involve some kind of modelling, either statistical (66) or dynamical (67). Both face difficulties with the most extreme events, especially when considering the nonlinear processes that often drive unprecedented events. Statistical models used in conventional attribution can break down when faced with such events due to the lack of appropriately similar historical samples (190), while numerical climate models generally used in attribution studies are typically coarse ($O(100 \text{ km})$ horizontal resolution), and poorly represent important processes involved in the development of extreme weather events, such as blocking (79) and atmospheric rivers (191). Even with a ‘perfect’ model of the earth system, the unconditioned nature of the vast majority of climate model simulations used in attribution means that obtaining enough analogues of unprecedented events (192) to avoid the same issue faced by statistical modelling of the observational record requires very large ensembles, possibly beyond current computational limits (193). Crucially, the role of climate change in an individual event may differ from that in other events of the same class due to the specific physical processes behind it (75, 76).

The storyline framework overcomes some of these issues, and the risk of a false

negative, by examining the impact of climate change on the causal drivers of an event deterministically. For instance, one might separate out the thermodynamic (typically high confidence in response to climate change and well-represented in numerical models) and dynamic (typically much lower confidence in response to climate change, and more poorly represented in numerical models) drivers of an event, for example by conditioning on the concurrent large scale atmospheric circulation. One approach for applying such conditioning is to ‘nudge’ climate model simulations towards the large-scale flow observed during a particular extreme (81, 82). The storyline approach does not, however, provide quantitative information about how climate change has affected the probability of the event in question, which is of interest to the general public and relevant to policymakers for adaptation planning.

We propose a forecast-based approach that could synthesise the probabilistic and storyline frameworks to extreme event attribution (109). Although they belong to the same class of dynamical model and often share components (194), operational weather forecast models are typically run at much higher resolutions than climate models, improving their overall physical representation of extremes. They are validated for producing predictions that span the range of possible weather by the centres that produce them to a much higher degree than climate models, where other aspects are more important. In addition to this high level of explicit validation, using a model that has successfully predicted an event ensures that the model is able to accurately represent all the processes involved in the event in question, increasing the reliability of attribution statements based upon it (76). Stepping back through lead times allows for a robust storyline-like framing by examining how climate change has affected the causal drivers of the specific event within the limits of predictability. Probabilistic attribution can be performed using a reliable forecast ensemble, with the level of conditioning set by the lead time - the limiting case of long lead times is equivalent to a conventional unconditioned analysis. There has been some previous work into forecast-based attribution, using seasonal forecast models (155–157, 195, 196) and exploring the conceptual framework (185, 197, 198). To our knowledge, however, this study is the first time that a complete forecast-based attribution has been carried out in a coupled operational forecast model at such a high resolution.

In this study we use the coupled operational ECMWF model (details in the Supplement) to analyse the Pacific Northwest heatwave, taking advantage of its successful predictions of this unprecedented event at leads of over a week. We perform counterfactual forecasts of the event by perturbing the initial and boundary conditions of the model in order to simulate how the heatwave might have emerged had it occurred in a cooler pre-industrial world, or a warmer future world. We then compare the counterfactual and operational forecasts to assess the impact of anthropogenic climate change on both the magnitude and probability-of-occurrence of the event. We believe that this forecast-based approach opens the door to not only a reliable and practical operational attribution system, but

also to a robust way of generating projections of future weather explicitly referenced to the forecasts used already by adaptation planners ([173](#)).

4.4 The Pacific Northwest heatwave

At the end of June 2021, a large fraction of the Pacific Northwest region of the US and Canada experienced unprecedented high temperatures, including the cities of Portland, Salem, Seattle and Vancouver (Figure 4.1). This heatwave (the ‘PNW heatwave’) has been directly linked to many hundred excess deaths during and following it, making it the deadliest weather event on record for both Canada and Washington state ([199](#)). The heatwave peak was observed between the 28th & 30th June, though temperatures were still exceptionally high on the days immediately before and after this period ([200](#), [201](#)). Many local maximum temperature records were broken during this period, including the Canadian all-time record by a margin of 4.6 °C.

Based on current understanding, the heatwave arose from an optimal combination of proximal drivers ([188](#), [202–204](#)). Development of an omega block between the 23rd-27th coincided with the landfall of an atmospheric river (AR) on the 25th. Warm air was drawn up from the tropical West Pacific, heated diabatically through condensation in the river and then further heated adiabatically through subsidence: both the temperature and lapse rate at 500 hPa reached or approached record levels in the regions affected. This atmospheric heating was enhanced by soil moisture feedbacks ([205](#)) and high insolation at the land surface during the hottest hours of the day (Figure 4.4). Given the unprecedented nature of the observed heatwave, any dynamical numerical model would need to capture all these processes, including the coupling between them, in order to produce an accurate representation of the event.

Despite the observed temperatures lying far outside the historical record, the heatwave was well predicted by numerical weather forecast models such as from ECMWF at lead times of more than a week. The seasonal forecast from ECMWF captured one important aspect of the event: it predicted a thicker troposphere than average (measured by 500 hPa geopotential height) over the Pacific Northwest during the summer. A key change in the predictability of the exceptional temperatures occurred around June 21st, being the earliest point at which the penetration of the AR over land was well represented ([204](#)). The success of these forecast models provides an opportunity to use them to examine the influence of anthropogenic climate change on the event as it actually occurred.

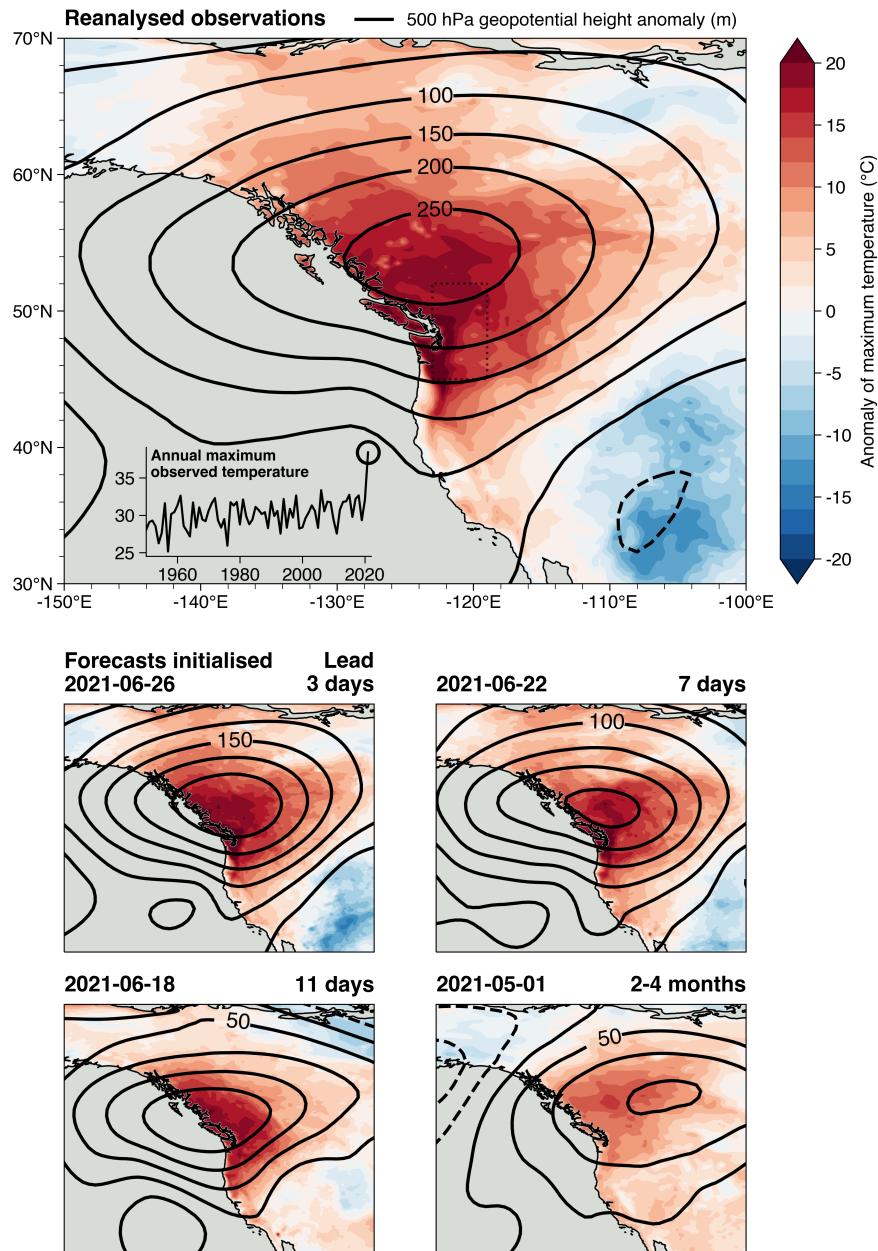


Figure 4.1: Features and forecasts of the Pacific Northwest heatwave. Top panel: Surface temperature anomalies at the time of the peak heat during the heatwave within the region enclosed by 45–52 N, 119–123 W (indicated by the dotted rectangle). Solid black contours show the 500 hPa geopotential height anomaly averaged over 26–30th June 2021. Data are from ERA5 reanalysis (86). **Inset:** timeseries of annual maximum temperatures for the same dotted region. **Bottom panels:** As above, but taken from the ensemble member within the forecast initialised on the date given above each panel that predicted the nearest temperature to the reanalysis within the dotted region.

4.5 Methods

Event definition

How the extreme event of interest is quantified - the event definition - is a key methodological decision that must be made in extreme event attribution studies. A significant amount of previous work has shown the impact of the event definition on the quantitative outcome of the analysis (148, 164, 206). In this study we use a definition consistent with a previous attribution study of the PNW heatwave (207) to allow for a comparison between our forecast-based approach and their probabilistic statistical and climate-model based approach.

We first average maximum temperatures over the region enclosed by 45–52 N, 119–123 W (indicated by the dotted rectangles in Figures 4.1 & 4.4). For the event as observed in the ERA5 reanalysis (86) we then take the peak temperature recorded during the heatwave, which occurred at 00 UTC on 2021-06-29. For the event as simulated in the medium-range forecast ensemble members, we take the peak temperature that occurred between the 26-30th June, the period over which the heatwave occurred in reality. For the event as simulated in the seasonal forecast ensemble members, which we would not expect to predict the precise timing of the heatwave, we take the peak temperature over the full summer season. The differences between the event definitions of the medium-range and seasonal cases lead to the discrepancies in the climatologies shown in Figure 4.5.

Experiment details

Model details The medium-range experiments we have performed use the version of the IFS EPS that was operational at the time of the PNW heatwave, CY47R2 (208). The forecast model atmosphere is run at a resolution of O640 (\sim 18 km) and has 137 vertical levels. The atmosphere is coupled to a 0.25 degree wave model (209), 0.25 degree sea ice model (210), LIM2, and 0.25 degree ocean model (211), NEMO v3.4, with 75 vertical levels (ORCA025Z75 configuration). We maintain the same number of ensemble members as the operational system, 51, throughout our experiments.

The seasonal experiments are performed with ECMWF's operational seasonal forecasting system, SEAS5 (212). This uses IFS CY43R1 (213) at a horizontal resolution of Tco319 (\sim 36 km) with 91 vertical levels. The seasonal configuration of IFS CY43R1 is coupled to a 0.5 degree wave model, LIM2, and NEMO v3.4 in the ORCA025Z75 configuration. We maintain the same number of ensemble members as the operational system, 51, throughout our experiments.

Simulation setup Our experiments all use the exact operational setup (model configuration and initial conditions) as their base. To this setup, we:

1. Change the CO₂ concentrations used to a ‘pre-industrial’ level of 285 ppm, and a ‘future’ level of 615 ppm. These represent the same fractional change in opposite directions from the present-day concentration of 420 ppm used in the operational forecast system.
2. Subtract (for the pre-industrial forecast) or add (for the future forecast) a perturbation of the estimated anthropogenic influence on the ocean state since the pre-industrial period from the initial conditions of the forecasts (through the ocean restart files). The estimation of this perturbation is described below. We use estimated perturbations for 3D temperature, sea ice concentration, and sea ice thickness.
3. Check the sea ice fields for unphysical values. In the perturbed restarts, we ensure that sea ice concentration does not exceed 1 or subceed 0. We ensure that sea ice thickness does not subceed 0. Values outside these bounds are set to their nearest bound. Finally, we set sea ice thickness to 0 where sea ice concentration is 0, and vice versa.
4. Modify ocean salinity such that in-situ ocean density is preserved following the 3D temperature perturbation as calculated using the equation of state from the forecast ocean model. The salinity compensation is achieved to machine precision using a simple gradient descent algorithm. The resulting coupled forecasts are thermodynamically consistent with the imposed ocean heat content anomalies without any adjustments to the initial ocean circulation, mixed layer depths, or horizontal pressure gradients. Importantly, and unlike uncoupled forecasts constrained by specified sea-surface temperatures, there are no infinite sources or sinks of heat in the resulting counterfactual forecasts. This approach is justifiable in shorter-range forecasts as there is no direct influence of salinity on the overlying atmosphere. This assumption may eventually break down at lead times comparable to ocean advective processes, for which there could be indirect feedbacks on the atmosphere associated with salinity-driven changes in the ocean state. Nevertheless, this approach works well for the medium-range and seasonal forecasts described in this study.

The perturbations used are computed using an optimal fingerprint analysis (5, 6, 140). We first calculate the Anthropogenic Warming Index (AWI) using anthropogenic and natural radiative forcings from AR6 (214) and the HadCRUT5 global mean surface temperature dataset (141). The AWI provides us with a plausible estimate of the

fingerprint of anthropogenic influence on other climate variables (5). For each perturbed variable, we then regress observed timeseries at each gridpoint onto the AWI, using the following data sources:

- Sea ice thickness: ORAS5 (1958:2019) ([215](#))
- Sea ice concentration: ORAS5 (1958:2019) ([215](#))
- Sea surface temperature: HadISSTv1.1 (1870-2019) ([130](#))
- Subsurface temperature: WOA18 (1950-2017) ([216](#))

We then scale the computed regression coefficients at each point by the change in AWI between the pre-industrial period of 1850-1900 and 2019 to produce our final estimated perturbations. The sea surface, and zonally and globally averaged temperature profiles are shown in Figure 4.2.

Finally, we combine the sea surface and subsurface temperature perturbations. We did not use a subsurface temperature dataset in isolation since observations of the sea surface temperature are considerably more abundant in the early 20th century than observations of subsurface temperatures, and since the temperatures at and near the surface are likely to be the most important for the medium-range forecasts performed, we leveraged the additional information contained in observed sea surface temperatures. We combine the two by relaxing the sea surface perturbation towards the subsurface perturbation using a relaxation depth scale of 60 m (the surface autocorrelation scale in WOA18).

We note that estimation of the perturbation, and in particular the subsurface temperatures, is associated with considerable uncertainty due to the lack of observations in the pre-ARGO era. Here we have used a single best-estimate perturbation due to constraints on the available computational resource, but to account for this uncertainty an ensemble of perturbations could be applied ([171](#)). A possible way in which such an ensemble could be derived would be to apply optimal fingerprinting to an ensemble of coupled climate models.

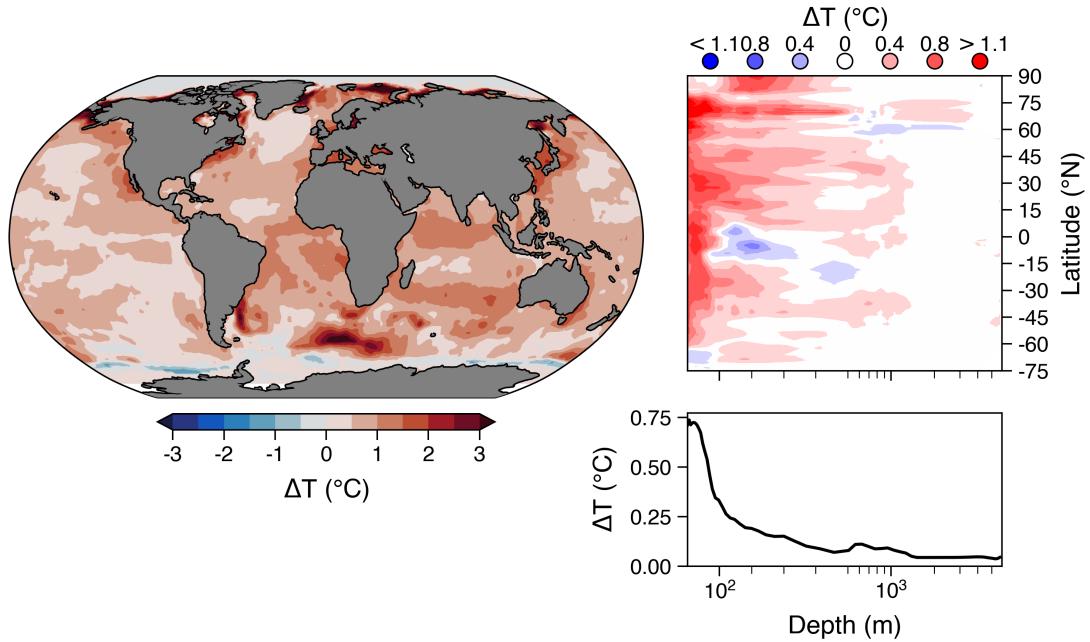


Figure 4.2: The initial ocean state perturbation applied. **Left panel:** map of the surface temperature perturbation. **Inset:** timeseries of annual maximum temperatures for the same dotted region. **Top right panel:** map of zonally averaged temperature perturbations as a function of depth. **Bottom right panel:** globally averaged temperature perturbation as a function of depth. Note that the x-axis switches from a linear to logarithmic scale at a depth of 500 m.

Bias correction of seasonal forecast ensembles

Climate drift can be an issue in the use of coupled seasonal forecast models (217). We find a non-negligible drift in the daily maximum temperature SEAS5 forecast ensemble initialised in May over the PNW region. This drift results in a positive temperature bias that grows with lead time. Hence, using the raw model output in our analysis would overestimate the probability of the PNW heatwave.

To account for the drift, we perform a simple bias-correction procedure on the seasonal forecast ensembles, informed by comparing the SEAS5 hindcasts over 1981-2020 with ERA5 reanalysis data over 1950-2020 (using the full time period that data is available and excluding the year of the event, 2021). We do this in three steps:

1. Remove the attributable forced trend from both the reanalysis and hindcasts by regressing mean JJA daily maximum temperatures onto the AWI (5).
2. Remove the drift from these detrended hindcasts, estimated by averaging the hindcasts for each lead time over all years and ensemble members, subtracting this from the corresponding reanalysis average over all years, and then regressing this timeseries onto the (linear) lead times, producing a linearly lead-time dependent drift correction (217).
3. The drift-corrected hindcasts still exhibit a positive bias during periods of extreme high temperatures. Hence, we finally remove the remaining mean bias in annual maximum temperatures in the hindcasts compared to reanalysis.

We apply this bias correction procedure to both the seasonal hindcasts shown in Figure 4.5 and used to estimate the return time of the event, and to the operational and perturbed seasonal forecasts of the 2021 summer. Figure 4.3 shows the results of this bias correction procedure, following Thompson et al. (218).

We note that validation of the bias correction procedure on the SEAS5 distribution of annual maximum temperatures (TXx) is challenging due to the unprecedented nature of the 2021 event. If we perform an analysis of the higher-order moments of the SEAS5 and ‘observed’ (ERA5 reanalysis over 1950-2020) distributions of TXx (218), we find that the bias-corrected ensemble tends to have larger values of higher-order moments than the observed timeseries. However, if the 2021 event is included in the observed distribution, then the opposite is found, due to the large impact of such an outlier on these moments. This sensitivity to inclusion / exclusion of the 2021 event, demonstrated in Figure 4.3, is why we have opted to perform a simple but physically motivated bias correction rather than a more complex statistical correction such as a quantile map.

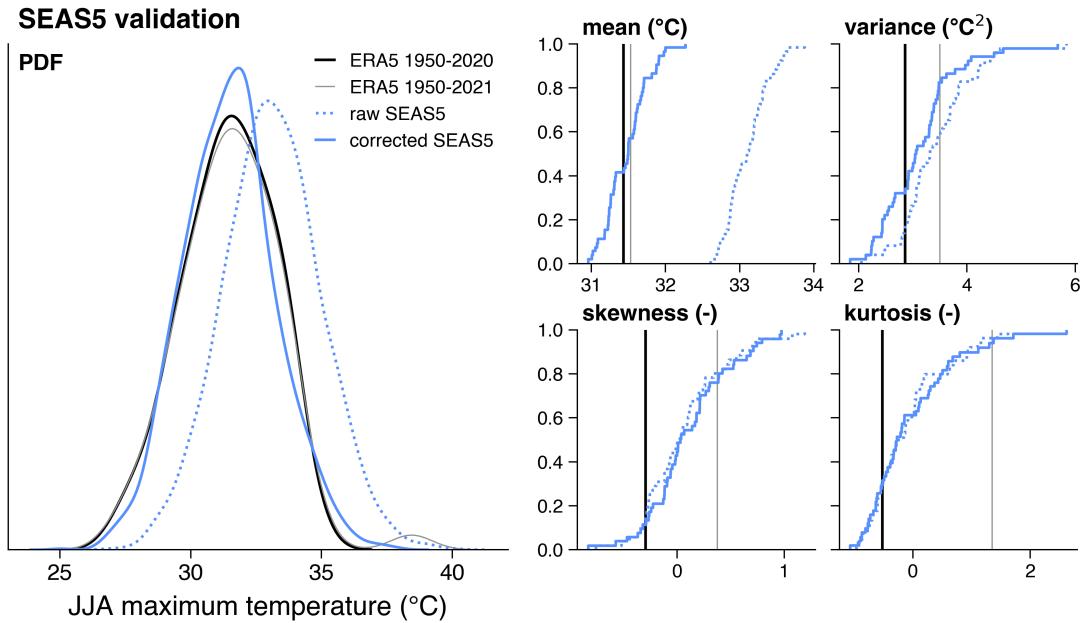


Figure 4.3: Validation of the bias correction applied to the SEAS5 seasonal forecast simulations, following Thompson et al. (218) Figure 2. **Left panel:** PDFs of summer maximum temperatures in detrended reanalysis, and raw and bias-corrected seasonal hindcasts. **Right panels:** CDFs of detrended proxy raw and bias-corrected seasonal hindcast timeseries mean, variance, skewness and kurtosis compared to reanalysis values. In all panels, we show reanalysis both including and excluding the 2021 PNW heatwave.

Statistical methodology

Intensity changes We calculate changes in intensity as the difference between the average of the nearest quintile of each ensemble to the event (in terms of peak temperatures). For the three longer leads, this is effectively the difference between the averages of the uppermost quintile of the two ensembles.

Risk changes We calculate the risk ratio by first fitting either a generalised extreme value (GEV) distribution to the full operational ensemble (for the shortest lead) or a straight line on a return-time diagram (i.e. an exponential tail) to the nearest quintile of either the operational ensemble (for the other two medium-range leads) or the model climatology (for the seasonal lead, since the tail of the operational ensemble lies considerably further below the event threshold than the tail of the much larger model climatology). We do this because while the shortest lead ensemble is well represented by a GEV distribution, the other three are not, and have generally heavier tails than estimated by likelihood-maximising GEV distributions. In these cases, where the event threshold lies in the extreme tail of the ensemble, the tail properties of the approximating distribution project considerably onto the estimated probability of the event. Hence, to avoid any undue assumptions on the tail shape, we fit a straight line on a return-time diagram such as Figure 4.5 (assuming an exponential tail) to the nearest quintile.

After fitting an appropriate distribution, we then shift the location of this distribution by the estimated attributable warming. We then calculate the probability of observing an event at least as intense as the PNW heatwave (the dashed line in Figure 4.5) in the original distribution and the shifted distribution. The risk ratio is the ratio of these two probabilities ($P_{\text{current}}/P_{\text{shifted}}$).

Throughout this chapter, confidence intervals are calculated using a 10,000 member non-parametric bootstrap with replacement.

4.6 Forecast-based attribution

The date at which we initialise our perturbed forecasts is a key choice that allows us to condition our attribution analysis on different synoptic drivers of the heatwave, which become predictable at different leads (203, 204). The climate change response of drivers already present in the initial conditions is clearly not incorporated into our attribution results for each lead time due to this conditioning. Starting with the operational configurations of the ECMWF forecast model, we chose to focus on three medium-range and one seasonal forecast lead: 3 days, 7 days, 11 days and 2-4 months. These leads highlight the following aspects of the attribution:

- 3 days (2021-06-26): a forecast very highly conditioned on the synoptic drivers of the event, with several key drivers prescribed in the initial conditions, and the rest forecast near perfectly. At this lead, our experiments could be considered analogous to a storyline attribution framing.
- 7 days (2021-06-22): a highly conditioned forecast, with most simulated processes mirroring reality closely. However, the shape and gradient reversal magnitude of the block shows considerable variation in this ensemble.
- 11 days (2021-06-18, depicted in Figure 4.4): while the exceptional thickness of the tropospheric block was well predicted in a large proportion of the ensemble, the shape and associated gradient reversal was only captured in a few members. The occurrence of the AR was well predicted, but its location and penetration over land less so, with most members predicting a more southerly landfall. The low soil moisture and cloud cover was well captured by the majority of the ensemble.
- 2–4 months (2021-05-01): a considerably less conditioned forecast. For this lead, we take the peak heat event over the whole summer period since we do not expect the forecast to predict the timing of the heatwave. Although the forecasts were unusually successful at predicting elevated geopotential height and temperatures over the summer in general, none of the peak heat events within individual ensemble members capture all of the detailed features of the PNW heatwave. At this lead, the ensemble can be viewed as being near-analogous to a high resolution unconditioned climate model simulation (though one that we know is able to represent the processes involved in the PNW heatwave accurately).

We then perturb the boundary and initial conditions of the operational forecast (technical details in the Supplement). First, we perturb the CO₂ concentrations in the atmosphere back to pre-industrial levels of 285 ppm, similar to (109). Then we remove a balanced estimate of anthropogenic change between pre-industrial and the present-day in surface and sub-surface ocean temperatures, sea ice concentration, and sea ice thickness (130, 215, 216) from the initial state of the model. Perturbing the temperatures over the entire ocean depth means that we produce forecasts that are thermodynamically consistent with the changes in upper ocean heat content, in contrast to prescribed SST approaches (129, 139). We do not alter the land-surface, noting the high uncertainties in past trends for indicators such as soil moisture in this region (219–221). Removing anthropogenic influence from the ocean state and reducing CO₂ levels produces a counterfactual ‘pre-industrial’ forecast; we also apply identical perturbations in the opposite direction to produce a ‘future’ forecast, in which the ocean state and CO₂ levels of 615 ppm correspond to approximately twice the level of global warming experienced at the present-day.

We find that despite the large impulse applied by the perturbed initial state upon forecast initialisation, the predictability of the heatwave is remarkably stable. The key synoptic drivers of the heatwave present in the original operational forecast remain intact. There are some changes consistent with the canonical response to global warming, including a thickening of the lower troposphere (17) and increased tropospheric water vapour (110) in the future forecast; and vice-versa in the pre-industrial forecast. As such, the perturbations have not altered the forecasts in such a way that they produce ‘different’ weather, and we can compare our forecasts to estimate the influence of anthropogenic global warming on the Pacific Northwest heatwave. This is consistent with (109), but is not guaranteed to be the case for every weather event.

This experiment design is consistent with the perturbed CO₂ experiments of (109) in another important respect: the adjustment to the new ‘pre-industrial’ or ‘future’ climate state occurs continually throughout the forecast. This adjustment typically means that as the lead time increases, the estimated attributable influence on the heatwave also increases. Interplay between dynamical noise and attributable signal in the forecasts, both of which increase with the lead time (short leads correspond to more confident but smaller attributable impacts and vice-versa) is discussed further in (109). The adjustment means that any attributable impacts estimated directly from the forecasts are lower-bounds on the true anthropogenic impact. However, we find that attributable impacts on the heatwave are approximately linear with the coincidental global warming level within the perturbed forecasts across the range of leads explored (Supplementary Figure S7), consistent with (222). Hence, in addition to the impacts estimated directly from the perturbed forecasts, we also present impacts scaled to the global warming level within the forecast at the time of the heatwave.

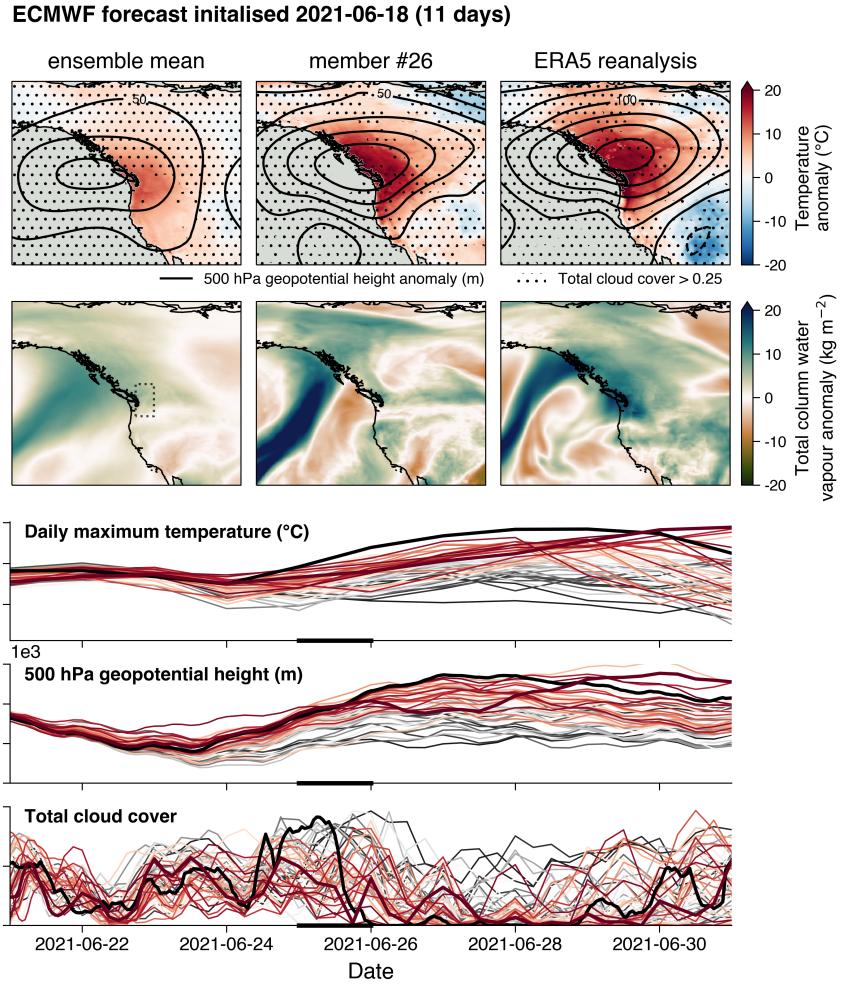


Figure 4.4: Drivers of the PNW heatwave and their predictability in the forecast initialised 2021-06-18 (11 days). **Top row:** temperature anomaly fields for the PNW heatwave in the ensemble mean, nearest member and reanalysis. Solid black contours indicate 500 hPa geopotential height anomalies and stippling indicates regions with total cloud cover greater than 25%. **Second row:** mean total column water vapour anomalies on the 25th June. The study region of 45–52 N, 119–123 W, over which fields are aggregated into timeseries, is indicated by the dotted rectangle. Anomalies shown are calculated relative to the 2001–2020 period. **Bottom three rows:** timeseries of daily maximum temperatures, total column water vapour and total cloud cover in each forecast ensemble member. The solid black line shows the reanalysis timeseries and the thick solid line shows the nearest member. The colour of each line indicates the rank of that ensemble member in terms of the peak temperature simulated during the heatwave period (dark grey = coolest, dark red = warmest). The solid black bar on the time axis of each panel indicates the averaging period used for the total column water vapour maps.

4.6.1 Results

The results of our forecast-based approach can be presented either as the attributable human influence on the intensity of the heatwave or the probability of the heatwave. We find that the intensity of the heatwave is reduced in the pre-industrial forecasts for all lead times (Figure 4.5). Due to the continual adjustment of the forecasts to the initial condition perturbations, the attributable influence on the heatwave peak temperature, estimated as half the difference between the pre-industrial and future forecasts to maximise the signal-to-noise ratio, increases as the lead time increases, ranging from 0.28 °C [0.25–0.33][†] using the 3-day lead to 0.7 °C [0.35–1.0] using the seasonal forecast. We account for the continual adjustment of the perturbed forecasts by scaling the attributable influence by the ratio of the coinciding global land warming level to the observed present-day level of 1.6 °C (223). This results in a best-estimate attributable impact on the heatwave intensity of 1.3 °C [0.5–1.9] for a current level of anthropogenic warming of 1.25 °C (140). This accounts for approximately 20% of the 7 °C 2021 anomaly over previous annual maxima.

We quantify the attributable change in probability due to anthropogenic global warming using relative risk (35), estimating the probability of observing an extreme at least as extreme as the observed 2021 heatwave using an appropriate extreme-value or tail distribution, and then shifting this distribution by the attributable change in intensity for each lead time. As with the heatwave intensity, the relative risk tends to increase with forecast lead time due to the adjustment to the initial conditions. Our results are consistent with a linear relationship between log probabilities and the coinciding global land warming level. If we account for this adjustment by scaling log probabilities by the current global land warming level of 1.6 °C, we find a best-estimate relative risk of a factor of 8 times [2–30] considering all lead times, or analogously a fraction of attributable risk of 0.9 [0.5–0.97].

Using the current rate of global warming over land (140) we can further estimate that the probability of observing an event at least as warm as the 2021 Pacific Northwest heatwave is doubling every 17 [10–50] years, and will continue to do so unless the rate of global warming decreases. Given the length of the historical record and our estimated change in probability over this period, such an event would be associated with a multi-decade to multi-century return period at the present-day, thus making this doubling time very relevant for adaptation planning.

[†]Numbers in square brackets [] represent a likely confidence interval (17–83%) throughout this chapter.

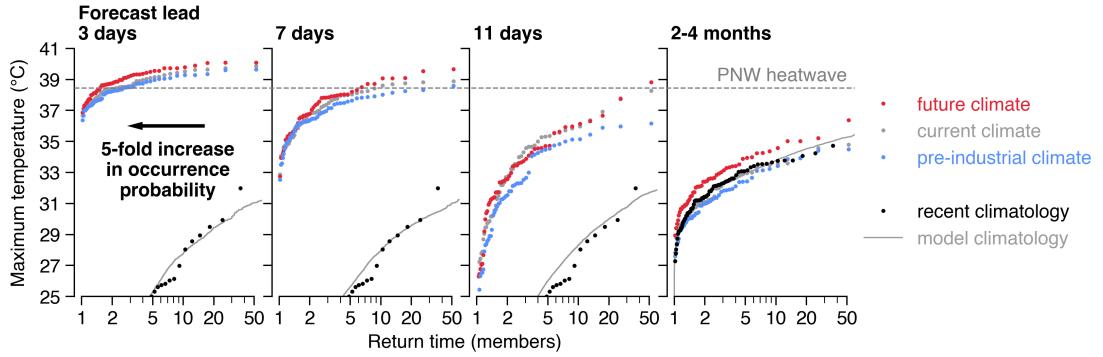


Figure 4.5: Return-time diagram of the PNW heatwave in the operational and counterfactual forecast ensembles. Each panel shows ensembles initialised at the lead given above the panel. Red, grey and blue dots indicate empirical return-time plots based on the ensemble members of the future, current and pre-industrial forecasts. The dashed grey line shows the temperature threshold observed during the PNW heatwave. The black dots indicate the recent climatology, based on detrended ERA5 reanalysis over 1950–2020. The solid grey line indicates the model climatology estimated using detrended hindcasts over 2001–2020 for the medium-range forecast, and using detrended and bias-corrected hindcasts over 1981–2020 for the seasonal forecast. The arrow in the left panel indicates, for illustration, the displacement along the log-scaled x-axis equivalent to a 5-fold increase in occurrence probability.

4.7 Discussion

The results presented here provide strong evidence of the impact of climate change on a specific extreme event, based on a model that has been demonstrated unequivocally to be able to simulate the event in question through a successful medium-range forecast. Our estimates of relative risk are lower than previous climate model-based estimates (207), albeit are not entirely incompatible within the context of the associated uncertainties and the fact that our estimates represent a lower bound on the impact of climate change on the heatwave (as was the case in 109). The primary reason is that our model (unlike a typical climate model) is capable of simulating the multiple physical factors that contributed to the heatwave that occurred, so we are not relying on extrapolation of distributions from physically dissimilar events. Moreover, our imposed perturbations do not include the total sum of human influence on the climate. It is known that land surface feedbacks are important in the development of extreme heatwaves (58), and is plausible that if we had removed the influence of anthropogenic climate change from the initial land state in addition to the ocean state, the resulting attribution statement might have been stronger.

Nevertheless, we argue that the forecast-based methodology presented here represents an important advance in both attribution in general, and operational attribution. Rather than relying on multiple lines of evidence that would each be unsatisfactory in isolation, here we have presented a single adequate line. The key to the adequacy of the result is the ability of the model used to represent the event in question, demonstrated through successful prediction. This not only means that we have increased confidence in the model's response to external forcing (75, 76), but also that the analysis is a genuine attribution of the specific event that occurred (rather than a mixture of events that share some characteristic like extreme temperatures, but differ in other important meteorological aspects). Forecast-based attribution provides many of the advantages of the storyline approach to attribution, but can still be used to provide quantitative estimates of the changing probability of extreme events with climate change. The use of an operational weather forecast model demonstrates how this approach could be easily adapted to provide an operational system for attribution in real-time (or potentially even in advance, 195). Such a system would involve re-running operational forecasts with perturbed initial and boundary conditions as in the counterfactual forecasts we have presented here (44).

There remain a number of ways in which the forecast-based approach explored here could be further developed. Firstly, analysis of the forecasts would be simplified if they were started from balanced states, rather than continually adjusting to the new initial conditions throughout the forecast. This could be done by either including additional perturbations to the initial conditions (ie. to the land-surface and atmospheric states, 44, 195, 224), or possibly by perturbing the initial state using the operational data-assimilation itself. Secondly, while here we have chosen to use the exact setup used operationally by

ECMWF, the uncertainty of forecast-based attribution statements could be reduced by increasing the ensemble size (we note that 51 members is a relatively small ensemble in the context of traditional attribution-specific experiments, 129, 139), particularly for the longer, relatively less-conditioned lead times.

The focus of this study was on the attribution question, but this forecast-based methodology could be applied to produce projections designed to inform climate change adaptation. Analogous to our ‘future’ counterfactual forecast, which we used here check the linearity of the climate change response, perturbations consistent with specific levels of global warming could be applied in order to, for example, simulate specific extreme events as if they occurred in a world of 2 °C. Such simulations of potential future extremes could be used to test the limits of regional adaptation in a targeted manner based on impactful events that have already occurred (173), complementing other approaches such as (193), which was designed to produce a rich set of different extreme events rather than specific ‘grey-swan’ type events.

Concluding remarks In this study, we have used a numerical weather forecast-based approach to determine the contribution of human influence to a specific unprecedented extreme event. We used a state-of-the-art coupled operational weather forecast model that was unequivocally able to simulate the event in question, demonstrated by a successful prediction. Our perturbed initial condition approach maintains consistency with the measured changes in upper ocean heat content, unlike many previous approaches. We view this forecast-based approach as synthesising the storyline and probabilistic approaches to event attribution, keeping the event specificity of the storyline approach while still providing meaningful estimates of the changing risk of the extreme in question. Given that it is increasingly clear that we need to go beyond the meteorology of event attribution, and into the societal impacts (189, 225), we suggest that our approach would be particularly well-placed to advance this agenda, especially in the context of extremes in a future climate.

4.8 Chapter close

In this chapter, the key takeaway is that we have been able to use a relatively simple methodology to produce counterfactual forecasts of an individual extreme weather, which we can then use to estimate the anthropogenic component of that event. This methodology is very similar to that used by Pall et al. (67), and is based on perturbing the forecast initial conditions, but has been adapted to allow coupled weather forecast models to be used. There remain a number of outstanding issues with the approach we have taken, most notably whether we should be — and how we should go about — perturbing the

initial atmospheric and land-surface states too. In terms of remaining scientific questions, a particularly interesting avenue would be to explore how atmospheric predictability changes in counterfactual forecasts. We have been surprised at how consistent the predictability of our case study extreme events is, even when ‘kicking’ the model as hard as we have done. It would be very interesting, and important for forecast-based approaches in general, to determine if there are particular situations where predictability breaks down in the counterfactual worlds. Unfortunately, I cannot address these questions within the scope of this thesis, but I discuss them further, including how they might be addressed in the future, in the Discussion chapter 6. I am, however, happy to have taken our forecast-based approach as far as the stage presented in this chapter!

Although the clear immediate application of the approach described in this chapter would be an operational attribution service, I am particularly interested in exploring how the same approach could be used for climate projection of extreme events. There are additional uncertainties associated with projection, including the socio-economic pathway taken, and the pattern of ocean warming, but such a forecast-based approach to projection would still confer many of the advantages I have argued that it does for attribution (e.g. 173). This shift in focus to climate projection fits in neatly with the following chapter 5, in which I explore a novel methodology for generating a rich variety of samples of future extreme weather that could be used for adaptation planning. I suggest that climate projection using counterfactual forecasts could complement the approach taken in chapter 5 by providing understanding and information about specific damaging extremes. This would be a natural direction for future work; as I discuss further in the Discussion 6.

Quote

— author

5

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 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>

*with the author contributing as follows. Data curation, Formal analysis, Investigation, Methodology, Visualisation, Writing – original draft and Writing — Review & Editing.

Contents

5.1	Chapter open	85
5.2	Abstract	85
5.3	Introduction	86
5.4	Study design & methods	88
5.4.1	Models	88
5.4.2	ExSamples experiment design	89
5.4.3	Statistical methods	95
5.5	Results	97
5.5.1	Comparison of HadAM4 and HadGEM3-GC3.05 baseline ensembles	97
5.5.2	Projections of future extremes	100
5.5.3	Sampling record-shattering subseasonal events	107
5.6	Discussion	109
5.7	Concluding remarks	111
5.8	Chapter close	112

5.1 Chapter open

This chapter arose somewhat by chance from a collaborative project between the *climatedprediction.net* team at Oxford and Bristol and the Met Office exploring the uncertainty surrounding winter extreme events. This project was looking for someone to carry out the formal analysis, and it so happened that at the time I was studying the February 2019 heat event, as described in chapter 3. Their interest in winter extremes was aligned with my own, and so I became involved. The aim of the project was to explore questions regarding the uncertainty surrounding the most extreme winters found within the UKCP18 coupled global model simulations: could they have been more extreme?; could a less complex atmosphere-only model produce similar extremes?; were the lower boundary forcings particularly conducive to producing extremes?. Although this work is a little detached from the rest of my thesis that concerns forecast-based approaches to attribution, I think that it provides a highly complementary way of understanding extremes. While extreme weather attribution is very specific and concerns individual events, the work here is pretty much the opposite in that it concerns understanding the full space of possible ways in which such extreme events can occur, and how best to sample from that space. Both approaches have clear utility for adaptation planning.

5.2 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 60 km 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.

5.3 Introduction

Weather extremes are one of the most damaging hazards that society faces at the present-day ([226](#)). 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 ([227](#)). 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 ([28](#), [142](#), [228](#)). 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 ([229](#), [230](#)) 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 *climatedprediction.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 (218, 231). 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 centres.

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 analogue 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 (190, 192). 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.

5.4 Study design & methods

5.4.1 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 (232). This PPE is based on the global HadGEM3-GC3.05 coupled ocean atmosphere model (230, 233). 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 (234–237). 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 (230), minus two members that displayed a significant weakening of the Atlantic Meridional Overturning Circulation (238). 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 (239, 240). Like its predecessor, HadAM3 (241), 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 (242, 243), 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 *climatedprediction.net* distributed computing system (244–246). This system has been used previously to run a range of Hadley Center Unified Model variants (84), including a coupled atmosphere-slab ocean model (247), a fully coupled model (248) and an atmosphere-only model (67) 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’.

5.4.2 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: 12 km regional and 2.2 km convection-permitting model projections (230, 249).
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 (244, 247), 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 ([250](#)).

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 ([251](#)). 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 (i.e. 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 5.1, there is one clear candidate for each type of extreme: UKCP18 PPE member 02868 (ID numbers as 236) year 2066 as a hot winter; and member 02242 year 2068 for the wet winter. The next most extreme hot winters shown in Figure 5.1A 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.

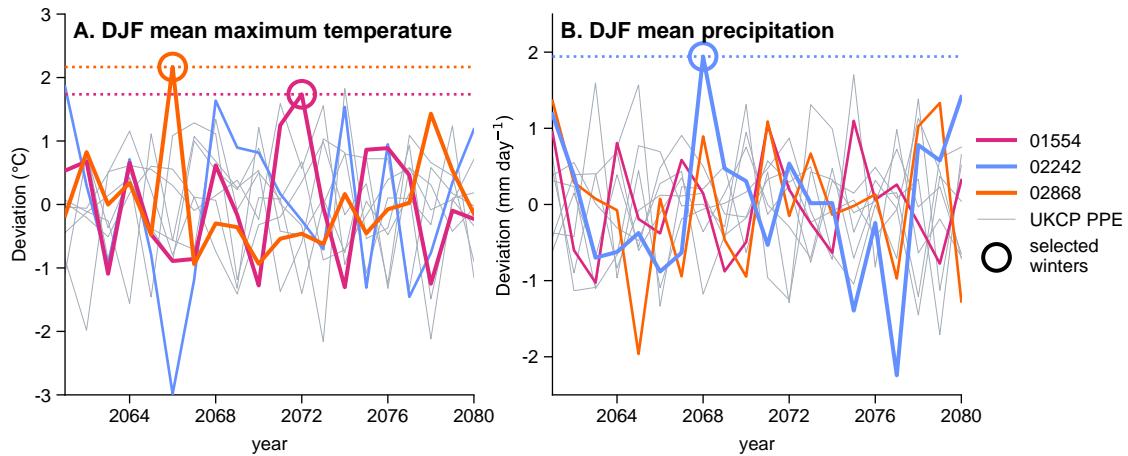


Figure 5.1: UKCP PPE 2061–2080 deviations from forced response. **A**, DJF mean of daily maximum temperatures averaged over the UK region. Coloured lines indicate the three UKCP runs from which the study winters were chosen. The study winters are circled and dotted horizontal lines indicate the deviation 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	
Future projections	02868	2066	HOT	HOT1
	01554	2072	HOT	HOT2
	02242	2068	WET	WET
Baseline ensembles	02868	2007–2016	-	HOT1-B
	01554	2007–2016	-	HOT2-B
	02242	2007–2016	-	WET-B

Table 5.1: 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 5.2 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 (252), 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 (162). During this winter, the El Niño 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 (253). 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. (252): a strong positive NAO with associated pressure high off the west coast of Spain. This

5.4. Study design & methods

weather pattern is associated with warm and wet weather over the UK ([162, 254–256](#)). There is a weak La Niña (negative) ENSO phase ; which has previously been linked to an increased likelihood of positive NAO ([257–260](#)). No other modes of SST variability are present. With regard 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 ([261, 262](#)), though this is still a subject of active scientific interest ([263–265](#)).

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.

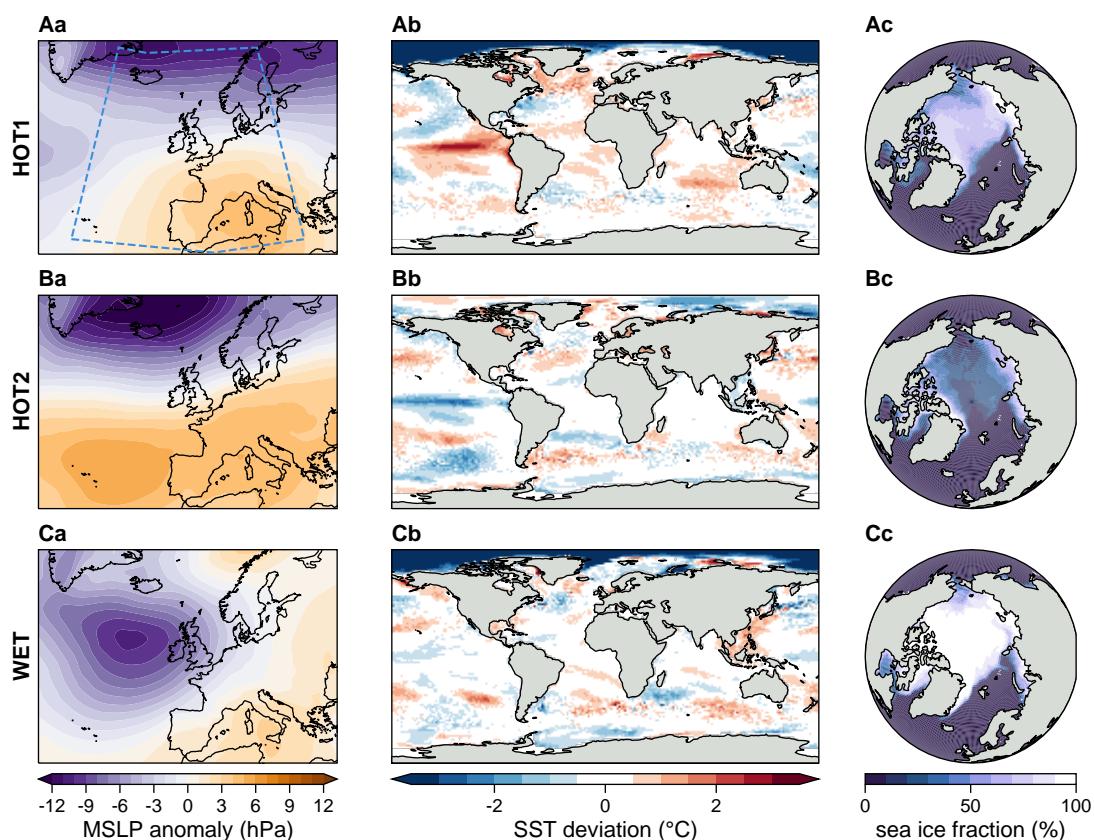


Figure 5.2: Synoptic characteristics of the study winters within the UKCP simulations. The row titles indicate the study winter. **a**, DJF mean MSLP anomalies for each winter. **b**, DJF mean SST deviations for each winter. Deviations are calculated for each gridpoint timeseries over 2061–2080. **c**, DJF mean Arctic sea ice fraction for each winter. The blue dashed line in **Aa** indicates the area used for analogue subsampling described in [5.4.3](#).

5.4.3 Statistical methods

Estimating distributions of extremes

We estimate distributions using the method of L-moments (145, 266, 267). 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 non-parametric 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 (268, 269) 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 (145) 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 (270, 271). 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 (272), though further work is needed to determine this conclusively.

Analogue construction

In order to assess the dynamical contributions to the extreme weather simulated during the study winters, we use an MSLP analogue approach (172, 273, 274). For each future ExSamples ensemble (and each corresponding baseline ensemble), we create a subsample of analogues 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 5.2Aa (-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 (252). We used a 3hPa threshold as this was the tightest constraint that resulted in analogue 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 analogues in which the mean large scale flow during the winter very closely matches the study winter. We can then use these ensembles of analogues to estimate the dynamical contribution and associated uncertainty to the extreme weather. ExSamples future and baseline ensemble distributions of MSLP distance from their corresponding UKCP winter are shown in Figure 5.3.

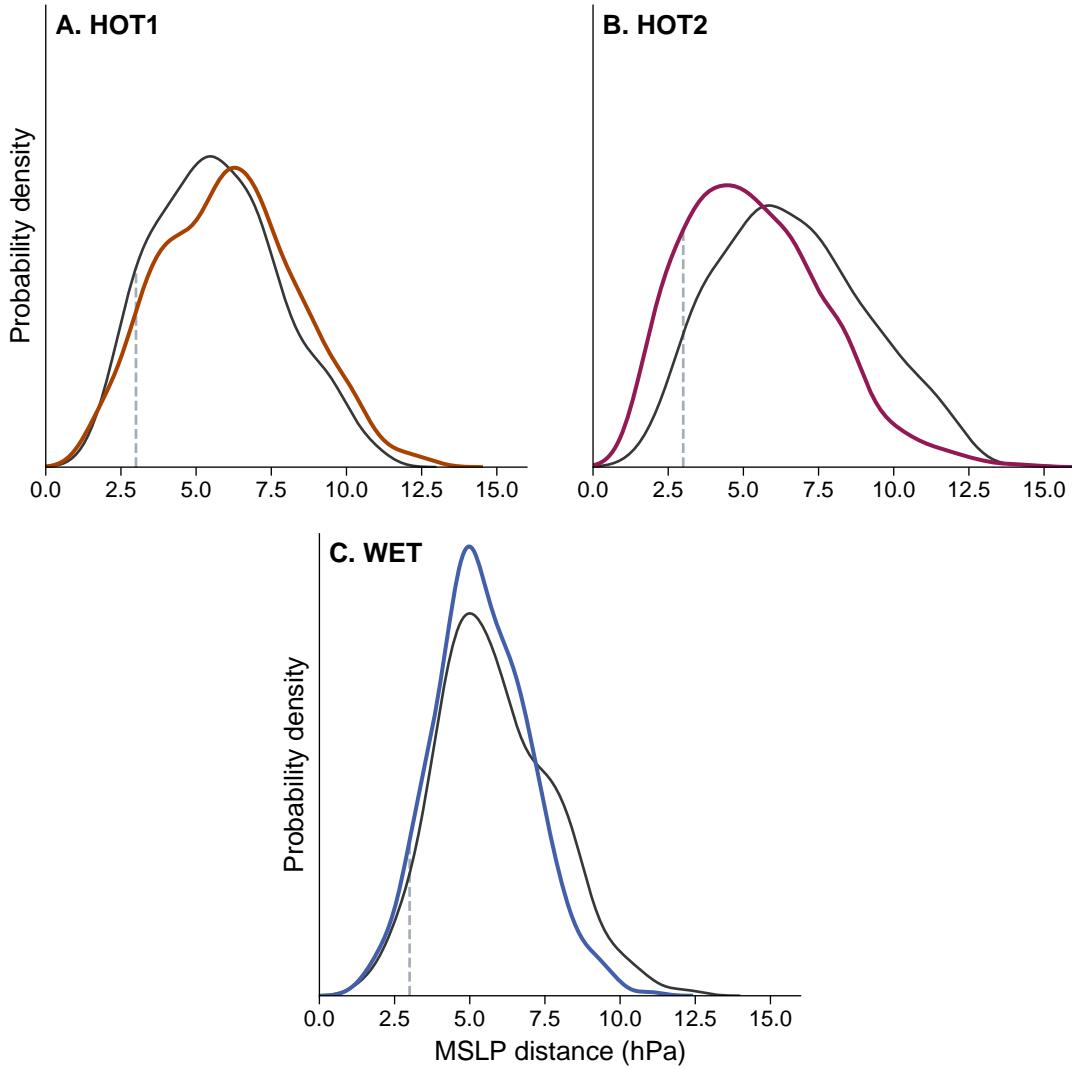


Figure 5.3: PDFs of Euclidean MSLP Distance between ExSamples ensemble members and corresponding UKCP18 extreme winters. Coloured lines show PDFs of ExSamples future ensembles, black lines show corresponding baseline ensembles. Dashed grey line indicates a distance of 3 hPa, the bounding limit of the analogue selection criterion.

5.5 Results

5.5.1 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-Smirnoff (K-S) test (275–278) 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 the scope of this paper, but has already been studied for both the UKCP18 PPE (230) and HadAM4 (242, 243).

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 5.5, 5.6, 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 mm day⁻¹) lower ensemble mean. They also have a slightly increased spread in winter rainfall. To check if these differences are due to differences in the simulated large-scale dynamics in HadGEM3-GC3.05 and HadAM4, we use a principal component (PC) analysis. Significant differences in these dynamics would be a concern, as it would represent a fundamental difference between the two models, and would make the ensemble comparisons in the main text less meaningful. We compute the principal components (PCs) and corresponding empirical orthogonal functions (EOFs) of DJF mean MSLP anomaly data from the UKCP18 PPE 1997–2016 over the region bounded by 35:70 N, -30:20 E (252). We then determined which PCs were most important in

explaining UK rainfall variance using an ordinary least-squares regression cross-validation. Regressing the top 20 (in terms of MSLP variance explained) MSLP PCs against DJF mean rainfall averaged over the UK from the same simulations, we exclude one PC at a time, and observe which exclusions reduce the total rainfall variance explained by the regression model by the largest amount. Three PCs are clearly more important than the rest: 0, 1 and 4 (which explain 43, 33 and 2.3 % of the overall variability in MSLP respectively). The regression model using these three PCs as the predictors explains 71 % of the total variance in UK rainfall. Their corresponding EOFs are shown in Figure 5.4. We then project each EOF onto all the ExSamples baseline simulations, and compare the distribution of the resulting pseudo-PC with the distribution of the corresponding PC in the UKCP baseline. We find no statistically significant differences between the ExSamples and UKCP distributions of these PCs (Figure 5.4). Hence, we conclude 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. Despite the apparent 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 5.7) 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 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.

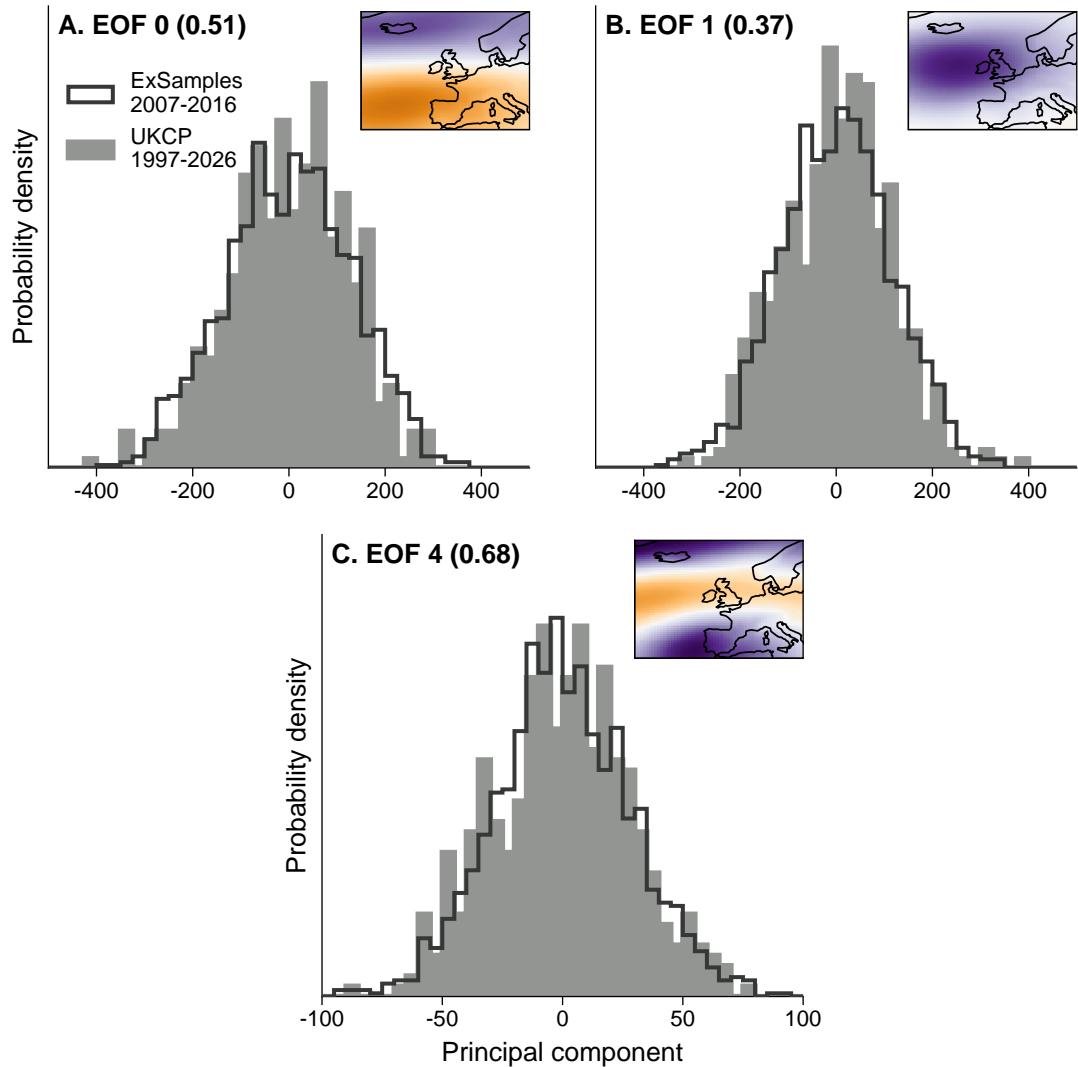


Figure 5.4: Rainfall variability-explaining PC histograms and EOF patterns in ExSamples and UKCP baselines. Grey filled histogram shows the original PCs from decomposition of UKCP18 PPE 1997-2026 MSLP. Black line histogram shows corresponding pseudo-PCs from EOF projection onto the aggregated ExSamples 2007-2016 baseline. The bracketed value in each subplot title states the p-value of a two sample K-S test, $P(H_0 : A == B)$. Inset shows MSLP pattern of each EOF.

5.5.2 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 5.5.

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 5.5: 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 5.5B). 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 (25, 35). 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 5.5: 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 analogues 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 5.5B) by a factor of 6.2 [5.3–6.9][†]. A similar level of dynamical conditioning is seen in the baseline ensemble. The analogue-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 analogues in the HOT1 ensemble is 20 % lower than in the HOT1-B ensemble (Figure 5.3). Note that this change in analogue 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).

Study winter	Variable	UKCP18 quantile (return period)		
		0.9 (1-in-10 year)	0.95 (1-in-20)	0.99 (1-in-100)
HOT1	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]
HOT2	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]
WET	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 5.2: Ratio of exceedance likelihood of three extreme thresholds between the ExSamples future ensembles and the UKCP18 PPE 2061–2080 deviations.

[†]Numbers in square brackets [] represent a 90% confidence interval throughout this chapter.

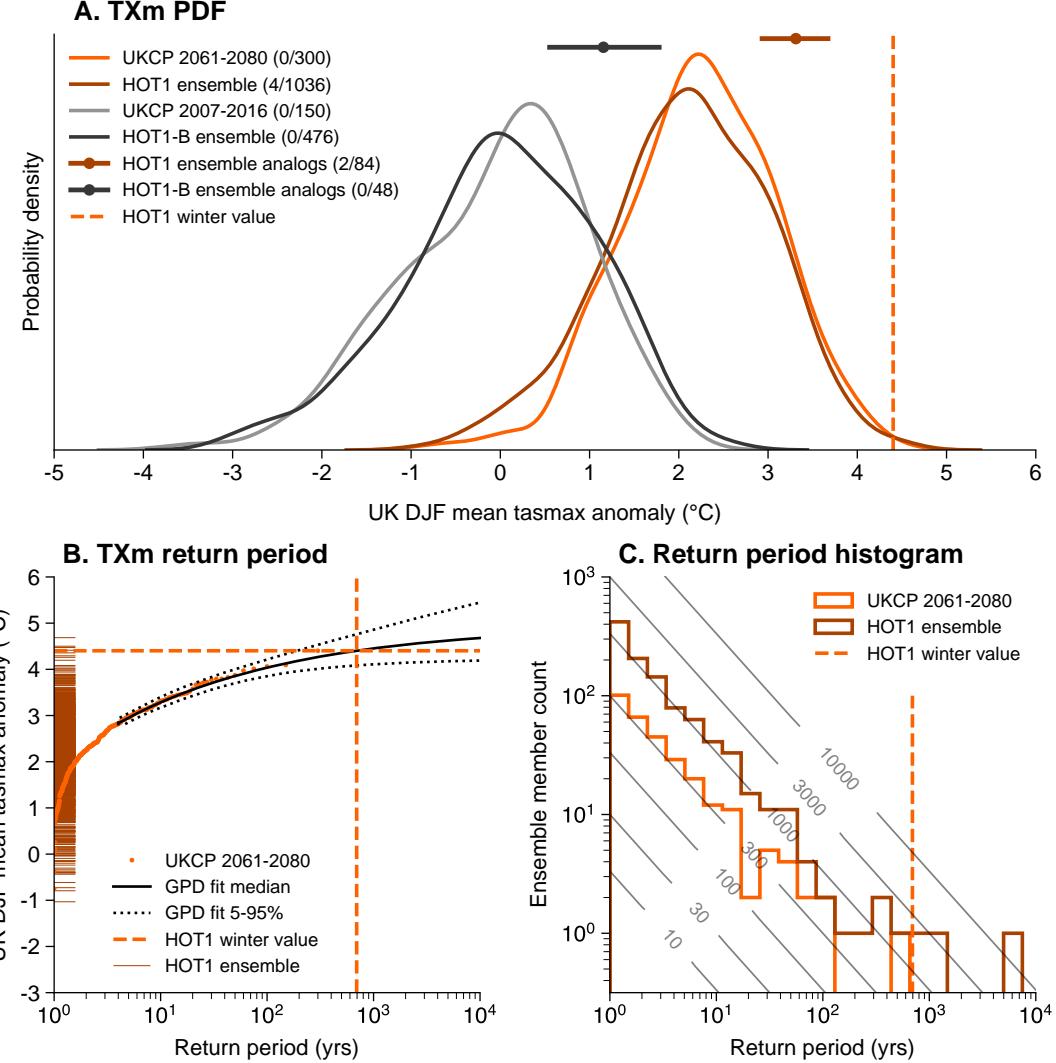


Figure 5.5: Comparing statistics of DJF mean of daily maximum temperatures (TXm) averaged over the UK region for the HOT1 winter. **A**, PDFs of baseline and future ensembles. The light orange PDF shows UKCP 2061–2080 deviations, with the distribution mean set to the ensemble mean anomaly between 2007–2016 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 analogue 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.

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**.

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 5.6A, C, Table 2), suggesting that it was conditioned towards such extremes by the prescribed SST/SIC. We can see from Figure 5.6C 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 analogue subsamples in Figure 5.6A. Based on the number of analogues 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 (Figure 5.3). This would fit within the canonical picture that the negative La Niña ENSO phase is associated with positive NAO (257, 279).

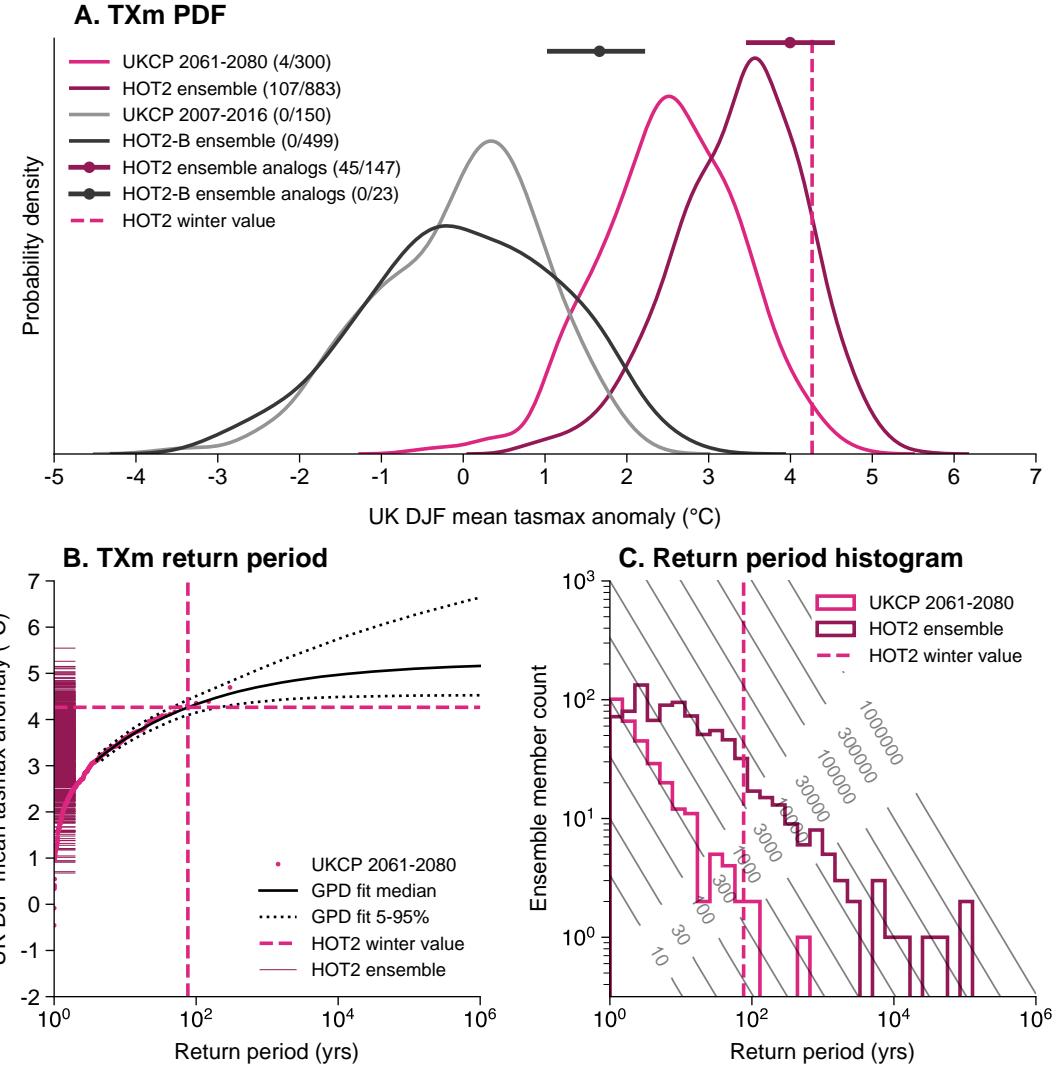


Figure 5.6: Comparing statistics of DJF mean of daily maximum temperatures (TXm) averaged over the UK region for the HOT1 winter. As Figure 5.5, but for the HOT2 winter.

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 5.7. 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 analogue-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 5.7A. This may be due to the SST/SIC conditioning in the future ensemble.

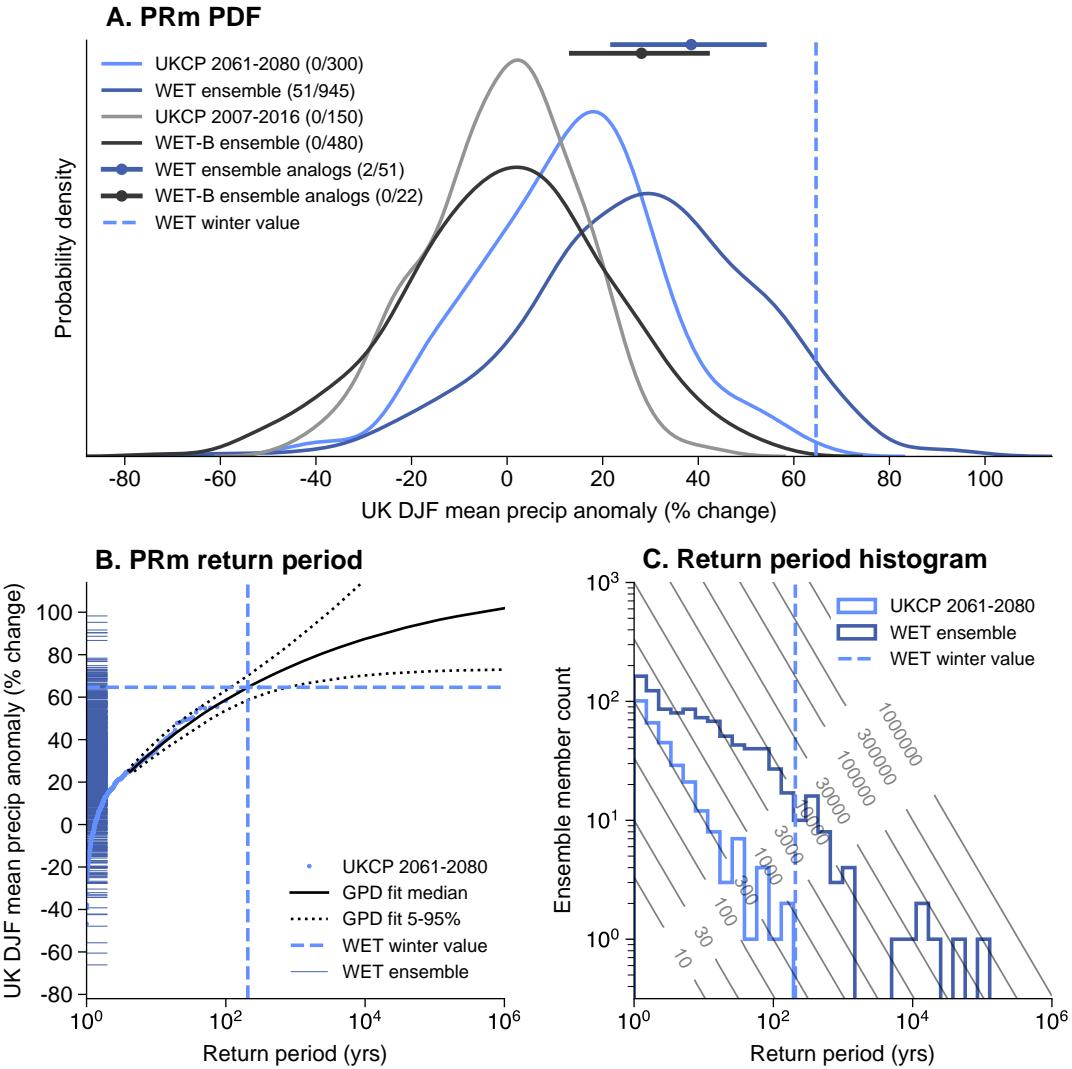


Figure 5.7: Comparing statistics of DJF mean precipitation (PRm) averaged over the UK region for the WET winter. As Figure 5.5, but of DJF mean precipitation averaged over the UK region for the WET winter.

5.5.3 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 (280–282). 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 (TX_x as previously defined). Figure 5.8 shows a return period diagram of UKCP 2061–2080 TX_x deviations (centred 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 (283). 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. (284). The dashed dark orange line in Figure 5.8 shows the TX_x for HOT1 ensemble member c0qu, which lies considerably above (by 2.3 °C) any UKCP18 PPE samples. This event is roughly 5 standard deviations above the mean of the UKCP18 deviations distribution shown in Figure 5.8. 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 (268), it sets an upper bound on TX_x , consistent with previous studies of extreme heat events (190). 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 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 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 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 (108, 162). 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.

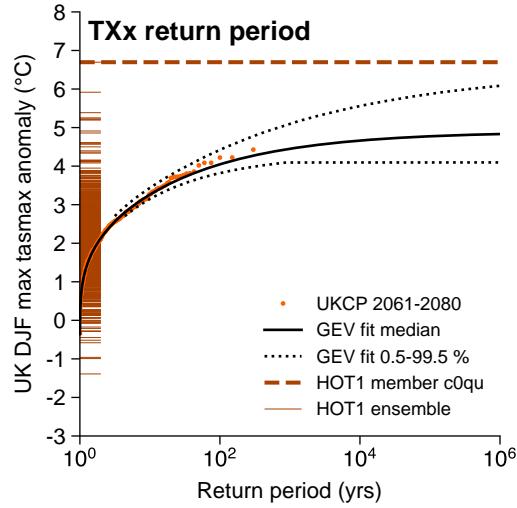


Figure 5.8: Examining statistics of subseasonal extreme weather events. As return period diagram of Figure 5.5, but of DJF maximum of maximum daily temperatures averaged over the UK region for the HOT1 winter. The statistical model indicated by the solid and dotted black lines is a generalised extreme 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.

5.6 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 of these 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-centred deviations to define the UKCP18 maxima as shown in Figures 5.5–5.7, 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

TX_x 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 (282). 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 (146, 147, 285, 286). 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 (285) 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 5.5-5.7, 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 (i.e. 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 (178, 287), 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 (288). 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.

5.7 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 ([230](#), [243](#)).

5.8 Chapter close

This work has described and demonstrated a novel methodology for producing a large number of samples of extreme winters. I suggest that the primary outcome lies in demonstrating how this method is able to produce a rich variety of extreme events at a relatively low computational cost. This rich variety of multivariate extremes could be used to estimate potential limits to adaptation, which are relevant to policymakers. However, it is important for me to emphasise that this is a proof-of-concept, rather than something to be used immediately. Although HadAM4 has been shown to validate well in the winter, further model validation focussed on extreme events would be necessary to increase confidence in the realism of the future samples. In addition, providing samples at several global warming levels would further increase the information provided by this methodology. The clear outstanding science question is whether it is possible to predict which lower boundary conditions from the coupled model will condition the atmosphere-only simulation towards extremes, and understanding the physical reasons why. This would enable the ExSamples methodology to produce samples of extremes even more efficiently.

If I now consider the thesis as a whole, the work presented in this chapter provides a contrast to the others: while forecast-based attribution concentrates on understanding single weather events as reliably and in as much detail as possible, the ExSamples approach aims to span the uncertainty associated with an extreme in order to produce as broad a range of plausible extremes as possible. I argue that both are important: in order to effectively adapt to climate change, we need to understand how specific extreme weather is changing in as much detail as possible; but we also need to quantify the breadth of possible future extreme weather to avoid damaging surprises. I discuss how forecast-based approaches could be used for projecting future extremes in more detail in the Discussion; but believe that the methodology presented in this chapter could provide highly complementary information in a computationally efficient manner to such a forecast-based approach.

Quote

— author

6

Discussion

In this final chapter, I consider this thesis as a whole. Starting with a brief overview of everything that I have covered within the science chapters, I then move on to discussing related precedent works, including how each is similar to and different from my own, and the advantages and disadvantages of various approaches. I examine the limitations of the forecast-based approach to attribution that has been my focus — and how these limitations could be overcome in further work. I expand this discussion of future research directions to include potential applications beyond the attribution of extreme weather. I end the chapter, and thesis, with some closing thoughts.

Contents

6.1	An overview of this thesis	114
6.2	This thesis in the context of previous work	116
6.3	Limitations	121
6.4	Future research directions	124
6.4.1	Addressing the rapid atmospheric adjustment	124
6.4.2	Expanding the scope of this work	127
6.4.3	Alternative applications	128
6.5	Concluding remarks	131

6.1 An overview of this thesis

I started this thesis with a discussion of the precise question that this thesis would be examining, posing it as: ‘how has human influence on the climate affected both the probability and intensity of specific extreme weather events?’. Following this, I provided several reasons why this is an interesting and useful question to answer, including: quantifying the fraction of damages from individual weather events that GHG emitters are liable for in both tort law and loss and damage frameworks; engaging the public in climate change research and impacts; improving our understanding of extreme weather to improve both adaptation planning and the models we use to simulate such weather. I then described the two most common frameworks that have previously been used to answer this question: the probabilistic and storyline approaches. Both of these approaches have advantages and disadvantages, but I suggested that they could be synthesised and complemented through the use of counterfactual weather forecasts. One way of achieving such a forecast-based approach is the main contribution of this thesis to the literature. I concluded with a description of the underlying physics behind heatwaves, and the basis for how they are being affected by human influence on the climate.

In chapter 2 I presented an attribution analysis of the 2018 heatwave in Europe. This analysis used a conventional probabilistic approach to explore whether two seemingly conflicting quantitative attribution results publicised in the media could be reconciled by digging into the technical details. I found that this conflict could indeed be largely resolved when considering the different temporal scales used to define the event in those previous results. This analysis provided an introduction to the methods used widely in probabilistic attribution. However, it also resulted in a number of questions over the approach surrounding model validation and the specificity of the analysis: in particular whether I could genuinely claim to be attributing the specific 2018 event given that none of the meteorological context had been taken into account in the probabilistic approach followed in this chapter. One apparent solution to these questions would be to use weather forecast models that had successfully predicted the event in question to assess the human contribution to the event — as opposed to the climate models that are commonly used. I developed this idea of forecast-based attribution in the two following chapters.

Chapter 3 presented the first example of my forecast-based approach to attribution. Motivated by the remarkably accurate forecast of an exceptional heatwave in February 2019 by the operational ECMWF ensemble prediction system, I perturbed the CO₂ concentration in this system back to pre-industrial levels to examine how the heatwave would be impacted. I argued that — in addition to the high-resolution of the forecast model — this successful prediction is a key feature of the forecast-based approach, as it largely guarantees that the model is able to represent all the key processes that drove the heatwave. An important limitation of this study is that it only addressed the contribution

of a single component of anthropogenic influence on the climate: the direct radiative impact of increased CO₂ concentrations above pre-industrial levels. However, despite this limitation, I still found a detectable CO₂ signal in both the intensity and probability of the heatwave. One surprise was how rapidly the signal was detectable within the forecasts: it could be detected even at a very short lead of 3 days. Another interesting feature was how stable the predictability of the heatwave was: despite the relatively large perturbation made to the forecast boundary conditions, it remained predictable in both reduced and increased CO₂ counterfactual forecasts. Although there were several possible directions for future work, I identified that trying to produce a more complete estimate of human influence on an individual event was the most important one, and so the following chapter explores doing exactly that.

The 2021 Pacific Northwest Heatwave is the case study used in chapter 4. This was a truly unprecedented event that not only shattered local temperature records by several degrees, but also presented conventional attribution methods with a considerable challenge. Statistical models often suggested that the event should have been impossible, and it was unclear if climate models were even able to simulate such an extreme event. I argued the reason for these challenges were the very specific processes that drove the heatwave: an atmospheric river coincided with an anticyclonic block, providing both diabatic and adiabatic heating to air parcels that then descended to the surface and heated further through soil-moisture feedbacks and insolation. This optimal combination of processes had likely not been seen before in the historical record — hence the difficulty faced by statistical approaches. However, the event was incredibly well-forecast given this context, with suggestions of such extreme temperatures appearing over 10 days before they actually occurred. I took the same ECMWF model used in chapter 3, and modified both the CO₂ concentration boundary conditions and the initial ocean conditions to produce an initial state consistent with a climate without human influence. These perturbations resulted in a pre-industrial counterfactual forecast; and I also applied exactly opposite perturbations to produce a ‘future’ counterfactual forecasts. These counterfactual forecasts indicated that the heatwave was made at least 8 [2–30] times more likely due to climate change — and that this risk is doubling every 17 [10–50] years at the current rate of global warming. Although there remain scientific challenges for this approach to overcome, this chapter demonstrated that an operational attribution service based on counterfactual forecasts is not far out of reach, and could provide extremely valuable information for adaptation planning.

Chapter 5 provided a contrasting but complementary study. While extreme event attribution is typically backward-looking, examining changes in weather since the pre-industrial period, here I looked at the projection of climate extremes. And while extreme event attribution focuses in on a single specific event, here I explored how the full space of uncertainty in all such extremes could be sampled from. I performed this

exploration upon a novel set of large-ensemble atmosphere-only simulations forced by lower boundary conditions from three of the most extreme winters simulated within the UKCP18 coupled global runs. These ~1000-member ensembles allowed me to better understand the uncertainty surrounding these extreme winters, including to what extent they were forced or whether they were driven by atmospheric internal variability. They demonstrated that the atmosphere-only model used was not only able to simulate such extremes, but also to produce extremes beyond anything found in UKCP18 — even beyond what might be expected from statistical extrapolation. I suggested that this computationally efficient methodology could be used to provide a rich multivariate set of spatially, temporally and physically coherent samples of extremes, and that such a set could provide valuable information for adaptation planning and designing high-impact low-likelihood scenarios of future weather. It could be effectively complemented by specific ‘Tales of future weather’: counterfactual weather forecasts of damaging historical events as if they occurred in a warmer world, analogous to those in 4. Such Tales could provide detailed information about the future risk from and physical understanding of specific extremes, while the large ensembles in this chapter could provide broad information about the range of future extremes.

6.2 This thesis in the context of previous work

First approaches using weather forecast models Forecast-based attribution as an idea is not novel to this thesis, though this thesis does extend the approaches taken previously. One of the main reasons why weather forecast models have been used previously is their high resolution — though simply using weather forecast models is not necessarily the same as using weather forecasts. Such high resolution is required for representing localised dynamical weather extremes such as hurricanes or convective storms. Possibly the first study to explicitly use a weather forecast model for the purpose of attribution was Lackmann (159) *. Lackmann used the Weather Research and Forecasting (WRF) model (289) in a triply-nested setup driven by reanalysed SSTs to analyse how Hurricane Sandy may have evolved differently in both pre-industrial and future climates. The counterfactual climate forecasts were produced by perturbing the initial and boundary conditions in line with changes estimated from CMIP3 GCM simulations (290). With this experiment design, Lackmann showed that Sandy’s intensity would have reduced and its track would have shifted southward, conditioned on its predictable component at a lead of around 3 days. An analogous design was used by Meredith et al. (154), who used a similar triply-nested WRF model setup to examine the

*Although Hoerling et al. (65) did make use of weather forecasts in their storyline analysis of the 2011 Texas combined heatwave and drought, these were not used for attribution explicitly, but rather to explore the predictability of the event.

highly nonlinear influence of SST warming trends on a convective extreme near the Black Sea. Meredith et al. also used perturbed SSTs for their counterfactual simulations, but did not perturb other initial or boundary conditions. They additionally nudged the large-scale circulation to keep the dynamics consistent between their simulations.

The pseudo-global warming method The approach taken by Meredith et al. falls under a broad umbrella of approaches often referred to as ‘pseudo-global warming’ experiments (291). In the basic version of this experiment design, a regional model is nested within some time-evolving lateral boundary conditions (atmospheric and sea surface). These boundary conditions are then modified in line with global warming to examine the influence on the (freely-evolving) regional of interest. This can be used to study weather extremes or any other scenario of interest. The boundary condition modification in Schär et al. is similar to that of Lackmann, but the latter allows the atmospheric conditions in the outer domain to evolve freely after initialisation. On the other hand, Meredith et al. didn’t modify the atmospheric boundary conditions, but instead nudged them towards the scenario of interest. The pseudo-global warming approach was used in a study examining the impacts of Typhoon Haiyan by Takayabu et al. (292). They used a chain of successively more granular models, starting from a 60 km global ensemble forecast model, and ending at a 740 m ocean wave and storm surge model. Their use of an operational ensemble forecast system is interesting, though they did not focus on the probabilistic aspect of the system, instead using it to develop a plausible worst-case scenario of impacts from the typhoon. More recently, the related ‘hindcast attribution method’ has been developed, often to study convective dynamical events (197). This method represents a specific application of the pseudo-global warming approach in which either a regional model is forced by lateral boundary conditions from reanalysis of an extreme or a variable-resolution global model hindcast is used. The initial and lateral boundary conditions are then modified to be representative of a world without human influence in order to assess the conditional impact on the event in question. This hindcast attribution method was first established by Pall et al. (185) to examine anthropogenic contributions to an exceptional Colorado rainfall event in 2013. It has since been used extensively to study human influence on hurricanes by Michael F. Wehner and colleagues (158, 224, 293). At this point, it is worth mentioning a few differences between the pseudo-global warming and hindcast attribution methods and the forecast-based approach set out here. Although all of these studies employ ensemble simulations, and some do make probabilistic statements (185), the reliability of the models used is not assessed and therefore it is tricky to determine the robustness of such probabilistic statements (62, 92). This is a key limitation since it means that changes in probability cannot necessarily be accurately estimated from the ensembles (since the quantiles of individual ensembles may not represent the true outcome probabilities). In terms of the

nesting of models, if a reanalysis-forced regional model is used, then climate change responses and interactions outside the domain are not taken into account, which may be of importance for particular events. If a variable-resolution global model is used, biases present in the coarsest domain will be inherited and may be relevant to the simulation within the study domain. Finally, the prescribed-SST designs used in all the studies mentioned preclude ocean-atmosphere interactions that may also be important in the development of extremes such as hurricanes.

The subseasonal to seasonal forecast approach at BOM Certainly the most similar prior work to the forecast-based approach presented here is that of Pandora Hope and colleagues (155–157, 195, 196). In Hope et al. (155), a new method of extreme weather attribution was presented involving reinitialised coupled seasonal forecasts from the Australian Bureau of Meteorology’s POAMA model. Starting from the operational forecast model at the time in 2014, they first altered the CO₂ concentrations from the 2014 level of 400 ppm to a level consistent with the conditions in 1960 of 315 ppm. They then estimated the anomaly required to remove the ocean’s response to these two distinct atmospheric concentration levels by integrating two sets of free-running coupled simulations for 30 years. Two ‘recent’ climate simulations were initialised from observed initial conditions in 2000 and 2010, and two ‘1960s’ climate simulations from 1960 and 1970. An average over the last five years of the sets was used to create temperature and salinity perturbations consistent with the modelled ocean’s response to the different CO₂ levels. They found that the probability of the record warm Australian spring of 2014 was significantly reduced in these 1960s condition forecasts. Hope et al. (156) further extended this methodology by also modifying the initial land-surface and atmosphere conditions of the model. The land-surface (soil moisture and temperature) and atmosphere (humidity and temperature) perturbations are determined in a manner similar to the ocean, but with much shorter integrations of 2 months over each of the 15 years spanning 2000-2014 and using either true observed initial conditions or the modified CO₂ and ocean conditions as in Hope et al. (155). This complete perturbed CO₂-ocean-atmosphere-land seasonal forecast attribution system is described by Wang et al. (195). It was used by Hope et al. (156) to assess another record-breaking heat event, and by Hope et al. (157) to examine anthropogenic influence on the precursors of fire-weather. It now forms the basis for a near real-time ‘Event Explainer’ service, as described by Hope et al. (196).

There is evidently considerable overlap between the BOM approach described above and in Hope et al. (196) (henceforth H22) and the methodology developed in this thesis throughout the course of chapters 3 and 4 (henceforth L22). Both make use of weather forecast models to analyse extreme events within the limit at which they are predictable within those models. Both perturb both the initial and the boundary conditions of the weather models to assess the impact of human influence on such extremes. However,

there are a number of clear differences. H22 has previously focussed on subseasonal to seasonal forecasts and timescales, while L22 has concentrated on shorter medium-range timescales. This different temporal focus impacts the types of event that can be analysed by each approach — with H22 able to examine events taking place over periods longer than the integration length of medium-range forecasts, and L22 able to study sharper and more short-lived events that may not be predictable on seasonal timescales. For example, the H22 approach is suited to analysing monthly mean temperatures (155); while L22 more suited to analysing heatwave peak temperatures as in chapter 4. The longer timescales involved in seasonal forecasts also mean that model drifts may be important, something that I found in chapter 4. The atmospheric model used in the POAMA seasonal forecast system in H22 is coarse, at a horizontal resolution of 250 km. This coarse resolution means that this model will share the difficulties faced by many coupled climate models when it comes to simulating extreme events. An upgraded system is under development at a resolution of 60 km, but this is still insufficient for simulating some key classes of extreme weather such as hurricanes (196, 293). On the other hand, the operational ensemble prediction system at ECMWF used in L22 has a resolution of approximately 18 km. In terms of the ‘completeness’ of the estimated human influence, H22 is ahead of L22. Thus far, L22 perturbs both the initial ocean and sea-ice conditions, and the CO₂ concentration boundary conditions. This is analogous to the perturbations made by Hope et al. (155, though they did not perturb the sea ice). However, in the most recent studies, H22 also perturb the land-surface and atmosphere, thus allowing for a more complete estimate of human influence on the extreme event of interest. In particular, leaving the atmosphere unperturbed requires it to adjust over the course of the forecast, which is a clear limitation on the interpretation of results from this approach. I discuss this limitation and potential methods to resolve it below. It would be very desirable for the two approaches to reach a point at which they were consistent enough to start providing multi-model attribution results, in order to further increase the robustness of forecast-based statements.

Spanning uncertainty in climate projections of extreme weather Although the focus of this section is looking at work related to the forecast-based approach to attribution that I have developed, here I also briefly discuss some prior studies done on projections of climate extremes relevant to chapter 5, particularly on how to capture the associated range of uncertainty. I have split these into two broad groups for clarity: dynamical and statistical modelling studies.

The primary challenge posed by climate extremes is that they are, by their nature, rare, and therefore typically require a considerable amount of data to make inferences about. One increasingly common approach to addressing this requirement is to use large ensembles of climate model simulations. These may be multi-model ensembles, such as

the CMIP ensembles (131, 178); single model initial condition large ensembles, which are used to separate internal climate variability from forced responses (294); perturbed parameter ensembles, such as the global UKCP18 ensemble used in chapter 5 (229); or combinations of the above that aim to span as complete an uncertainty space as possible (247, 248). All of these types of large ensemble explore slightly different components of climate projection uncertainty, and can clearly answer a range of useful questions about future climate extremes. However, none are designed specifically for the study of extremes in the same way that the ExSamples methodology described in chapter 5 is. The ExSamples method also explores a different component of uncertainty to each of these large ensemble approaches: atmospheric internal variability (as opposed to combined atmosphere-ocean internal variability). The most similar approach to ExSamples, unsurprisingly, is that which it was based on: the very large atmosphere-only simulations performed by *climatedprediction.net* for attribution of extreme weather (67, 295). However, unlike ExSamples, all the previous work using this approach simulated the present and the past, rather than the future. One interesting recent approach developed explicitly for studying and simulating the most extreme events possible is ‘ensemble boosting’. As described by Gessner et al. (190), ensemble boosting locates the most extreme events in pre-existing climate projections, and then reinitialises an initial condition ensemble a few days before, in order to see if it could have been even more extreme. Ensemble boosting could therefore be very useful for delving into what the maximum possible extremes might be in the future — but is not designed to provide the same richness in the variety of future extremes that ExSamples explores.

The main alternative to such an extreme-specific methodology as ExSamples is to apply statistical models to existing climate model projections (such as the large ensembles described above). One such alternative would be the extreme value-based approach of Brown et al. (283). They use an extreme value distribution that depends on the global mean temperature and accounts for model bias to predict changes in future regional extremes. This statistical approach can derive useful additional information from existing model simulations at low computational cost. However, it cannot provide the same multivariate and physically coherent extreme samples as ExSamples. In addition, my results in chapter 5 suggested that extreme value theory may underestimate the likelihood of the most extreme events. A different possible methodology for exploring the uncertainties in extremes could be to use a weather generator trained on existing model projections (296). Yiou and Jézéquel (297) used a circulation analogue-based approach to estimate how hot European summers could get in the present-day. This approach does provide some physical coherence by linking the estimated temperatures directly to the atmospheric circulation. However, it does not include any other feedbacks, such as soil moisture, would be complex to obtain multivariate information from, and cannot produce extremes driven by processes (or combinations of processes) that lie

outside of the training data. As such, it is more suited to providing information about long-term extremes (ie. on seasonal timescales) than about short-term extremes such as the Pacific Northwest heatwave. The lack of feedbacks may mean that the likelihood and intensity of the most extreme events is underestimated; though perhaps information about these feedbacks could be included in the future ([298](#))?

6.3 Limitations

Although I have discussed various limitations within the individual chapters that make up this thesis, in this section I consider some of the limitations of the forecast-based approach to attribution developed and explored here as a whole.

Forecast adjustment One key aspect of the counterfactual forecasts performed here is that the model — or more specifically the model atmosphere and land surface — adjusts continually to the imposed perturbations throughout the integration. This means that the further into the forecast the event of interest happens, the stronger the attributed impact of those perturbations is. At the same time, as the forecast evolves, this effect becomes more uncertain in general (though not necessarily always), due to the increasing dynamical noise arising from the chaotic nature of the weather system. This combination of this increasing strength and uncertainty can make analysing and interpreting the results of the counterfactual forecast experiments difficult. In chapter 4 I alleviated this difficulty by making use of the fact that the attributable regional impacts of climate change were near-linearly related to the coincidental measured level of global warming. This linear relationship allowed me to benchmark the estimated impact at each forecast lead time to the same level of global warming, regardless of how adjusted they were at the time of the event in question. However, this linear relationship is not guaranteed for every extreme event, and therefore it would be valuable to find methodologies by which this adjustment could either be reduced or removed entirely. I consider a few ideas to achieve this below.

Additional uncertainty dimensions In the experiments performed here, I have only considered uncertainty arising from the chaotic nature of the weather system. However, there are additional uncertainties associated with the approach I have developed. One dimension that has been explored in prior work is the uncertainty in the estimation of the “human fingerprint” that is removed from the model initial conditions. For example, Pall et al. ([67](#)), who removed such a fingerprint from the SSTs (and sea ice fields) of the initial conditions in their naturalised simulations, used four estimates of the warming pattern based on different coupled climate models. This allowed them to test the sensitivity of their attribution statements to the warming pattern used. This approach of using an ensemble

of coupled climate models to derive a corresponding ensemble of human fingerprints in order to more completely sample this dimension of uncertainty space has since been used in several other studies (295); though many attribution studies and systems still use a single estimate, often based on a multi-model mean pattern (129, 174).

In this thesis, I used a single pattern estimated from observations. I used observations to avoid over-reliance on a single coupled climate model, given the biases and known issues present in such models. Although it would have been extremely interesting to more completely explore this dimension of uncertainty (given observations of the ocean subsurface are by no means perfect, especially pre-2000), limits to computer resources prevented me from doing so within the scope of this thesis. However, I hope that such an exploration could be carried out in the future. Doing so would be conceptually straightforward, simply involving treating historical output from a set of different coupled climate models exactly as if they were the observations that I used. Quantitative attribution results derived from each coupled model estimate could be compared to test the sensitivity of such results to the estimate of the human fingerprint used. Because of the variation in the representation of transient historical climate within coupled climate models, each model-derived fingerprint might have to be scaled by (for example) the present-day level of global warming within the model for consistency (299).

Single model Within the core research of this thesis concerning forecast-based attribution, I have used a single model, ECMWF's IFS. There were a number of reasons for this limitation: ECMWF provided a mechanism for me to access the computing resources I required through their special project; there were also individuals at ECMWF who provided the technical expertise I needed to design and perform the counterfactual forecast experiments; the IFS is one of the (if not the) best performing numerical weather prediction models on a global basis (300); and applying the counterfactual forecast methodology to other models would have presented considerable technical challenges that lie beyond the scope of this thesis. However, the quantitative results presented here may be sensitive to this choice of model.

There is considerable variation in both the global and regional response to external forcing among climate models (3, 214, 222, 227, 301, 302). This variation is the reason why Philip et al. (61) suggest that having as large a set of different climate models as possible is important for a probabilistic attribution study. Although still a point that should not be overlooked, I argue that such a multi-model assessment is not as important when using the counterfactual forecast approach introduced here. Firstly, a successful prediction ensures (when combined with a limited validation that the prediction did not occur for the wrong reasons) that the model used is able to represent the physical processes of the event in question, and that vital processes are not missing, as may be the case in some climate models. This grounding in the specific physics of the event

means that the simulated response to external forcing is considerably more certain and less model dependent. Secondly, the use of a *reliable* forecast ensemble to assess probability ensures that these probabilities are representative of the full space of possible states of the climate system given the initial conditions of the forecast (303). This is not the case for climate model simulations, including perturbed parameter ensembles. However, despite these mitigating factors, exploring the sensitivity of the counterfactual forecast approach to the model used is an important question for future research. This could be done by carrying out an identical experiment in (for example) the Met Office's numerical weather prediction systems (304, 305).

Single event class This thesis has concentrated on extreme heat events. However, given the major contribution of the thesis to attribution literature has been the forecast-based approach taken, rather than understanding the specific type of extreme event studied, this does represent a limitation. There are good reasons for this focus on heatwaves, given the possible scope of a thesis: they have severe associated impacts; are generally well understood; and have been the subject of a large body of prior attribution literature. However, demonstrating that the forecast-based approach can be used for other classes of extreme event will be vital for the method to be taken up widely. A possible candidate for the next class to study would be a high precipitation event. Attribution of high precipitation extremes is generally more challenging than of heatwaves, due to the smaller spatial scales involved, though there still exists a considerable amount of prior work that addresses this question. The high resolution of weather forecast models certainly makes them a more appropriate tool than coarse climate models for studying localised extremes.

Considering additional forcing agents In chapter 4, the ‘complete’ estimate of human influence on the Pacific Northwest Heatwave was derived by removing human influence on ocean heat content (by reducing the 3D ocean temperature) and reducing the levels of CO₂ in the atmosphere back to their pre-industrial levels. Although we argue that this represents a good estimate of the total sum of human influence, there are a number of additional sources of anthropogenic forcing on the climate system that may need to be considered in future work. Increased levels of other greenhouse gases such as methane or nitrous oxide have a similar radiative effect to CO₂, though the forcing from these other agents is relatively small in magnitude compared to CO₂ (214). The other significant human contribution arises from aerosol emissions. Unlike greenhouse gases, historical aerosol emissions have reduced the energy imbalance of the earth, thus masking some of the global warming caused by greenhouse gases. Additionally, while greenhouse gases are well-mixed throughout the atmosphere, aerosols are highly localised in space due to their short lifetime. This means that their effect on local climate can vary considerably from region to region. In this thesis we only considered forcing from

increases in CO₂ concentrations since i) forcings from these other sources approximately cancel each other out on global scales, and ii) the IFS does not include an interactive atmospheric chemistry model, but instead uses an aerosol climatology (306). There has been relatively little research into the effects of aerosols on heatwaves specifically (111), but it has been found that aerosol reductions in the future exacerbate increases in heatwave magnitude arising from continued greenhouse gas emissions (307). Including the effect of these additional forcing agents on specific extreme weather events would be an extremely interesting direction for future research. The effect of aerosols may be especially interesting for precipitation extremes, since aerosols are known to have direct impacts on cloud formation. However, while including the radiative effects from other greenhouse gases would be straightforward, and could be done exactly as has been for CO₂, including the effects from aerosol emissions would likely be considerably more technically complicated and subject to large uncertainty, though might be possible using a version of IFS that includes a tropospheric aerosol scheme (308).

6.4 Future research directions

6.4.1 Addressing the rapid atmospheric adjustment

In this section, I discuss possibilities for how the methodology used here could be altered in order to remove the issues associated with the rapid adjustment of the forecast model, in the atmosphere and at the land surface, to the perturbed initial state. Removing (or alleviating) this adjustment would considerably simplify the interpretation and analysis of counterfactual forecast experiments.

Perturbing the initial atmospheric state The simplest way in which to initialise the model from an atmospheric (and land-surface) state that is closer to thermodynamic equilibrium would be to attempt to perturb it such that it is consistent with the changes made to the oceanic and boundary (CO₂) conditions. This approach is based on the pseudo-global warming framework (291), and has been used by a number of recent attribution studies (158, 185, 197, 224, 293). However, unlike these studies and the original framework, which perturb the lateral boundary and initial conditions of a high-resolution nested regional model, in our case we would need to perturb the forecast model globally. These studies typically perturb 3D thermodynamic fields such as temperature, humidity and geopotential based on simulated climate change in GCMs. However, it would be possible to avoid reliance on such models by estimating anthropogenic fingerprints in these fields from long-term reanalysis data (86, 309) exactly as was done to estimate the ocean state perturbations in chapter 4. The disadvantage of this approach would be that the imposed atmospheric perturbations could cause unexpected

changes to the physical processes driving the event in question that would be difficult to distinguish from real attributable changes to the event. For example, changes to the temperature and moisture fields could affect local atmospheric circulation in a way that is not necessarily physically consistent with how the same event might have evolved in a climate without human influence.

Assimilating the perturbations The simple perturbation approach could be extended to counter some of the issues with physical consistency by coupling it to the data assimilation procedure used to generate the model initial conditions. Data assimilation aims to create the best possible forecast initial state by combining recent model predictions with observations (310). It may therefore have the potential to generate a balanced — but also physically consistent — initial state for the counterfactual forecasts. Data assimilation has been previously proposed as a technique that could be used for extreme event attribution by Hannart et al. (160), though they suggested using likelihoods output from the data assimilation procedure directly, rather than using the procedure to create initial conditions for counterfactual forecasts. The basic idea would be to replace the operational version of the forecast model with an ‘unforced’ version during the data assimilation cycle. This unforced version would essentially be identical to the perturbed initial and boundary condition model run to produce the counterfactual forecasts in chapter 4. The aim behind this unforced data assimilation cycle would be to produce a physically consistent initial state in the unforced model that is as close as possible (in such a climate without human influence) to the observed state of the climate system. How similar the original operational and unforced counterfactual initial states produced would be would depend heavily on the real-world state at the time of the data assimilation. Although this is a promising idea in theory, it would likely face significant technical challenges. For example, some of the observations used in the data assimilation may also have to be perturbed for the procedure to succeed (especially when using a perturbed ocean state if observations of the ocean are used). Given the incomplete nature of the observations and variety of their sources, altering observations before the data assimilation step could be extremely difficult. These potential challenges mean that collaboration with an expert in the operational data assimilation system used would be essential.

An operational approach using successive forecasts The previous two methods could be used to perform a counterfactual forecast for a single event. My final suggestion is potentially simpler than both, but would only work in the case of an operational, regularly run counterfactual forecast system for attribution and projection. This idea would be to use the previous forecast to calculate the perturbation required to create a balanced initial state, and is probably clearest expressed mathematically. If we call the operational (assimilated) initial state at time τ be $\chi(\tau)$, the operational ‘forecast operator’,

that transforms an initial state into a prediction at time t after initialisation G_t^1 , and the equivalent counterfactual operator G_t^0 , then operational and counterfactual forecast states X at time t after initialisation time τ can be written respectively as

$$\begin{aligned} X^0(t|\tau) &= G_t^0[\chi(\tau)] \text{ and} \\ X^1(t|\tau) &= G_t^1[\chi(\tau)]. \end{aligned}$$

And the difference between the factual operational state and counterfactual state, as estimated by the physics of the forecast model can be written

$$\Delta X(t|\tau) = X^1(t|\tau) - X^0(t|\tau). \quad (6.1)$$

Now for a sufficiently small time t , such that the model has significant skill and dynamical noise is low, this ΔX represents a good, and physically consistent, estimate of the difference between the factual and counterfactual worlds at time $\tau + t$. Hence if we then want to issue a successive forecast at time $\tau + t$, rather than simply using $\chi(\tau + t)$ to initialise both forecasts, we could use $\chi(\tau + t)$ for the operational forecast, and $\chi(\tau + t) - \Delta X(t|\tau)$ for the counterfactual forecast. As this routine is applied to several successive operational and counterfactual forecasts in a row, the differences between $\Delta X(t|\tau)$ and $\Delta X(t|\tau + t)$ should tend to a small (though non zero) value, as ΔX tends towards the ‘real’ difference between a balanced factual initial state and an analogue state in the counterfactual world. This difference between successive ΔX s won’t ever reach zero, since the difference between factual and counterfactual worlds at a certain time is dependent on the climate state at that same time. After performing this routine several times, this ΔX should tend towards the difference between assimilated factual and counterfactual initial states (ie. as would be obtained by assimilating the perturbations, described above).

The need to continually apply this adjustment to successive forecasts, perhaps a few days apart, is why this routine would only work in the case of an operational system. Its relative simplicity in comparison to the perturbed data assimilation approach, and physical consistency compared with the simple perturbation approach make it attractive. Conceptually, it is somewhat similar to the methodology developed by Wang et al. (195). They used separate long-running integrations of the same forecast model at different CO₂ levels to derive the perturbations applied to the initial conditions for their counterfactual simulations. The main differences are that i) this approach does not require costly separate simulations to determine the perturbations since the perturbations are developed through several successive forecasts, and ii) the perturbations generated here would be more closely linked to the actual state of the climate system at the time the forecasts are initialised. However, despite the advantages of this approach, it is still very likely that there would be technical challenges to address. For example, we would have to ensure

that errors in ΔX , possibly arising from dynamical noise or forecast error, do not grow between successive forecasts. If this happened, the factual and counterfactual initial states would move further and further apart, and any differences between the factual and counterfactual forecasts could then not be attributed to human influence because of the confounding error present. The shorter the time between successive runs of such an operational attribution system, the less likely for issues like this to occur. The other clear challenge would be determining what variables to perturb. Although it would be possible to perturb everything in the initial conditions, it might be more robust to only apply this adjustment to thermodynamic variables that have a physical basis for being perturbed (just as in the pseudo-global warming approach). Nevertheless, this approach is an intriguing prospect for robust operational attribution using a weather forecast system.

6.4.2 Expanding the scope of this work

This section explores various tests for the methodology used here that should be carried out in further work to more completely assess the robustness of the approach. These tests could be completed without any major changes to the methodology itself.

Alternative extremes As mentioned in the Limitations above, this thesis has focused on heatwaves. However, there are many other weather extremes that are of scientific and public interest. Hence, testing the robustness of the forecast-based approach to other extremes would be a natural next step to take. Given their significant coverage in the literature, a high precipitation event would be a good choice for such a test. I note that there would be potential additional considerations when examining a precipitation extreme compared to a heatwave. Firstly, the smaller spatial scale of such precipitation extremes means that these scales might have to be taken into account during the attribution step — what if the center of the extreme shifts in the counterfactual world (83)? Such shifts mean that, especially if linking the precipitation to flooding impacts, pooled catchments, rather than catchments on an individual basis, may have to be considered. The other significant difference is that precipitation forecasts are typically less skillful than temperature forecasts for lead times of more than a few days (311–316). This reduced skill (particularly during the 1–2 week period, where temperature forecasts are generally still good) means that shorter lead times may have to be used to ensure that the forecast model is still able to capture the extreme event within its ensemble — even if the forecast is reliable. Reducing the lead time would make addressing the forecast adjustment to the perturbed initial conditions even more important.

Alternative forecast models Another limitation discussed above was the use of a single model, ECMWF’s IFS. As such, testing the sensitivity of this approach to

the particular forecast model used would be an important step. It would make sense to begin with a forecast system that uses the same ocean model as the IFS, NEMO 3.4 in ORCA025Z75 configuration (211). This would allow bit-identical perturbations to be used, thus minimising differences that arise from experimental setup as opposed to those that we are interested in that arise from model choice. However, a shared ocean model is not an absolute necessity as one of the steps taken to produce the ocean state perturbations in chapter 4 was to interpolate the perturbations onto the ORCA025Z75 model grid — in theory any ocean grid (and thus any ocean model) could have been used. The Met Office’s medium-range MOGREPS-G and seasonal GloSea5 ensemble forecasting systems both use the NEMO ocean model on the same grid, and so may make good candidates for applying the forecast-based approach in an alternative model (305).

Incorporating perturbation uncertainty Including uncertainties associated with the estimation of the perturbed initial state in the forecast-based approach may be important, as discussed above and shown in previous work (67, 171). The main difficulty of including these uncertainties comes from the additional computational cost: if we were to simply repeat our experiments using climate model-derived estimates of these perturbations the cost would scale with the number of climate models used. In order to get a reasonable representation of the uncertainty, $O(10)$ model estimates would be required (as in 171). This would immediately increase the computational cost by a factor of 10 — possibly feasible for a single experiment, but much less desirable from an operational perspective. However, given the current operational ensemble prediction system at ECMWF already uses a 5-member ensemble of ocean analyses from which the 51 forecast ensemble members are initialised from, it is possible that we could use a similar ensemble of perturbations within a single forecast ensemble. There are a number of additional outstanding questions: i) how to choose the climate models from which these perturbations are estimated; ii) whether to give the climate models equal weight, or weight them based on some form of model evaluation (this is especially relevant for the latest generation of climate models, the CMIP6 ensemble, 178, 317); and iii) if using both observation- and climate model-based estimates, how to combine them?

6.4.3 Alternative applications

This thesis has largely focussed on the physical attribution of extreme weather events — which is an intrinsically backward-looking question. However, the approach explored here has potential to inform future projections of climate change as well. In this section, I discuss forward-looking applications of this work and also how it might further research into attribution and projection of the societal impacts arising from extreme weather, which is a rapidly developing field at the moment.

Projections of future extremes The question that this thesis has been concerned with answering is how human influence on the climate over the past century or so has affected the probability and severity of extreme events occurring in the present-day. This question is of considerable importance for the numerous reasons detailed in chapter 1. However, an arguably more policy-relevant question is that of how extreme weather events may change in the future — this is especially critical for adaptation planning (64). Providing information that is specific enough to be useful in a policy context often requires more granularity than coarse climate models are able to provide. Hence, statistical or dynamical downscaling is typically used in order to increase the utility of climate model simulations — for example in the United Kingdom Climate Projections (UKCP) reports (229, 230, 281, 318). However, given the structural errors in current climate models, this approach cannot be expected to provide entirely robust and reliable probabilistic information, especially on the scales that adaptation planners require. This was the reason why in 2015 Hazeleger et al. suggested that a complementary approach would be to construct ‘what if’ scenarios using high-resolution weather forecast models that would be able to provide the specific information required; especially given their position in current extreme weather hazard warning systems (83). This approach was called ‘Tales of future weather’ (173).

I suggest that the forecast-based approach developed in this thesis could be an attractive methodology for constructing such Tales. The idea would be to take a set of damaging historical extreme weather events (that were successfully forecast), perturb the forecast initial conditions exactly as done in chapter 4, and thus produce realisations of these events in a warmer world. These future forecasts could then be used to examine how future impacts might be worse than in the present, and thus how adaptation policies could be implemented in order to mitigate such impacts. I have already essentially done this future forecast experiment — though here they were used to test the linearity of the response, rather than for climate projection specifically. The methodology I have used to calculate the anthropogenic fingerprints to be removed from the forecast initial conditions could not just determine the estimated perturbation between pre-industrial and present-day climates, but between climates separated by specified levels of global warming (5). For instance, one could construct forecast-based Tales for policy-relevant future warming levels of 1.5, 2, 3 and 4 °C. One advantage that this forecast-based approach has over a storyline approach (eg. 82) in this context is that ensemble forecasts do not just reproduce the event as it unfolded, but also possible alternative realisations that may be even more extreme. This ‘ensemble-boosting’ aspect of such a forecast-based approach to extreme weather projections could help to ensure potential impacts are not underestimated as a result of limiting our view to the outcome that did occur by exploring the range of physically consistent possible outcomes (190).

One issue with using an optimal fingerprinting approach to estimating the perturbations

required is that it assumes that the pattern of global warming remains constant into the future, which is not certain to be the case, in particular for the higher levels of warming. The discussion above on incorporating perturbation uncertainty is relevant to this (319). For example, perturbations to particular levels of global warming could be derived from coupled climate models in addition to observations in order to more completely span the space of possible future patterns of warming. Another apparent issue with this forecast-based approach arises due to the reliance on historically damaging events. The length of the historical record means that regional coverage of such events will vary considerably. For example, while many regions may have experienced 1-in-100 to 1-in-1000 year heatwaves over the course of the historical record, many will not have. This could potentially leave these regions under-informed in terms of the risk from climate change exacerbated extremes. However, there are already-developed approaches to counter this issue in the literature. One of the most relevant is the UNSEEN approach (218, 231). Briefly, this approach uses seasonal ensemble hindcasts to considerably increase the effective sample size of such events within the historical record. I suggest that one way in which useful Tales could be constructed would be to not only look for damaging events in the historical record, but also within the seasonal and medium-range hindcast ensembles that are available. Such ‘unseen’ events could then be re-forecast within a future climate to explore how they may change under continued global warming. A disadvantage of this approach is that such unseen events were not necessarily successfully forecast (which is one of the key features of the forecast-based approaches explored in this thesis), though model fidelity and reliability could be validated in other ways in this case (320).

Impact assessment Impact attribution is a rapidly growing field of research (42, 321). Linking the attributable physical changes due to climate change to the socio-economic impacts can be extremely powerful as a communication tool and as a way to drive policy (46). Some examples of this linkage include economic damages from hurricanes (322, 323) and mortality from heatwaves (94, 95, 189). Despite its clear importance, impact attribution has only taken off recently, possibly due to the difficulties that arise as a result of the additional uncertainties (and non-linearities) associated with linking physical to socioeconomic impacts. This section will not be a lengthy discussion, but I suggest that there are a number of reasons why forecast-based approaches could complement and advance current approaches.

One key reason, that I have mentioned previously, is that weather forecasts are already built into the modelling chains used to assess risk from extreme weather by combining physical hazard and vulnerability information (83). Given how important the vulnerability aspect of extreme weather risk is (225, 324), using models already familiar to those with relevant expertise is a significant advantage. In addition to this familiarity advantage, the fact that weather forecasts are already key components of many well-

validated impact prediction systems (for example, the GloFAS flood warning system [325](#)) means that impact attribution may be able to be carried out with very little technical work — simply by switching operational weather forecasts for counterfactual ones. For example, Wilkinson et al. ([326](#)) developed a scheme for translating weather forecasts into damages. Such a scheme could essentially be used ‘as is’ to generate estimates of attributable damages to climate change. Trustworthy estimates of attributable damages could further support litigation relating to increased extreme weather risk. The arguments that I have made at length for forecast-based approaches in this thesis are also very relevant here: significantly increased resolution, well-established reliability, implicit and explicit model validation and event specificity. This is certainly not to say that climate models cannot provide useful information about general impacts arising from climate change (for example, through the storyline framework), but for assessing risks from many types of extreme weather, I argue that weather forecast models would be a more robust tool ([76](#)).

6.5 Concluding remarks

Appendices

A

Resources

Examples of output from the thesis besides research:

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

Media articles concerning research from this thesis

- CarbonBrief
- Science article

B

Appendix 2

Quote

— author

References

¹S. Arrhenius, “On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground”, *Philosophical Magazine and Journal of Science Series 5*, 237–276 (1896) (pp. 3, 21).

Widely recognised as the first paper to quantify the contribution of carbon dioxide to the greenhouse effect. However, it is predated by the discovery of the effect of carbon dioxide and water vapour on atmospheric temperatures, which is attributed to Eunice Newton Foote in 1856.

²N. L. Bindoff, P. A. Stott, K. M. AchutaRao, M. R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I. I. Mokhov, J. Overland, J. Perlitz, R. Sebbari, and X. Zhang, “Detection and Attribution of Climate Change: from Global to Regional”, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013), pp. 867–952 (p. 3).

³V. Eyring, N. Gillett, K. Achuta Rao, R. Barimalala, M. Barreiro Parrillo, N. Bellouin, C. Cassou, P. Durack, Y. Kosaka, S. McGregor, S. Min, O. Morgenstern, and Y. Sun, “Human Influence on the Climate System”, 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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 423–552 (pp. 3, 122).

⁴P. A. Stott and J. F. B. Mitchell, “Detection and attribution of climate change”, *Weather, wea.4027* (2021) (p. 3).

⁵K. Hasselmann, “Optimal Fingerprints for the Detection of Time-dependent Climate Change”, *Journal of Climate 6, 1957–1971* (1993) (pp. 3, 8, 69, 70, 72, 129).

An paper of exceptional importance that laid the foundations for modern approaches to detection and attribution of climate change, despite its somewhat opaque notation.

⁶K. Hasselmann, “Multi-pattern fingerprint method for detection and attribution of climate change”, *Climate Dynamics 13, 601–611* (1997) (pp. 3, 8, 69).

⁷M. R. Allen and S. F. B. Tett, “Checking for model consistency in optimal fingerprinting”, *Climate Dynamics 15, 419–434* (1999) (p. 3).

⁸M. R. Allen and P. A. Stott, “Estimating signal amplitudes in optimal fingerprinting, part I: theory”, *Climate Dynamics 21, 477–491* (2003) (p. 3).

- ⁹P. A. Stott, M. R. Allen, and G. S. Jones, “Estimating signal amplitudes in optimal fingerprinting. Part II: application to general circulation models”, *Climate Dynamics* **21**, 493–500 (2003) (p. 3).
- ¹⁰G. C. Hegerl, K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, “Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change”, *Climate Dynamics* **13**, 613–634 (1997) (p. 3).
- ¹¹P. A. Stott, S. F. B. Tett, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, “Attribution of twentieth century temperature change to natural and anthropogenic causes”, *Climate Dynamics* **17**, 1–21 (2001) (p. 3).
- ¹²N. P. Gillett, M. Kirchmeier-Young, A. Ribes, H. Shiogama, G. C. Hegerl, R. Knutti, G. Gastineau, J. G. John, L. Li, L. Nazarenko, N. Rosenbloom, Ø. Seland, T. Wu, S. Yukimoto, and T. Ziehn, “Constraining human contributions to observed warming since the pre-industrial period”, *Nature Climate Change* **11**, 207–212 (2021) (p. 3).
- ¹³N. P. Gillett, D. A. Stone, P. A. Stott, T. Nozawa, A. Y. Karpechko, G. C. Hegerl, M. F. Wehner, and P. D. Jones, “Attribution of polar warming to human influence”, *Nature Geoscience* **1**, 750–754 (2008) (p. 3).
- ¹⁴C. J. W. Bonfils, B. D. Santer, J. C. Fyfe, K. Marvel, T. J. Phillips, and S. R. H. Zimmerman, “Human influence on joint changes in temperature, rainfall and continental aridity”, *Nature Climate Change* **10**, 726–731 (2020) (pp. 3, 22).
- ¹⁵X. Gu, Q. Zhang, J. Li, V. P. Singh, J. Liu, P. Sun, and C. Cheng, “Attribution of Global Soil Moisture Drying to Human Activities: A Quantitative Viewpoint”, *Geophysical Research Letters* **46**, 2573–2582 (2019) (pp. 3, 22).
- ¹⁶P. A. Stott, R. T. Sutton, and D. M. Smith, “Detection and attribution of Atlantic salinity changes”, *Geophysical Research Letters* **35**, 10.1029/2008GL035874 (2008) (p. 3).
- ¹⁷N. Christidis and P. A. Stott, “Changes in the geopotential height at 500 hPa under the influence of external climatic forcings”, *Geophysical Research Letters* **42**, 10, 798–10, 806 (2015) (pp. 3, 76).
- ¹⁸S. Pahl, S. Sheppard, C. Boomsma, and C. Groves, “Perceptions of time in relation to climate change”, *WIREs Climate Change* **5**, 375–388 (2014) (p. 3).
- ¹⁹M. Allen, “Liability for climate change”, *Nature* **421**, 891–892 (2003) (pp. 3–7, 42). **A seminal paper that introduces the idea of extreme event attribution, including the litigative motivation behind such attribution, and a proposed attribution methodology, grounded in probabilistic concepts from epidemiology.**
- ²⁰M. 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) (pp. 3, 5). **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.**

- ²¹T. Knutson, J. Kossin, C. Mears, J. Perlitz, and M. Wehner, “Detection and attribution of climate change”, in *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, edited by D. Wuebbles, D. Fahey, K. Hibbard, D. Dokken, B. Stewart, and T. Maycock (U.S. Global Change Research Program, Washington, DC, USA, 2017), pp. 114–132 (p. 3).
- ²²Y. Imada, M. Watanabe, H. Kawase, H. Shiogama, and M. Arai, “The July 2018 High Temperature Event in Japan Could Not Have Happened without Human-Induced Global Warming”, *SOLA 15A*, 8–12 (2019) (p. 3).
- ²³A. R. Solow, “Extreme weather, made by us?”, *Science* 349, 1444–1445 (2015) (p. 3).
- ²⁴F. E. Otto, “Attribution of Weather and Climate Events”, *Annual Review of Environment and Resources* 42, 627–646 (2017) (pp. 3, 12).
- ²⁵P. A. Stott, D. A. Stone, and M. R. Allen, “Human contribution to the European heatwave of 2003”, *Nature* 432, 610–614 (2004) (pp. 3, 7, 31, 47, 53, 57, 58, 64, 100). **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.**
- ²⁶R. Dole, M. Hoerling, J. Perlitz, J. Eischeid, P. Pegion, T. Zhang, X. W. Quan, T. Xu, and D. Murray, “Was there a basis for anticipating the 2010 Russian heat wave?”, *Geophysical Research Letters* 38, 10.1029/2010GL046582 (2011) (pp. 3, 4).
- ²⁷F. E. L. Otto, N. Massey, G. J. van Oldenborgh, R. G. Jones, and M. R. Allen, “Reconciling two approaches to attribution of the 2010 Russian heat wave”, *Geophysical Research Letters* 39, n/a–n/a (2012) (p. 4).
- ²⁸S. 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) (pp. 4, 86).
- ²⁹P. A. Stott, M. Allen, N. Christidis, R. M. Dole, M. Hoerling, C. Huntingford, P. Pall, J. Perlitz, and D. Stone, “Attribution of Weather and Climate-Related Events”, in *Climate Science for Serving Society: Research, Modeling and Prediction Priorities*, edited by G. R. Asrar and J. W. Hurrell (Springer Netherlands, Dordrecht, 2013), pp. 307–337 (p. 4).
- ³⁰P. A. Stott, N. Christidis, F. E. L. Otto, Y. Sun, J.-P. Vanderlinden, G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou, and F. W. Zwiers, “Attribution of extreme weather and climate-related events”, *Wiley Interdisciplinary Reviews: Climate Change* 7, 23–41 (2016) (p. 4).
- ³¹F. E. L. Otto, G. Jan van Oldenborgh, J. Eden, P. A. Stott, D. J. Karoly, and M. R. Allen, “The attribution question”, *Nature Climate Change* 6, 813–816 (2016) (p. 4).
- ³²D. 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) (pp. 4, 6).
- ³³D. R. Easterling, K. E. Kunkel, M. F. Wehner, and L. Sun, “Detection and attribution of climate extremes in the observed record”, *Weather and Climate Extremes, Observed and Projected (Longer-term) Changes in Weather and Climate Extremes* 11, 17–27 (2016) (p. 4).

- ³⁴a. M. National Academies of Sciences Engineering, *Attribution of Extreme Weather Events in the Context of Climate Change* (National Academies Press, Washington, D.C., 2016) (pp. 4, 18, 63, 64).
- Possibly the most thorough review of methodologies used to carry out extreme event attribution. Although the field is developing so rapidly that it is starting to become a little outdated, I have found it an invaluable reference throughout this thesis.**
- ³⁵D. A. Stone and M. R. Allen, “The end-to-end attribution problem: From emissions to impacts”, *Climatic Change* **71**, 303–318 (2005) (pp. 5, 53, 78, 100).
- ³⁶R. 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. 5).
- ³⁷S. Marjanac, L. Patton, and J. Thornton, “Acts of God, human influence and litigation”, *Nature Geoscience* **10**, 616–619 (2017) (p. 5).
- ³⁸E. 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. 5).
- ³⁹E. A. Lloyd and N. Oreskes, “Climate Change Attribution: When Is It Appropriate to Accept New Methods?”, *Earth’s Future* **6**, 311–325 (2018) (p. 5).
- ⁴⁰E. 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. 5).
- ⁴¹E. A. Lloyd and T. G. Shepherd, “Climate change attribution and legal contexts: evidence and the role of storylines”, *Climatic Change* **167**, 28 (2021) (pp. 5, 18).
- ⁴²M. Burger, J. Wentz, and R. Horton, “The Law and Science of Climate Change Attribution”, *Columbia Journal of Environmental Law* **45**, 10.7916/cjel.v45i1.4730 (2020) (pp. 5, 130).
- 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.**
- ⁴³S. 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. 5).
- ⁴⁴M. F. Wehner and K. A. Reed, “Operational extreme weather event attribution can quantify climate change loss and damages”, *PLOS Climate* **1**, e0000013 (2022) (pp. 5, 19, 80).
- ⁴⁵M. Mace and R. Verheyen, “Loss, Damage and Responsibility after COP21: All Options Open for the Paris Agreement”, *Review of European, Comparative & International Environmental Law* **25**, 197–214 (2016) (p. 5).
- ⁴⁶B. 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) (pp. 5, 130).
- ⁴⁷F. E. L. Otto, R. B. Skeie, J. S. Fuglestvedt, T. Berntsen, and M. R. Allen, “Assigning historic responsibility for extreme weather events”, *Nature Climate Change* **7**, 757–759 (2017) (p. 5).

As far as I am aware, the first attribution study that begins to explore the question of how different collections of people (in this case, countries) are responsible for extreme weather events to different extents. It has been followed up by Frame et al. (2020) and Lott et al. (2021), but I expect the base idea presented could gain considerable traction over the next few years, for example in quantifying to what degree individual companies are responsible for damaging weather extremes.

- ⁴⁸F. C. Lott, A. Ciavarella, J. J. Kennedy, A. D. King, P. A. Stott, S. F. B. Tett, and D. Wang, “Quantifying the contribution of an individual to making extreme weather events more likely”, *Environmental Research Letters* **16**, 104040 (2021) (p. 5).
- ⁴⁹T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013) (p. 6).
- ⁵⁰IPCC, *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change*, edited by V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (2018) (p. 6).
- ⁵¹IPCC, *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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, Vol. In Press (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021) (p. 6).
- ⁵²A. Fouillet, G. Rey, F. Laurent, G. Pavillon, S. Bellec, C. Guihenneuc-Jouyaux, J. Clavel, E. Jouglard, and D. Hémon, “Excess mortality related to the August 2003 heat wave in France”, *International Archives of Occupational and Environmental Health* **80**, 16–24 (2006) (p. 6).
- ⁵³J. 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. 6).
- ⁵⁴S. Osaka and R. Bellamy, “Natural variability or climate change? Stakeholder and citizen perceptions of extreme event attribution”, *Global Environmental Change* **62**, 102070 (2020) (p. 6).
- ⁵⁵M. Bergquist, A. Nilsson, and P. W. Schultz, “Experiencing a Severe Weather Event Increases Concern About Climate Change”, *Frontiers in Psychology* **10** (2019) (p. 6).
- ⁵⁶L. Gärtner and H. Schoen, “Experiencing climate change: revisiting the role of local weather in affecting climate change awareness and related policy preferences”, *Climatic Change* **167**, 31 (2021) (p. 6).

- ⁵⁷E. M. Fischer, S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär, “Soil Moisture–Atmosphere Interactions during the 2003 European Summer Heat Wave”, *Journal of Climate* **20**, 5081–5099 (2007) (pp. 6, 21, 22).
- ⁵⁸E. M. Fischer, S. I. Seneviratne, D. Lüthi, and C. Schär, “Contribution of land-atmosphere coupling to recent European summer heat waves”, *Geophysical Research Letters* **34**, 10.1029/2006GL029068 (2007) (pp. 6, 21, 80).
- ⁵⁹K. Wehrli, B. P. Guillo, M. Hauser, M. Leclair, and S. I. Seneviratne, “Identifying Key Driving Processes of Major Recent Heat Waves”, *Journal of Geophysical Research: Atmospheres* **124**, 11746–11765 (2019) (pp. 6, 21).
- ⁶⁰J. 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. 7, 11, 16, 47).
- ⁶¹S. 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. 7–9, 17, 57, 58, 64, 122).
An extremely valuable paper that describes the probabilistic methodology followed by the most prominent extreme event attribution collaboration. This work is taken as the reference text for how probabilistic attribution is carried out in the present day.
- ⁶²O. Bellprat, V. Guemas, F. Doblas-Reyes, and M. G. Donat, “Towards reliable extreme weather and climate event attribution”, *Nature Communications* **10**, 1732 (2019) (pp. 7, 11, 17, 35, 37, 42, 47, 117).
- ⁶³N. 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. 7, 8, 44, 47).
- ⁶⁴L. J. Harrington, K. L. Ebi, D. J. Frame, and F. E. L. Otto, “Integrating attribution with adaptation for unprecedented future heatwaves”, *Climatic Change* **172**, 1–7 (2022) (pp. 7, 64, 129).
- ⁶⁵M. 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. 7, 12, 42, 64, 116).
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 CO₂ concentrations.
- ⁶⁶G. J. Van Oldenborgh, “How unusual was autumn 2006 in Europe?”, *Climate of the Past* **3**, 659–668 (2007) (pp. 8, 10, 64).
An important work that pioneers the observational trend-based approach to the probabilistic framework for extreme event attribution. The author demonstrates the groundwork for a methodology that becomes integral to the World Weather Attribution project attribution protocol, which I take as the standard way of carrying out probabilistic event attribution in this thesis.

- ⁶⁷P. 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. 9, 10, 47, 63, 64, 81, 89, 120, 121, 128).
The first paper that made use of explicit counterfactual, rather than transient, simulations to estimate the anthropogenic contribution to flood risk in autumn 2000. The author’s modified boundary condition approach to creating naturalised counterfactual simulations of season has been replicated in many extreme event attribution studies since, and serves as inspiration for the approach ultimately taken in this thesis.
- ⁶⁸G. J. van Oldenborgh, K. van der Wiel, S. Kew, S. Philip, F. Otto, R. Vautard, A. King, F. Lott, J. Arrighi, R. Singh, and M. van Aalst, “Pathways and pitfalls in extreme event attribution”, *Climatic Change* **166**, 13 (2021) (pp. 9, 10, 18).
A study that presented a novel attribution methodology based on event probabilities inferred from time-slices of historical coupled model simulations. Of note due to the use of this methodology by the UK Met Office attribution group, among others, since.
- ⁶⁹S. C. Lewis and D. J. Karoly, “Anthropogenic contributions to Australia’s record summer temperatures of 2013”, *Geophysical Research Letters* **40**, 3705–3709 (2013) (p. 9).
A study that presented a novel attribution methodology based on event probabilities inferred from time-slices of historical coupled model simulations. Of note due to the use of this methodology by the UK Met Office attribution group, among others, since.
- ⁷⁰D. 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**, 10.1186/1476-069X-9-37 (2010) (pp. 10, 29).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.
- ⁷¹K. E. Trenberth, J. T. Fasullo, and T. G. Shepherd, “Attribution of climate extreme events”, *Nature Climate Change* **5**, 725–730 (2015) (pp. 11, 12, 14, 64).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.
- ⁷²M.-E. Demory, P. L. Vidale, M. J. Roberts, P. Berrisford, J. Strachan, R. Schiemann, and M. S. Mizielinski, “The role of horizontal resolution in simulating drivers of the global hydrological cycle”, *Climate Dynamics* **42**, 2201–2225 (2014) (pp. 11, 16).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.
- ⁷³R. Schiemann, M.-E. Demory, L. C. Shaffrey, J. Strachan, P. L. Vidale, M. S. Mizielinski, M. J. Roberts, M. Matsueda, M. F. Wehner, and T. Jung, “The Resolution Sensitivity of Northern Hemisphere Blocking in Four 25-km Atmospheric Global Circulation Models”, *Journal of Climate* **30**, 337–358 (2017) (pp. 11, 16).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.
- ⁷⁴J. Derrington, K. Strommen, and F. Fabiano, “How well does CMIP6 capture the dynamics of Euro-Atlantic weather regimes, and why?”, *Weather and Climate Dynamics Discussions*, 1–41 (2021) (p. 11).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.
- ⁷⁵T. N. Palmer, “A Nonlinear Dynamical Perspective on Climate Prediction”, *Journal of Climate* **12**, 575–591 (1999) (pp. 11, 64, 80).
An important perspective piece that lucidly sets out the practical justification for the storyline approach to attribution with examples: the authors suggest that we should aim to separate out what we can know with confidence from what we cannot given the current state of science. In extreme event attribution, this generally reduces to examining the influence of large-scale thermodynamic changes on a particular event, given specific dynamic or synoptic drivers.

- ⁷⁶T. 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. 11, 17, 42, 47, 64, 65, 80, 131).
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.
- ⁷⁷O. 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. 11, 17, 42, 47).
- ⁷⁸T. G. Shepherd, “A Common Framework for Approaches to Extreme Event Attribution”, *Curr Clim Change Rep* **2**, 28–38 (2016) (pp. 11–13, 42, 47, 48, 50, 53, 54, 57–59, 64).
A clear discussion of the different approaches to attribution that sets out the case for increased use of the storyline approach. However, argues that the best framing and approach depends upon the question being asked.
- ⁷⁹G. Masato, B. J. Hoskins, and T. Woollings, “Winter and Summer Northern Hemisphere Blocking in CMIP5 Models”, *Journal of Climate* **26**, 7044–7059 (2013) (pp. 12, 64).
- ⁸⁰T. 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) (pp. 13, 58).
- ⁸¹L. 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. 14, 43, 47, 48, 65).
- ⁸²A. S. Benítez, H. Goessling, F. Pithan, T. Semmler, and T. Jung, “The July 2019 European heatwave in a warmer climate: Storyline scenarios with a coupled model using spectral nudging”, *Journal of Climate* **-1**, 1–51 (2022) (pp. 14, 65, 129).
- ⁸³N. Schaller, J. Sillmann, M. Müller, R. Haarsma, W. Hazeleger, T. J. Hegdahl, T. Kelder, G. van den Oord, A. Weerts, and K. Whan, “The role of spatial and temporal model resolution in a flood event storyline approach in western Norway”, *Weather and Climate Extremes* **29**, 100259 (2020) (pp. 15, 19, 127, 129, 130).
- ⁸⁴A. 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) (pp. 15, 89).
- ⁸⁵E. N. Lorenz, “Deterministic Nonperiodic Flow”, *Journal of the Atmospheric Sciences*, **10**.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2 (1963) (p. 15).
A work beloved by climate scientists and meteorologists in need of a low-order chaotic system.

- ⁸⁶H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J. N. Thépaut, “The ERA5 global reanalysis”, *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020) (pp. 16, 45, 67, 68, 124).
- ⁸⁷K. D. Williams, A. Bodas-Salcedo, M. Déqué, S. Fermepin, B. Medeiros, M. Watanabe, C. Jakob, S. A. Klein, C. A. Senior, and D. L. Williamson, “The Transpose-AMIP II Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models”, *Journal of Climate* **26**, 3258–3274 (2013) (p. 17).
- ⁸⁸T. N. Palmer, F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, “TOWARD SEAMLESS PREDICTION Calibration of Climate Change Projections Using Seasonal Forecasts”, *Bulletin of the American Meteorological Society* **89**, 459–470 (2008) (p. 17).
- ⁸⁹M. Matsueda, A. Weisheimer, and T. N. Palmer, “Calibrating Climate Change Time-Slice Projections with Estimates of Seasonal Forecast Reliability”, *Journal of Climate* **29**, 3831–3840 (2016) (p. 17).
- ⁹⁰A. Weisheimer, N. Schaller, C. O’reilly, D. A. MacLeod, and T. Palmer, “Atmospheric seasonal forecasts of the twentieth century: multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution”, *Quarterly Journal of the Royal Meteorological Society Q. J. R. Meteorol. Soc* **143**, 917–926 (2017) (p. 17).
- ⁹¹F. C. Lott and P. A. Stott, “Evaluating Simulated Fraction of Attributable Risk Using Climate Observations”, *Journal of Climate* **29**, 4565–4575 (2016) (p. 17).
- ⁹²A. Weisheimer and T. N. Palmer, “On the reliability of seasonal climate forecasts”, *Journal of The Royal Society Interface* **11**, 20131162 (2014) (pp. 17, 117).
- ⁹³R. Vautard, M. van Aalst, O. Boucher, A. Drouin, K. Haustein, F. Kreienkamp, G. J. van Oldenborgh, F. E. L. Otto, A. Ribes, Y. Robin, M. Schneider, J.-M. Soubeyroux, P. Stott, S. I. Seneviratne, M. M. Vogel, and M. Wehner, “Human contribution to the record-breaking June and July 2019 heatwaves in Western Europe”, *Environmental Research Letters* **15**, 094077 (2020) (p. 18).
- ⁹⁴D. Mitchell, C. Heaviside, S. Vardoulakis, C. Huntingford, G. Masato, B. P. Guillod, P. Frumhoff, A. Bowery, D. Wallom, and M. Allen, “Attributing human mortality during extreme heat waves to anthropogenic climate change”, *Environmental Research Letters*, **10**. 1088/1748-9326/11/7/074006 (2016) (pp. 19, 130).
- ⁹⁵Y. T. E. Lo, D. M. Mitchell, R. Thompson, E. O’Connell, and A. Gasparrini, “Estimating heat-related mortality in near real time for national heatwave plans”, *Environmental Research Letters* **17**, 024017 (2022) (pp. 19, 130).

- ⁹⁶B. Schauberger, S. Archontoulis, A. Arneth, J. Balkovic, P. Ciais, D. Deryng, J. Elliott, C. Folberth, N. Khabarov, C. Müller, T. A. M. Pugh, S. Rolinski, S. Schaphoff, E. Schmid, X. Wang, W. Schlenker, and K. Frieler, “Consistent negative response of US crops to high temperatures in observations and crop models”, *Nature Communications* **8**, 13931 (2017) (p. 19).
- ⁹⁷A. P. Davis, D. Mieulet, J. Moat, D. Sarmu, and J. Haggar, “Arabica-like flavour in a heat-tolerant wild coffee species”, *Nature Plants* **7**, 413–418 (2021) (p. 19).
- ⁹⁸*Mapped: How climate change affects extreme weather around the world*, Feb. 2021 (p. 19).
- ⁹⁹E. B. Garriott, *Long-range Weather Forecasts* (U.S. Government Printing Office, 1904) (p. 20).
- ¹⁰⁰L.-A. Kautz, O. Martius, S. Pfahl, J. G. Pinto, A. M. Ramos, P. M. Sousa, and T. Woollings, “Atmospheric blocking and weather extremes over the Euro-Atlantic sector – a review”, *Weather and Climate Dynamics* **3**, 305–336 (2022) (p. 20).
- ¹⁰¹T. Woollings, D. Barriopedro, J. Methven, S.-W. Son, O. Martius, B. Harvey, J. Sillmann, A. R. Lupo, and S. Seneviratne, “Blocking and its Response to Climate Change”, *Current Climate Change Reports* **4**, 287–300 (2018) (p. 20).
- ¹⁰²A. Schneidereit, S. Schubert, P. Vargin, F. Lunkeit, X. Zhu, D. H. W. Peters, and K. Fraedrich, “Large-Scale Flow and the Long-Lasting Blocking High over Russia: Summer 2010”, *Monthly Weather Review* **140**, 2967–2981 (2012) (p. 20).
- ¹⁰³R. J. Greatbatch, G. Gollan, T. Jung, and T. Kunz, “Tropical origin of the severe European winter of 1962/1963”, *Quarterly Journal of the Royal Meteorological Society* **141**, 153–165 (2015) (p. 20).
- ¹⁰⁴G. Di Capua, S. Sparrow, K. Kornhuber, E. Rousi, S. Osprey, D. Wallom, B. van den Hurk, and D. Coumou, “Drivers behind the summer 2010 wave train leading to Russian heatwave and Pakistan flooding”, *npj Climate and Atmospheric Science* **4**, 1–14 (2021) (p. 20).
- ¹⁰⁵H. Wang and D. Luo, “Summer Russian heat waves and their links to Greenland’s ice melt and sea surface temperature anomalies over the North Atlantic and the Barents–Kara Seas”, *Environmental Research Letters* **15**, 114048 (2020) (p. 20).
- ¹⁰⁶E. Nabizadeh, P. Hassanzadeh, D. Yang, and E. A. Barnes, “Size of the Atmospheric Blocking Events: Scaling Law and Response to Climate Change”, *Geophysical Research Letters* **46**, 13488–13499 (2019) (p. 20).
- ¹⁰⁷D. Steinfeld, M. Sprenger, U. Beyerle, and S. Pfahl, “Response of moist and dry processes in atmospheric blocking to climate change”, *Environmental Research Letters* **17**, 084020 (2022) (pp. 20, 21).
- ¹⁰⁸M. 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. 21, 44, 107).
- ¹⁰⁹N. J. Leach, A. Weisheimer, M. R. Allen, and T. Palmer, “Forecast-based attribution of a winter heatwave within the limit of predictability”, *Proceedings of the National Academy of Sciences* **118**, 10.1073/pnas.2112087118 (2021) (pp. 21, 65, 75, 76, 80).

- ¹¹⁰M. R. Allen and W. J. Ingram, “Constraints on future changes in climate and the hydrologic cycle”, *Nature* **419**, 228–232 (2002) (pp. 21, 76).
- ¹¹¹R. M. Horton, J. S. Mankin, C. Lesk, E. Coffel, and C. Raymond, “A Review of Recent Advances in Research on Extreme Heat Events”, *Current Climate Change Reports* **2**, 242–259 (2016) (pp. 21, 124).
- ¹¹²L. R. V. Zeppetello, D. S. Battisti, and M. B. Baker, “The Physics of Heat Waves: What Causes Extremely High Summertime Temperatures?”, *Journal of Climate* **35**, 2231–2251 (2022) (p. 21).
- ¹¹³M. M. Vogel, J. Zscheischler, and S. I. Seneviratne, “Varying soil moisture–atmosphere feedbacks explain divergent temperature extremes and precipitation projections in central Europe”, *Earth System Dynamics* **9**, 1107–1125 (2018) (p. 21).
- ¹¹⁴D. G. Miralles, A. J. Teuling, C. C. van Heerwaarden, and J. Vilà-Guerau de Arellano, “Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation”, *Nature Geoscience* **7**, 345–349 (2014) (p. 21).
- ¹¹⁵P. M. Sousa, D. Barriopedro, R. García-Herrera, C. Ordóñez, P. M. M. Soares, and R. M. Trigo, “Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves”, *Communications Earth & Environment* **1**, 1–13 (2020) (p. 21).
- ¹¹⁶K. Marvel, B. I. Cook, C. J. W. Bonfils, P. J. Durack, J. E. Smerdon, and A. P. Williams, “Twentieth-century hydroclimate changes consistent with human influence”, *Nature* **569**, 59–65 (2019) (p. 22).
- ¹¹⁷C. Johnston, *Heatwave temperatures may top 45C in southern Europe*, Aug. 2018 (p. 27).
- ¹¹⁸NESDIS, *Record Summer Heat Bakes Europe*, Aug. 2018 (p. 27).
- ¹¹⁹F. 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. 27).
- ¹²⁰J. Watts, *Wildfires rage in Arctic Circle as Sweden calls for help*, July 2018 (p. 27).
- ¹²¹Publico, *Nueve fallecidos por la ola de calor en España*, Aug. 2018 (p. 27).
- ¹²²C. Harris, *Heat, hardship and horrible harvests: Europe’s drought explained*, Aug. 2018 (p. 27).
- ¹²³World Weather Attribution, *Heatwave in northern Europe, summer 2018*, July 2018 (p. 27).
- ¹²⁴Press Office, *Chance of summer heatwaves now thirty times more likely*, Dec. 2018 (p. 27).
- ¹²⁵M. 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. 27).
- ¹²⁶J. 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. 28, 44, 45).

- ¹²⁷J. 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. 29).
- ¹²⁸N. 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. 29, 47, 48).
- ¹²⁹A. 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. 29, 47, 48, 57, 75, 81, 122).
A model description paper that sets out a modelling and evaluation framework for the Met Office’s attribution system. I highlight it because of its provision of a very clear and rigorous discussion of the various possible ways in which attribution analyses are often conditioned.
- ¹³⁰N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, “Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century”, *Journal of Geophysical Research* **108**, 4407 (2003) (pp. 29, 70, 75).
- ¹³¹K. 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) (pp. 29, 51, 120).
- ¹³²R. 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. 29).
- ¹³³D. 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. 29).
- ¹³⁴M. 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. 29).
- ¹³⁵E. 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. 29).

- ¹³⁶G. 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. 29).
- ¹³⁷A. M. Thomson, K. V. Calvin, S. J. Smith, G. P. Kyle, A. Volke, P. Patel, S. Delgado-Arias, B. Bond-Lamberty, M. A. Wise, L. E. Clarke, and J. A. Edmonds, “RCP4.5: a pathway for stabilization of radiative forcing by 2100”, *Climatic Change* **109**, 77 (2011) (p. 29).
- ¹³⁸R. 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) (pp. 29, 44).
- ¹³⁹N. 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. 29, 47, 48, 75, 81).
- ¹⁴⁰K. 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) (pp. 30, 69, 78).
- ¹⁴¹C. 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**, 10.1029/2019JD032361 (2021) (pp. 30, 69).
- ¹⁴²N. 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. 30, 86).
- ¹⁴³S. 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. 30, 58).
- ¹⁴⁴G. 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. 31).
- ¹⁴⁵J. 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. 31, 46, 95).
- ¹⁴⁶E. 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. 36, 47, 57, 110).
- ¹⁴⁷J. 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. 36, 110).

- ¹⁴⁸M. 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. 37, 44, 68).
- ¹⁴⁹H. 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 CO₂ concentrations increase extreme event risk in a 1.5 °C world”, *Nature Climate Change* **8**, 604–608 (2018) (pp. 41, 50).
- ¹⁵⁰T. 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. 42).
- ¹⁵¹M. Hulme, “Attributing weather extremes to ‘climate change’: A review”, *Progress in Physical Geography* **38**, 499–511 (2014) (p. 42).
- ¹⁵²S. 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. 42).
- ¹⁵³S. 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. 42).
- ¹⁵⁴E. 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. 43, 47, 116, 117).
- ¹⁵⁵P. 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. 43, 58, 59, 65, 118, 119).
An underappreciated and unassuming paper which I believe is the earliest work that utilizes a forecast-based approach similar to that explored in this thesis. In this study, the authors use perturbed CO₂ and initial ocean conditions (estimated using separate free-running coupled integrations at different CO₂ levels) to quantify the human contribution since 1960 to the record hot 2014 springtime temperatures in Australia.
- ¹⁵⁶P. 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. 43, 59, 65, 118).
- ¹⁵⁷P. 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 CO₂ on the Record Fire Weather in Eastern Australia in February 2017”, *Bulletin of the American Meteorological Society* **100**, S111–S117 (2019) (pp. 43, 58, 59, 65, 118).
- ¹⁵⁸K. 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) (pp. 43, 117, 124).
- ¹⁵⁹G. M. Lackmann, “Hurricane Sandy before 1900 and after 2100”, *Bulletin of the American Meteorological Society* **96**, 547–560 (2015) (pp. 43, 116, 117).

- ¹⁶⁰A. 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) (pp. 43, 125).
- ¹⁶¹P. 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. 44).
- ¹⁶²M. 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. 44, 93, 94, 107).
- ¹⁶³N. 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. 44, 55).
- ¹⁶⁴P. 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) (pp. 44, 68).
- ¹⁶⁵R. J. Dunn, K. M. Willett, P. W. Thorne, E. V. Woolley, I. Durre, A. Dai, D. E. Parker, and R. S. Vose, “HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011”, *Climate of the Past* **8**, 1649–1679 (2012) (p. 45).
- ¹⁶⁶R. J. Dunn, K. M. Willett, C. P. Morice, and D. E. Parker, “Pairwise homogeneity assessment of HadISD”, *Climate of the Past* **10**, 1501–1522 (2014) (p. 45).
- ¹⁶⁷R. J. H. Dunn, K. M. Willett, D. E. Parker, and L. Mitchell, “Expanding HadISD: quality-controlled, sub-daily station data from 1931”, *Geoscientific Instrumentation, Methods and Data Systems* **5**, 473–491 (2016) (p. 45).
- ¹⁶⁸A. Smith, N. Lott, and R. Vose, “The integrated surface database: Recent developments and partnerships”, *Bulletin of the American Meteorological Society* **92**, 704–708 (2011) (p. 45).
- ¹⁶⁹ECMWF, *IFS evolution - summary of cycle 45R1*, 2018 (p. 46).
- ¹⁷⁰ECMWF, *ECMWF Confluence- M-climate, the ENS Model Climate*, 2018 (p. 46).
- ¹⁷¹S. 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) (pp. 47, 70, 128).
- ¹⁷²P. 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. 47, 57, 95).
- ¹⁷³W. 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**, 10 . 1038/NCLIMATE2450 (2015) (pp. 48, 66, 81, 82, 129).

A perspective paper suggesting that rather than rely on coarse climate model ensembles and statistical downscaling for climate projection, we should instead employ weather forecast models to provide simulations of high-impact weather in counterfactual future scenarios. The forecast-based approach presented in this thesis could be leveraged to do exactly this.

- ¹⁷⁴D. 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. 50, 59, 63, 122).
- ¹⁷⁵M. 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. 50).
- ¹⁷⁶W. L. Gates, K. H. Cook, and M. E. Schlesinger, “Preliminary analysis of experiments on the climatic effects of increased CO₂ with an atmospheric general circulation model and a climatological ocean”, *Journal of Geophysical Research* **86**, 6385 (1981) (p. 50).
- ¹⁷⁷J. F. Mitchell, “The seasonal response of a general circulation model to changes in CO₂ and sea temperatures”, *Quarterly Journal of the Royal Meteorological Society* **109**, 113–152 (1983) (p. 50).
- ¹⁷⁸V. 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. 51, 111, 120, 128).
- ¹⁷⁹M. Rugenstein, J. Bloch-Johnson, J. Gregory, T. Andrews, T. Mauritsen, C. Li, T. L. Frölicher, D. Paynter, G. Danabasoglu, S. Yang, J.-L. 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**, 10.1029/2019GL083898 (2020) (p. 51).
- ¹⁸⁰C. 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. 51).
- ¹⁸¹T. 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. 51).
- ¹⁸²D. 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. 51).
- ¹⁸³E. 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. 53, 57, 59).
- ¹⁸⁴A. 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. 57, 58).

- ¹⁸⁵P. 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) (pp. 57, 65, 117, 124).
- ¹⁸⁶S. 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. 58).
- ¹⁸⁷S. 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. 58).
- ¹⁸⁸R. White, S. Anderson, J. Booth, G. Braich, C. Draeger, C. Fei, C. Harley, S. Henderson, M. Jakob, C.-A. Lau, L. M. Admasu, V. Narinesingh, C. Rodell, E. Roocroft, K. Weinberger, and G. West, “The Unprecedented Pacific Northwest Heatwave of June 2021”, *nature portfolio under review*, [10.21203/rs.3.rs-1520351/v1](https://doi.org/10.21203/rs.3.rs-1520351/v1) (2022) (pp. 63, 66).
- ¹⁸⁹D. Mitchell, “Climate attribution of heat mortality”, *Nature Climate Change* **11**, 467–468 (2021) (pp. 64, 81, 130).
- ¹⁹⁰C. 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. 64, 88, 107, 120, 129).
- ¹⁹¹A. E. Payne and G. Magnusdottir, “An evaluation of atmospheric rivers over the North Pacific in CMIP5 and their response to warming under RCP 8.5”, *Journal of Geophysical Research: Atmospheres* **120**, 11, 173–11, 190 (2015) (p. 64).
- ¹⁹²E. M. Fischer, S. Sippel, and R. Knutti, “Increasing probability of record-shattering climate extremes”, *Nature Climate Change*, 1–7 (2021) (pp. 64, 88).
- ¹⁹³N. J. Leach, P. A. G. Watson, S. N. Sparrow, D. C. H. Wallom, and D. M. H. Sexton, “Generating samples of extreme winters to support climate adaptation”, *Weather and Climate Extremes* **36**, 100419 (2022) (pp. 64, 81).
- ¹⁹⁴C. D. Roberts, R. Senan, F. Molteni, S. Boussetta, M. Mayer, and S. P. E. Keeley, “Climate model configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS cycle 43r1) for HighResMIP”, *Geoscientific Model Development* **11**, 3681–3712 (2018) (p. 65).
- ¹⁹⁵G. Wang, P. Hope, E.-P. Lim, H. H. Hendon, and J. M. Arblaster, “An Initialized Attribution Method for Extreme Events on Subseasonal to Seasonal Time Scales”, *Journal of Climate* **34**, 1453–1465 (2021) (pp. 65, 80, 118, 126).
- ¹⁹⁶P. Hope, M. Zhao, S. Abhik, G. Tolhurst, R. C. McKay, S. P. Rauniar, L. Bettio, A. Ramchurn, E.-P. Lim, A. S. Pepler, T. Cowan, and A. B. Watkins, “Subseasonal to Seasonal Climate Forecasts Provide the Backbone of a Near-Real-Time Event Explainer Service”, *Bulletin of the American Meteorological Society* **103**, S7–S13 (2022) (pp. 65, 118, 119).

- ¹⁹⁷M. F. Wehner, C. Zarzycki, and C. Patricola, “Estimating the Human Influence on Tropical Cyclone Intensity as the Climate Changes”, in *Hurricane Risk*, edited by J. M. Collins and K. Walsh, Hurricane Risk (Springer International Publishing, Cham, 2019), pp. 235–260 (pp. 65, 117, 124).
- ¹⁹⁸J. S. Tradowsky, L. Bird, P. V. Kreft, S. M. Rosier, I. Soltanzadeh, D. A. Stone, and G. E. Bodeker, “Toward Near-Real-Time Attribution of Extreme Weather Events in Aotearoa New Zealand”, *Bulletin of the American Meteorological Society* **103**, S105–S110 (2022) (p. 65).
- ¹⁹⁹S. B. Henderson, K. E. McLean, M. J. Lee, and T. Kosatsky, “Analysis of community deaths during the catastrophic 2021 heat dome: Early evidence to inform the public health response during subsequent events in greater Vancouver, Canada”, *Environmental Epidemiology* **6**, e189 (2022) (p. 66).
- ²⁰⁰M. J. Menne, I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R. S. Vose, B. E. Gleason, and T. G. Houston, *Global Historical Climatology Network - Daily (GHCN-Daily)*, Version 3.26, 2012 (p. 66).
- ²⁰¹M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, “An Overview of the Global Historical Climatology Network-Daily Database”, *Journal of Atmospheric and Oceanic Technology* **29**, 897–910 (2012) (p. 66).
- ²⁰²J. E. Overland, “Causes of the Record-Breaking Pacific Northwest Heatwave, Late June 2021”, *Atmosphere* **12**, 1434 (2021) (p. 66).
- ²⁰³H. Lin, R. Mo, and F. Vitart, “The 2021 Western North American Heatwave and Its Subseasonal Predictions”, *Geophysical Research Letters* **49**, e2021GL097036 (2022) (pp. 66, 74).
- ²⁰⁴R. Mo, H. Lin, and F. Vitart, “An anomalous warm-season trans-Pacific atmospheric river linked to the 2021 western North America heatwave”, *Communications Earth & Environment* **3**, 1–12 (2022) (pp. 66, 74).
- ²⁰⁵V. Thompson, A. T. Kennedy-Asser, E. Vosper, Y. T. E. Lo, C. Huntingford, O. Andrews, M. Collins, G. C. Hegerl, and D. Mitchell, “The 2021 western North America heat wave among the most extreme events ever recorded globally”, *Science Advances* **8**, eabm6860 (2022) (p. 66).
- ²⁰⁶O. Angélil, D. Stone, S. Perkins-Kirkpatrick, L. V. Alexander, M. Wehner, H. Shiogama, P. Wolski, A. Ciavarella, and N. Christidis, “On the nonlinearity of spatial scales in extreme weather attribution statements”, *Climate Dynamics* **50**, 2739–2752 (2018) (p. 68).
- ²⁰⁷S. Y. Philip, S. F. Kew, G. J. van Oldenborgh, F. S. Anslow, S. I. Seneviratne, R. Vautard, D. Coumou, K. L. Ebi, J. Arrighi, R. Singh, M. van Aalst, C. Pereira Marghidan, M. Wehner, W. Yang, S. Li, D. L. Schumacher, M. Hauser, R. Bonnet, L. N. Luu, F. Lehner, N. Gillett, J. Tradowsky, G. A. Vecchi, C. Rodell, R. B. Stull, R. Howard, and F. E. L. Otto, “Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021”, *Earth System Dynamics Discussions*, 1–34 (2021) (pp. 68, 80).
- ²⁰⁸IFS Documentation CY47R2, IFS Documentation (ECMWF, 2020) (p. 68).
- ²⁰⁹P. Janssen, *The Interaction of Ocean Waves and Wind* (Cambridge University Press, Cambridge, 2004) (p. 68).

- ²¹⁰T. Fichefet and M. A. M. Maqueda, “Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics”, *Journal of Geophysical Research: Oceans* **102**, 12609–12646 (1997) (p. 68).
- ²¹¹G. Madec, *NEMO ocean engine*, Project Report (Institut Pierre-Simon Laplace (IPSL), 2008) (pp. 68, 128).
- ²¹²S. J. Johnson, T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, S. P. Keeley, K. Mogensen, H. Zuo, and B. M. Monge-Sanz, “SEAS5: The new ECMWF seasonal forecast system”, *Geoscientific Model Development* **12**, 1087–1117 (2019) (p. 68).
- ²¹³*IFS Documentation CY43R1*, IFS Documentation (ECMWF, 2016) (p. 68).
- ²¹⁴P. Forster, T. Storelvmo, K. Armour, W. Collins, J.-L. Dufresne, D. Frame, D. Lunt, T. Mauritsen, M. Palmer, M. Watanabe, M. Wild, and H. Zhang, “The Earth’s Energy Budget, Climate Feedbacks, and Climate Sensitivity”, 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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 923–1054 (pp. 69, 122, 123).
- ²¹⁵H. Zuo, M. A. Balmaseda, S. Tietsche, K. Mogensen, and M. Mayer, “The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment”, *Ocean Science* **15**, 779–808 (2019) (pp. 70, 75).
- ²¹⁶R. A. Locarnini, A. V. Mishonov, O. K. Baranova, T. P. Boyer, M. M. Zweng, H. E. García, J. R. Reagan, D. Seidov, K. A. Weathers, C. R. Paver, and I. V. Smolyar, *World Ocean Atlas 2018, Volume 1: Temperature*, tech. rep. 82 (NOAA National Centers for Environmental Information, 2019), p. 50 (pp. 70, 75).
- ²¹⁷T. N. Stockdale, “Coupled Ocean–Atmosphere Forecasts in the Presence of Climate Drift”, *Monthly Weather Review* **125**, 809–818 (1997) (p. 72).
- ²¹⁸V. 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) (pp. 72, 73, 87, 130).
- ²¹⁹H. Douville, K. Raghavan, J. Renwick, R. Allan, P. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mbà, D. Rosenfeld, J. Tierney, and O. Zolina, “Water Cycle Changes”, 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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 1055–1210 (p. 75).

- ²²⁰S. Gulev, P. Thorne, J. Ahn, F. Dentener, C. Domingues, S. Gerland, D. Gong, D. Kaufman, H. Nnamchi, J. Quaas, J. Rivera, S. Sathyendranath, S. Smith, B. Trewin, K. von Schuckmann, and R. Vose, “Changing State of the Climate System”, 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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 287–422 (p. 75).
- ²²¹J. Gutiérrez, R. Jones, G. Narisma, L. Alves, M. Amjad, I. Gorodetskaya, M. Grose, N. Klutse, S. Krakovska, J. Li, D. Martínez-Castro, L. Mearns, S. Mernild, T. Ngo-Duc, B. van den Hurk, and J.-H. Yoon, “Atlas”, 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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 1927–2058 (p. 75).
- ²²²S. I. Seneviratne and M. Hauser, “Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles”, *Earth’s Future* **8**, e2019EF001474 (2020) (pp. 76, 122).
- ²²³T. J. Osborn, P. D. Jones, D. H. Lister, C. P. Morice, I. R. Simpson, J. P. Winn, E. Hogan, and I. C. Harris, “Land Surface Air Temperature Variations Across the Globe Updated to 2019: The CRUTEM5 Data Set”, *Journal of Geophysical Research: Atmospheres* **126**, e2019JD032352 (2021) (p. 78).
- ²²⁴K. A. Reed, M. F. Wehner, and C. M. Zarzycki, “Attribution of 2020 hurricane season extreme rainfall to human-induced climate change”, *Nature Communications* **13**, 1905 (2022) (pp. 80, 117, 124).
- ²²⁵D. Mitchell, L. Hawker, J. Savage, R. Bingham, N. S. Lord, M. J. U. Khan, P. Bates, F. Durand, A. Hassan, S. Huq, A. S. Islam, Y. Krien, J. Neal, C. Sampson, A. Smith, and L. Testut, “Increased population exposure to Amphan-scale cyclones under future climates”, *Climate Resilience and Sustainability* **1**, e36 (2022) (pp. 81, 130).
- ²²⁶WEF, *The Global Risks Report 2021*, tech. rep. (World Economic Forum, Jan. 2021) (p. 86).
- ²²⁷S. 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. Vicente-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. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 1513–1766 (pp. 86, 122).

- ²²⁸M. 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. 86).
- ²²⁹J. 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, Nov. 2018) (pp. 86, 120, 129).
- ²³⁰J. 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. 86, 88, 89, 97, 112, 129).
- ²³¹T. 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) (pp. 87, 130).
- ²³²K. 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. 88).
- ²³³K. 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. 88).
- ²³⁴A. 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. 88).
- ²³⁵D. 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. 88).
- ²³⁶D. 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. 88, 92).
- ²³⁷K. 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. 88).
- ²³⁸D. 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. 88).

- ²³⁹M. 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. 88).
- ²⁴⁰K. 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. 88).
- ²⁴¹V. D. Pope, M. L. Gallani, P. R. Rowntree, and R. A. Stratton, “The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3”, *Climate Dynamics* **16**, 123–146 (2000) (p. 88).
- ²⁴²E. 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. 88, 97).
- ²⁴³P. 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, Mar. 2020) (pp. 88, 97, 112).
- ²⁴⁴M. Allen, “Do-it-yourself climate prediction”, *Nature* **401**, 642–642 (1999) (p. 89).
- ²⁴⁵D. Anderson, “BOINC: a system for public-resource computing and storage”, in *Fifth IEEE/ACM International Workshop on Grid Computing* (Nov. 2004), pp. 4–10 (p. 89).
- ²⁴⁶D. 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. 89).
- ²⁴⁷D. 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) (pp. 89, 120).
- ²⁴⁸D. 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) (pp. 89, 120).
- ²⁴⁹E. 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, Sept. 2019) (p. 89).
- ²⁵⁰S. Sparrow, D. Sexton, N. J. Leach, P. A. G. Watson, and D. C. H. Wallom, “ExSamples Simulation Dataset”, *Scientific Data* **in prep.** (2021) (p. 90).
- ²⁵¹D. 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. 90).

- ²⁵²R. 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. 93, 95, 97).
- ²⁵³C. 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. 93).
- ²⁵⁴W. 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. 94).
- ²⁵⁵D. 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. 94).
- ²⁵⁶D. 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. 94).
- ²⁵⁷C. 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. 94, 103).
- ²⁵⁸M. 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. 94).
- ²⁵⁹M. 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. 94).
- ²⁶⁰J. López-Parages, B. Rodríguez-Fonseca, D. Domménget, 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. 94).
- ²⁶¹J. A. Francis and S. J. Vavrus, “Evidence linking Arctic amplification to extreme weather in mid-latitudes”, *Geophysical Research Letters* **39**, 10.1029/2012GL051000 (2012) (p. 94).
- ²⁶²R. 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. 94).
- ²⁶³M. 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. 94).
- ²⁶⁴J. A. Screen and I. Simmonds, “Exploring links between Arctic amplification and mid-latitude weather”, *Geophysical Research Letters* **40**, 959–964 (2013) (p. 94).
- ²⁶⁵J. A. Screen, “The missing Northern European winter cooling response to Arctic sea ice loss”, *Nature Communications* **8**, 14603 (2017) (p. 94).

- ²⁶⁶J. 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. 95).
- ²⁶⁷J. R. M. Hosking and J. R. Wallis, *Regional Frequency Analysis* (Cambridge University Press, Cambridge, Apr. 1997) (p. 95).
- ²⁶⁸S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics (Springer-Verlag, London, 2001) (pp. 95, 107).
- ²⁶⁹J. R. M. Hosking and J. R. Wallis, “Parameter and Quantile Estimation for the Generalized Pareto Distribution”, *Technometrics* **29**, 339–349 (1987) (p. 95).
- ²⁷⁰J. 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 (July 2007), pp. 43–50 (p. 95).
- ²⁷¹W. Z. Wan Zin, A. A. Jemain, and K. Ibrahim, “The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment”, *Theoretical and Applied Climatology* **96**, 337–344 (2009) (p. 95).
- ²⁷²M. Marani and M. Ignaccolo, “A metastatistical approach to rainfall extremes”, *Advances in Water Resources* **79**, 121–126 (2015) (p. 95).
- ²⁷³J. 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**, 10.1029/2010GL044613 (2010) (p. 95).
- ²⁷⁴R. 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. 95).
- ²⁷⁵J. L. Hodges, “The significance probability of the smirnov two-sample test”, *Arkiv för Matematik* **3**, 469–486 (1958) (p. 97).
- ²⁷⁶A. N. Kolmogorov, “Sulla Determinazione Empirica di Una Legge di Distribuzione”, *Giornale dell’Istituto Italiano degli Attuari* **4**, 83–91 (1933) (p. 97).
- ²⁷⁷N. 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. 97).
- ²⁷⁸N. Smirnoff, “Sur les écarts de la courbe de distribution empirique”, *Matematicheskii Sbornik* **6(48)**, 3–26 (1939) (p. 97).
- ²⁷⁹S. Brönnimann, “Impact of El Niño–Southern Oscillation on European climate”, *Reviews of Geophysics* **45**, 10.1029/2006RG000199 (2007) (p. 103).
- ²⁸⁰D. King, D. Schrag, Z. Dadi, Q. Ye, and A. Ghosh, *Climate change: A risk assessment*, tech. rep. (UK Foreign & Commonwealth Office, London, July 2015) (p. 107).
- ²⁸¹J. 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) (pp. 107, 129).

- ²⁸²S. 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, Nov. 2015), p. 144 (pp. 107, 110).
- ²⁸³S. 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) (pp. 107, 120).
- ²⁸⁴S. 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. 107).
- ²⁸⁵J. 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. 110).
- ²⁸⁶B. 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. 110).
- ²⁸⁷C. source id values, *CMIP6 source id values*, 2022 (p. 111).
- ²⁸⁸R. 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. 111).
- ²⁸⁹C. Skamarock, B. Klemp, J. Dudhia, O. Gill, M. Barker, W. Wang, and G. Powers, “A Description of the Advanced Research WRF Version 2”, [10.5065/D6DZ069T](https://doi.org/10.5065/D6DZ069T) (2005) (p. 116).
- ²⁹⁰G. A. Meehl, C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, “THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research”, *Bulletin of the American Meteorological Society* **88**, 1383–1394 (2007) (p. 116).
- ²⁹¹C. Schär, C. Frei, D. Lüthi, and H. C. Davies, “Surrogate climate-change scenarios for regional climate models”, *Geophysical Research Letters* **23**, 669–672 (1996) (pp. 117, 124).
- ²⁹²I. Takayabu, K. Hibino, H. Sasaki, H. Shiogama, N. Mori, Y. Shibutani, and T. Takemi, “Climate change effects on the worst-case storm surge: a case study of Typhoon Haiyan”, *Environmental Research Letters* **10**, 064011 (2015) (p. 117).
- ²⁹³C. M. Patricola and M. F. Wehner, “Anthropogenic influences on major tropical cyclone events”, *Nature* **563**, 339–346 (2018) (pp. 117, 119, 124).
- ²⁹⁴N. Maher, S. Milinski, and R. Ludwig, “Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble”, *Earth System Dynamics* **12**, 401–418 (2021) (p. 120).

- ²⁹⁵N. Schaller, A. L. Kay, R. Lamb, N. R. Massey, J. Van Oldenborgh, F. E. L. Otto, S. N. Sparrow, R. Vautard, P. Yiou, I. Ashpole, A. Bowery, S. M. Crooks, K. Haustein, C. Huntingford, W. J. Ingram, R. G. Jones, T. Legg, J. Miller, J. Skeggs, D. Wallom, A. Weisheimer, S. Wilson, P. A. Stott, and M. R. Allen, “Human influence on climate in the 2014 southern England winter floods and their impacts”, [10.1038/NCLIMATE2927 \(2016\)](https://doi.org/10.1038/NCLIMATE2927) (pp. 120, 122).
- ²⁹⁶P. Yiou, “AnaWEGE: a weather generator based on analogues of atmospheric circulation”, [Geoscientific Model Development 7, 531–543 \(2014\)](https://doi.org/10.5194/GMD-7-531-2014) (p. 120).
- ²⁹⁷P. Yiou and A. Jézéquel, “Simulation of extreme heat waves with empirical importance sampling”, [Geoscientific Model Development 13, 763–781 \(2020\)](https://doi.org/10.5194/GMD-13-763-2020) (p. 120).
- ²⁹⁸L. Suarez-Gutierrez, W. A. Müller, C. Li, and J. Marotzke, “Dynamical and thermodynamical drivers of variability in European summer heat extremes”, [Climate Dynamics, 1–16 \(2020\)](https://doi.org/10.1007/s00344-020-03000-w) (p. 121).
- ²⁹⁹K. B. Tokarska, M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, and R. Knutti, “Past warming trend constrains future warming in CMIP6 models”, [Science Advances 6, eaaz9549 \(2020\)](https://doi.org/10.1126/sciadv.aaaz9549) (p. 122).
- ³⁰⁰R. Hagedorn, R. Buizza, T. M. Hamill, M. Leutbecher, and T. N. Palmer, “Comparing TIGGE multimodel forecasts with reforecast-calibrated ECMWF ensemble forecasts”, [Quarterly Journal of the Royal Meteorological Society 138, 1814–1827 \(2012\)](https://doi.org/10.1002/qj.3111) (p. 122).
- ³⁰¹G. A. Meehl, C. A. Senior, V. Eyring, G. Flato, J.-F. Lamarque, R. J. Stouffer, K. E. Taylor, and M. Schlund, “Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models”, [Science Advances 6, eaab1981 \(2020\)](https://doi.org/10.1126/sciadv.aaab1981) (p. 122).
- ³⁰²F. Doblas-Reyes, A. Sörensson, M. Almazroui, A. Dosio, W. Gutowski, R. Haarsma, R. Hamdi, B. Hewitson, W.-T. Kwon, B. Lamptey, D. Maraun, T. Stephenson, I. Takayabu, L. Terray, A. Turner, and Z. Zuo, “Linking Global to Regional Climate Change”, 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](https://doi.org/10.1017/9781009154700), edited by V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021), pp. 1363–1512 (p. 122).
- ³⁰³A. H. Murphy, “A New Vector Partition of the Probability Score”, [Journal of Applied Meteorology and Climatology 12, 595–600 \(1973\)](https://doi.org/10.1175/1520-0450-12-1001) (p. 123).
- ³⁰⁴D. Walters, I. Boutle, M. Brooks, T. Melvin, R. Stratton, S. Vosper, H. Wells, K. Williams, N. Wood, T. Allen, A. Bushell, D. Copsey, P. Earnshaw, J. Edwards, M. Gross, S. Hardiman, C. Harris, J. Heming, N. Klingaman, R. Levine, J. Manners, G. Martin, S. Milton, M. Mittermaier, C. Morcrette, T. Riddick, M. Roberts, C. Sanchez, P. Selwood, A. Stirling, C. Smith, D. Suri, W. Tennant, P. L. Vidale, J. Wilkinson, M. Willett, S. Woolnough, and P. Xavier, “The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations”, [Geoscientific Model Development 10, 1487–1520 \(2017\)](https://doi.org/10.5194/GMD-10-1487-2017) (p. 123).

- ³⁰⁵C. MacLachlan, A. Arribas, K. A. Peterson, A. Maidens, D. Fereday, A. A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. E. Comer, J. Camp, P. Xavier, and G. Madec, “Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system”, *Quarterly Journal of the Royal Meteorological Society* **141**, 1072–1084 (2015) (pp. 123, 128).
- ³⁰⁶A. Bozzo, A. Benedetti, J. Flemming, Z. Kipling, and S. Rémy, “An aerosol climatology for global models based on the tropospheric aerosol scheme in the Integrated Forecasting System of ECMWF”, *Geoscientific Model Development* **13**, 1007–1034 (2020) (p. 124).
- ³⁰⁷A. Zhao, M. A. Bollasina, and D. S. Stevenson, “Strong Influence of Aerosol Reductions on Future Heatwaves”, *Geophysical Research Letters* **46**, 4913–4923 (2019) (p. 124).
- ³⁰⁸S. Rémy, Z. Kipling, J. Flemming, O. Boucher, P. Nabat, M. Michou, A. Bozzo, M. Ades, V. Huijnen, A. Benedetti, R. Engelen, V.-H. Peuch, and J.-J. Morcrette, “Description and evaluation of the tropospheric aerosol scheme in the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS-AER, cycle 45R1)”, *Geoscientific Model Development* **12**, 4627–4659 (2019) (p. 124).
- ³⁰⁹P. Laloyaux, E. de Boisseson, M. Balmaseda, J. R. Bidlot, S. Broennimann, R. Buizza, P. Dalhgren, D. Dee, L. Haimberger, H. Hersbach, Y. Kosaka, M. Martin, P. Poli, N. Rayner, E. Rustemeier, and D. Schepers, “CERA-20C: A Coupled Reanalysis of the Twentieth Century”, *Journal of Advances in Modeling Earth Systems* **10**, 1172–1195 (2018) (p. 124).
- ³¹⁰R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Journal of Basic Engineering* **82**, 35–45 (1960) (p. 125).
- ³¹¹M. J. Rodwell and F. J. Doblas-Reyes, “Medium-Range, Monthly, and Seasonal Prediction for Europe and the Use of Forecast Information”, *Journal of Climate* **19**, 6025–6046 (2006) (p. 127).
- ³¹²F. Vitart, “Evolution of ECMWF sub-seasonal forecast skill scores”, *Quarterly Journal of the Royal Meteorological Society* **140**, 1889–1899 (2014) (p. 127).
- ³¹³R. Swinbank, M. Kyouda, P. Buchanan, L. Froude, T. M. Hamill, T. D. Hewson, J. H. Keller, M. Matsueda, J. Methven, F. Pappenberger, M. Scheuerer, H. A. Titley, L. Wilson, and M. Yamaguchi, “The TIGGE Project and Its Achievements”, *Bulletin of the American Meteorological Society* **97**, 49–67 (2016) (p. 127).
- ³¹⁴S. Monhart, C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, “Skill of Subseasonal Forecasts in Europe: Effect of Bias Correction and Downscaling Using Surface Observations”, *Journal of Geophysical Research: Atmospheres* **123**, 7999–8016 (2018) (p. 127).
- ³¹⁵N. Mishra, C. Prodhomme, and V. Guemas, “Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe”, *Climate Dynamics* **52**, 4207–4225 (2019) (p. 127).
- ³¹⁶T. Haiden, M. Janousek, F. Vitart, Z. Ben-Boualleugue, L. Ferranti, and F. Prates, *Evaluation of ECMWF forecasts, including the 2021 upgrade*, tech. rep. 884 (ECMWF, 2021) (p. 127).

- ³¹⁷Z. Hausfather, K. Marvel, G. A. Schmidt, J. W. Nielsen-Gammon, and M. Zelinka, “Climate simulations: recognize the ‘hot model’ problem”, *Nature* **605**, 26–29 (2022) (p. 128).
- ³¹⁸J. Murphy, D. Sexton, G. Jenkins, P. Boorman, B. B. Booth, C. Brown, R. Clark, M. Collins, G. Harris, E. Kendon, R. A. Betts, S. J. Brown, T. Howard, K. Humphrey, M. McCarthy, R. McDonald, A. Stephens, C. Wallace, R. Warren, R. Wilby, and R. A. Wood, *UK Climate Projections Science Report: Climate change projections*, tech. rep. (Met Office Hadley Centre, Exeter, 2009) (p. 129).
- ³¹⁹C. Zhou, M. D. Zelinka, A. E. Dessler, and M. Wang, “Greater committed warming after accounting for the pattern effect”, *Nature Climate Change*, 1–5 (2021) (p. 130).
- ³²⁰T. Kelder, N. Wanders, K. v. d. Wiel, T. I. Marjoribanks, L. J. Slater, R. I. Wilby, and C. Prudhomme, “Interpreting extreme climate impacts from large ensemble simulations—are they unseen or unrealistic?”, *Environmental Research Letters* **17**, 044052 (2022) (p. 130).
- ³²¹S. E. Perkins-Kirkpatrick, D. A. Stone, D. M. Mitchell, S. Rosier, A. D. King, Y. T. E. Lo, J. Pastor-Paz, D. Frame, and M. Wehner, “On the attribution of the impacts of extreme weather events to anthropogenic climate change”, *Environmental Research Letters* **17**, 024009 (2022) (p. 130).
- ³²²D. J. Frame, M. F. Wehner, I. Noy, and S. M. Rosier, “The economic costs of Hurricane Harvey attributable to climate change”, *Climatic Change* **160**, 271–281 (2020) (p. 130).
- ³²³B. Strauss, P. Orton, K. Bittermann, M. K. Buchanan, R. E. Kopp, S. Kulp, C. Massey, and H. de Moel, “Economic damages from Hurricane Sandy attributable to sea level rise caused by anthropogenic climate change”, *Nature Communications* **12**, 15 (2021) (p. 130).
- ³²⁴E. Raju, E. Boyd, and F. Otto, “Stop blaming the climate for disasters”, *Communications Earth & Environment* **3**, 1–2 (2022) (p. 130).
- ³²⁵L. Alfieri, P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen, and F. Pappenberger, “GloFAS — global ensemble streamflow forecasting and flood early warning”, *Hydrology and Earth System Sciences* **17**, 1161–1175 (2013) (p. 131).
- ³²⁶S. Wilkinson, S. Dunn, R. Adams, N. Kirchner-Bossi, H. J. Fowler, S. González Otálora, D. Pritchard, J. Mendes, E. J. Palin, and S. C. Chan, “Consequence forecasting: A rational framework for predicting the consequences of approaching storms”, *Climate Risk Management* **35**, 100412 (2022) (p. 131).