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

Nicholas J. Leach

St. Cross College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2022

Abstract

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

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

Nicholas J. Leach
St. Cross College
University of Oxford

Supervised by

Antje Weisheimer
Myles R. Allen

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

Acknowledgements

Personal

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.

Institutional

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.

Abstract

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

Contents

List of Figures

List of Abbreviations

SST . . . . . Sea surface temperatures.

Quote

— author

1 Introduction

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

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

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

Contents

*with the author contributing as follows.

1.1 Section

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

Quote

— author

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 probablistic event attribution – the definition of the event – projects strongly onto the quantitative results. In the closing remarks, I reflect on potential issues with the approach taken within the chapter, and suggest ways in which these could be overcome.

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

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

Contents

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

2.1 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.

2.2 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 (*1*, *2*). Impacts were felt across Europe, with wildfires burning in Sweden (*3*, *4*), heatstroke deaths in Spain (*5*), and widespread drought (*6*). During the summer, the World Weather Attribution (WWA) initiative released an analysis of the heat spell (*7*) 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 (*8*, *9*). These two estimates appear to quantitatively disagree; however, we show they can be reconciled by investigating the effects of using different spatial domains and temporal scales in the event definition. We also demonstrate that prescribed SST model simulations can underrepresent the variability of temperature extremes, especially near the coast, with implications for any derived attribution results.

2.2.1 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 statistic we use is the annual maximum of the 1-, 10-, and 90-day running mean of daily mean 2-m temperature (hereafter TM1x, TM10x, and TM90x respectively). We analyze three spatial scales: model grid box, regional, and European. For regional and European event definitions, the spatial mean is calculated before the running mean. Regional extents are taken from Christensen and Christensen (*10*), and European extent is the E-OBS (*11*) domain (land points within 25 – 71.5° N, 25° W – 45° E). The WWA used the annual maxima of 3-day mean daily maximum temperatures at specific grid points for its connection to local health effects (*12*), 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 (*13*).

2.3 Model simulations & validation

Three sets of simulations from the UKMO Hadley Centre HadGEM3-A global atmospheric model (*14*, *15*) are used. These are a historical ensemble (1960–2013; Historical) and factual (ACT) and counterfactual (a "natural" world without anthropogenic forcings; NAT) ensembles of 2018. We compare results from this factual-counterfactual analysis with those from a trend-based analysis of Historical, ensembles from EURO-CORDEX (*16*–*18*) (1971–2018) and RACMO (*19*, *20*) (1950–2018), and observations from E-OBS (1950–2018). A full model description is provided in the online supplemental information. Initially, we performed our analysis with the weather@home HadRM3P European-25 km setup (*21*) but found that this model overestimates the variability over all Europe for daily through seasonal-scale event statistics, and so it was omitted.

2.4 Methods

We calculate the return period (RP) for the 2018 event in a distribution fit to E-OBS using the generalized extreme value (GEV) distribution to model TM1x and TM10x, and the generalized logistic distribution to empirically model TM90x throughout. Since the distribution of temperature extremes changes as the climate does, to account for the non-stationarity of the time series we first remove the trend attributable to low-pass-filtered globally averaged mean surface temperature (GMST, from Berkeley Earth; Rohde et al. 2013) in an ordinary least squares regression (the regression coefficient or trend is shown in the supplemental material in Fig. ES1; *22*). We then find the temperature threshold corresponding to the RP in a distribution fit to the model's climatology. In the factual/counterfactual analysis, we do this by fitting parameters to a detrended (against GMST; trends shown in Figs. ES2c7–9) climatological ensemble of Historical plus 15 randomly sampled members of ACT. We finally calculate the probability (P) of exceeding this climatological temperature threshold in distributions fit to the ACT and NAT ensembles and calculate the probability ratio, $PR = P_{\mathrm{ACT}}/P_{\mathrm{NAT}}$, representing the increased likelihood of the 2018 event in the factual compared to the counterfactual world. Using estimated event probabilities rather than observed magnitudes constitutes a quantile bias correction (*23*), minimizing model biases in the mean and variability of the temperatures analyzed. A description of uncertainty calculation and the trend-based analysis discussed below is included in the supplemental material.

2.5 Results

Extreme daily heat events, measured by TM1x, are distributed heterogeneously throughout Europe (Fig. ES1i). This is paralleled in the factual/ counterfactual PRs seen in Fig. 1a, with large proportions 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 (Fig. 1d) with Scandinavia and the Iberian Peninsula respectively experiencing 1-in-150 [26–26,000] 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 record maximum daily temperatures likely to be repeated within 4 years. Considering the whole of Europe, the likelihood of the 2018 maximum of daily

European mean temperature occurring without climate change is zero. This result is consistent with Uhe et al. (2016) and Angélil et al. (2018), who showed that increasing spatial scale tends to increase the probability ratio.

Extreme 10-day heat events, TM10x, were also widespread in Europe, with the most extreme occurring in Scandinavia (Fig. ES1j). Regionally, the PRs become more uniform (Fig. 1d), although Scandinavia and the Iberian Peninsula still have very high bestestimate PRs of 185 [17–infinite] and 110 [18–56,000] respectively. The best-estimate PR for the average of Europe is still formally infinite.

The PR map for season-long heat events measured by TM90x is more uniform throughout Europe (Fig. 1c). Scandinavia, the British Isles, France, and central and eastern Europe, all of which experienced on the order of 1-in-10 year events (Fig. ES1l), and the corresponding best-estimate PRs are between 10 and 100 for all regions (Fig. 1d), including those with lower return periods. The PR for the European average is 1,000 [500–2,000].

Trend-based analysis [Figs. ES1m–p (observations) and Fig. ES2b (models)] yields similar results, although we note that for HadGEM-3A this results in generally higher PRs, due to the linear trend with GMST in the climatology being greater than the difference between the two ensembles used in the factual/counterfactual analysis. Observational and model analysis contradict in some grid boxes in northern Scandinavia for TM1x and TM10x, since the observed best-estimate trend against GMST is negative, reducing the event probability for the presentday compared to the preindustrial climate, therefore yielding PRs of less than 1. Comparing the regional factual/counterfactual model with observational analysis (Fig. 1d vs Fig. ES1p) shows that the large observational uncertainties overlap with the model results: the difference could be due to natural variability affecting the small observational sample size. However, we are cautious of drawing any conclusions regarding the change in likelihood of extreme heat events as defined here for these locations.

The PR increases with the event statistic time scale for the majority of grid points and regions (shown in Fig. 1). Figure 2 illustrates the cause using the British Isles region: as the time scale increases, the event statistic distribution variance decreases, while the mean shift between the factual and counterfactual distributions remains constant. Figure ES1t shows that the similarity in trends with GMST between the three time scales is also true for the observations. 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 PRs for TM1x in Scandinavia, the Iberian Peninsula, and the Netherlands. The solid and dotted black lines compare the temperature thresholds when using event return periods to anomaly magnitudes in E-OBS. This explains why the TM90x PR is much higher than the other time scales for the British Isles: in addition to the decreased variance, the seasonal-scale heat event was more unusual than the other time scales, with a longer return period (10.6 [5.7–21] years) than TM10x (2.6 [1.8–3.9] years) and TM1x (3.6 [2.5–6.2] years). These factors together result in PRs of 3.6 [2.9–4.8] for TM1x and 43 [27–84] for TM90x. We suggest that the change in variance between the time scales used 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 playing a minor role as we have demonstrated for the British Isles.Although higher return periods for TM90x do impact the PRs found, this effect is generally less significant than changes in variability between the time scales.

Figure 2 also demonstrates a relevant deficiency in the model: the model distributions are narrower than the observed distributions, meaning the model has lower variability than the real world. This reduced variance has a significant impact on attribution results (*24*) and means that the PRs for the British Isles presented here, especially for TM90x, are likely to be overestimated. Underrepresented variability often occurs in prescribed SST models (*25*, *26*) and is visible in HadGEM-3A for many coastal locations over Europe (Figs. ES2a7–9). Figure 2d shows the power spectrum of JJA summer temperatures over the British Isles, indicating that HadGEM3-A has similar spectral characteristics to E-OBS, but underrepresents the intraseasonal 2-m temperature variability at almost all frequencies, which will likely result in overestimated PRs. Power spectra for other model ensembles are shown for comparison, demonstrating that the fully bias-corrected EURO-CORDEX ensemble has the same variability characteristics and magnitude as the observations.

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. (2019), 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. The attribution results also show that the extreme daily temperatures experienced in parts of Scandinavia, the Netherlands, and the Iberian Peninsula would have been highly unlikely without anthropogenic warming. The prescribed SST model experiments used here tend 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 (*24*). 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 try 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 Closing remarks

Quote

— author

3 Attribution and projection

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

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

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

Contents

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

3.1 Abstract

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

3.2 Introduction

Weather extremes are one of the most damaging hazards that society faces at the present-day (*27*). 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 (*28*). 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 (*22*, *29*, *30*). 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 (*31*, *32*) and included a number of novel climate model ensembles: a set of transient global simulations from coupled climate models, with 15 simulations from a single-model perturbed parameter ensemble (PPE) and 13 additional simulations from CMIP5 models; a set of 12 regional climate model simulations; and a set of 12 convection permitting model projections. In this study, we focus on the PPE of 15 global simulations, and our analysis and results build upon the information provided by these runs.

In particular, we are interested in how effectively the UKCP18 PPE has sampled extreme weather during the UK winter, and in exploring methods for improving the sampling of extremes that could inform the design of future projections. To this end, we aim to provide proof-of-concept of a methodology for generating large ensembles

of extreme winters. The key advantage is that our ensembles provide multivariate spatially and physically coherent scenarios of extreme weather with high return periods for use in impacts assessment.

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

This provision of many samples of extremes is similar to the UNSEEN method for quantifying weather extremes (*33*, *34*). UNSEEN uses seasonal hindcast ensembles to estimate the likelihood of "unprecedented" extreme events with considerably more confidence than possible from the observational record in isolation. The key similarity between UNSEEN and the approach taken here is that both are methods that aim to drill into the uncertainty surrounding the most extreme events by providing very large ensembles of such extremes using a dynamical model. However, there are key differences: UNSEEN uses coupled simulations that are conditioned solely on the predictable component of the weather at the time the model was initialised by observations, while in ExSamples, the model is atmosphere-only and conditioned both on perturbed initial conditions and lower boundary forcing from a climate projection. Another difference lies in the distributed computing system used here, which enables 1000+ member ensembles of a single winter to be produced; compared to the O(100) members produced by operational seasonal forecasting centers.

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

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

In this paper, we first describe the models used, experiment design and statistical methodologies performed within the study. We then present the results of our experiments, first comparing the climate distributions of the two models over a present-day baseline period to assess whether there are any significant biases between them. Taking any biases into account, we compare the projections from our three future ensembles to the

UKCP18 PPE, focussing on how the extreme tail of the climate distribution is sampled. This comparison allows us to explore the sampling advantage given by, and influence of, the SST and SIC. The very large ensembles created also allow us to examine the influence of the large scale dynamics present during the study winters using a circulation analog approach. We then use a single ensemble member case study to highlight the importance of large ensembles for sampling unprecedented extreme events that cannot always be statistically extrapolated from smaller ensembles (*35*, *36*). Finally, we discuss the insights provided by these experiments, and how they might inform the design of future projections; also suggesting directions for future research that could further improve our approach.

3.3 Study design and methods

3.3.1 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 (*37*). This PPE is based on the global HadGEM3-GC3.05 coupled ocean atmosphere model (*32*, *38*). 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 (*39*–*42*). 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 (*32*), minus two members that displayed a significant weakening of the Atlantic Meridional Overturning Circulation (*43*). 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 (*44*, *45*). Like its predecessor, HadAM3 (*46*), 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 (*47*, *48*), which matches the horizontal resolution of the HadGEM3-GC3.05 simulations analysed here. HadAM4 has 38 vertical levels; and here the sea surface temperature (SST) and sea ice fraction (SIC) boundary conditions are taken from specific years and members of the HadGEM3-GC3.05 UKCP18 PPE simulations.

A key aspect of the HadAM4 simulations described here are that they are performed on the personal computers of volunteers using the climateprediction.net distributed computing system (*49*–*51*). This system has been used previously to run a range

of Hadley Center Unified Model variants (*52*), including a coupled atmosphere-slab ocean model (*53*), a fully coupled model (*54*) and an atmosphere-only model (*55*) similar to HadAM4. The near thousand member ensembles presented here would be prohibitively expensive to run using a standard supercomputer, and so we are only able to run the bespoke experiments presented in this study because of this distributed computing system, and the volunteers involved. However, the constraints of this system strongly motivate the choice of HadAM4: it is sufficiently memory-efficient that it can be run on personal computers at the same horizontal resolution as the state-of-the-art HadGEM3-GC3.05 model.

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

3.3.2 ExSamples experiment design

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

1. Select a single winter from within the UKCP18 PPE over the 2061-2080 period. This winter is chosen on the basis of being particularly "extreme"; more detail on how we selected the three future winters is given below in "Selecting the three 'extreme' study winters". The 2061-80 period is used as we wanted to test this proof-of-concept with a large underlying climate change signal; and this is the period for which there is additional UKCP18 data available: 12km regional and 2.2km convection-permitting model projections (*32*, *56*).

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 (*49*, *53*), 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 (*57*).

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

Definitions of key terms

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

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

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

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

Selecting the three "extreme" study winters

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

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

Motivated by the recent winter extremes of the record hottest UK winter day of February 26th 2019 and the record wet winter month of February 2020, we aimed to select two "hot" winters and one "wet" winter. However, the method could be applied to the winters with the coldest or driest deviations. As shown in Figure 1, there is one clear candidate for each type of extreme: UKCP18 PPE member 02868 (ID numbers as *41*) year 2066 as a hot winter; and member 02242 year 2068 for the wet winter. The next most extreme hot winters shown in Figure 1A all had similar deviations, so we distinguished between them on the basis of their anomalies, choosing member 01554 year 2072, which has the highest anomaly of any of the candidate winters.

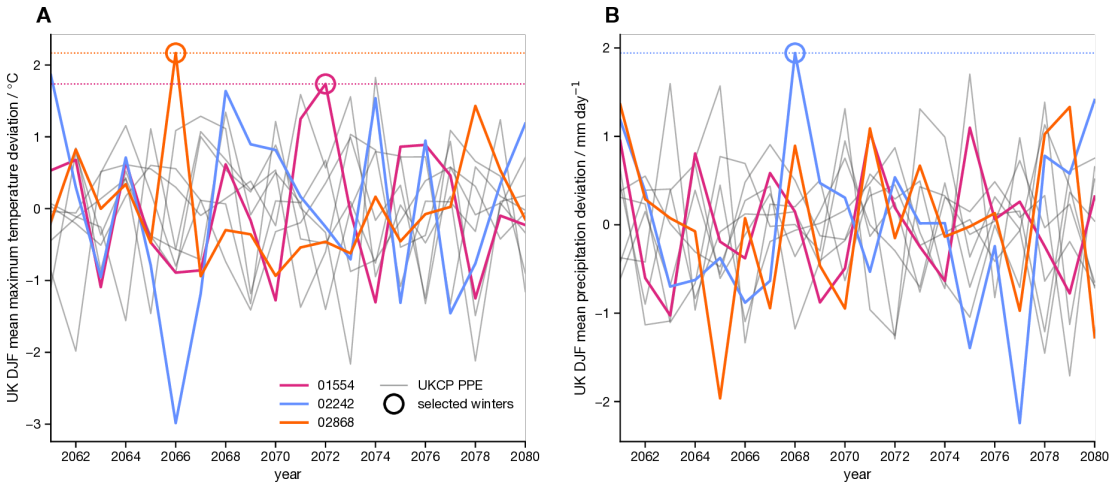

Figure 3.1: UKCP PPE 2061-2080 deviations. **A**, DJF mean of daily maximum temperatures averaged over the UK region. Coloured lines indicate the three UKCP runs from which the study winters were chosen. The study winters are circled and dotted horizontal lines indicate the deviation of each study winter. 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 used
	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 3.1: Summary of experiments performed for ExSamples project.

Synoptic characterisation of the study winters

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

The HOT1 winter displays a strong positive NAO pattern. Over the UK the flow is even more zonal, and has a weaker gradient; the positive NAO pattern is also weaker. In terms of the 30 weather patterns derived by (*59*), 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 (*60*). During this winter, the El Nino Southern Oscillation (ENSO) pattern of global SST variability was in a strong positive phase , alongside moderately positive Atlantic Multidecadal Variability and negative phase Pacific Decadal Oscillation (*61*). 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.* (*59*): a strong positive NAO with associated pressure high off the west coast of Spain. This weather pattern is associated with warm and wet weather over the UK (*60*, *62*–*64*). There is a weak La Nina (negative) ENSO phase ; which has previously been linked to an increased likelihood of positive NAO (*65*–*68*). No other modes of SST variability are present. With regards to SIC, this particular PPE member has virtually lost all winter Arctic sea ice by 2072. It has been suggested that Arctic sea ice loss may be linked with more persistent mid-latitude weather patterns (*69*, *70*), though this is still a subject of active scientific interest (*71*–*73*).

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

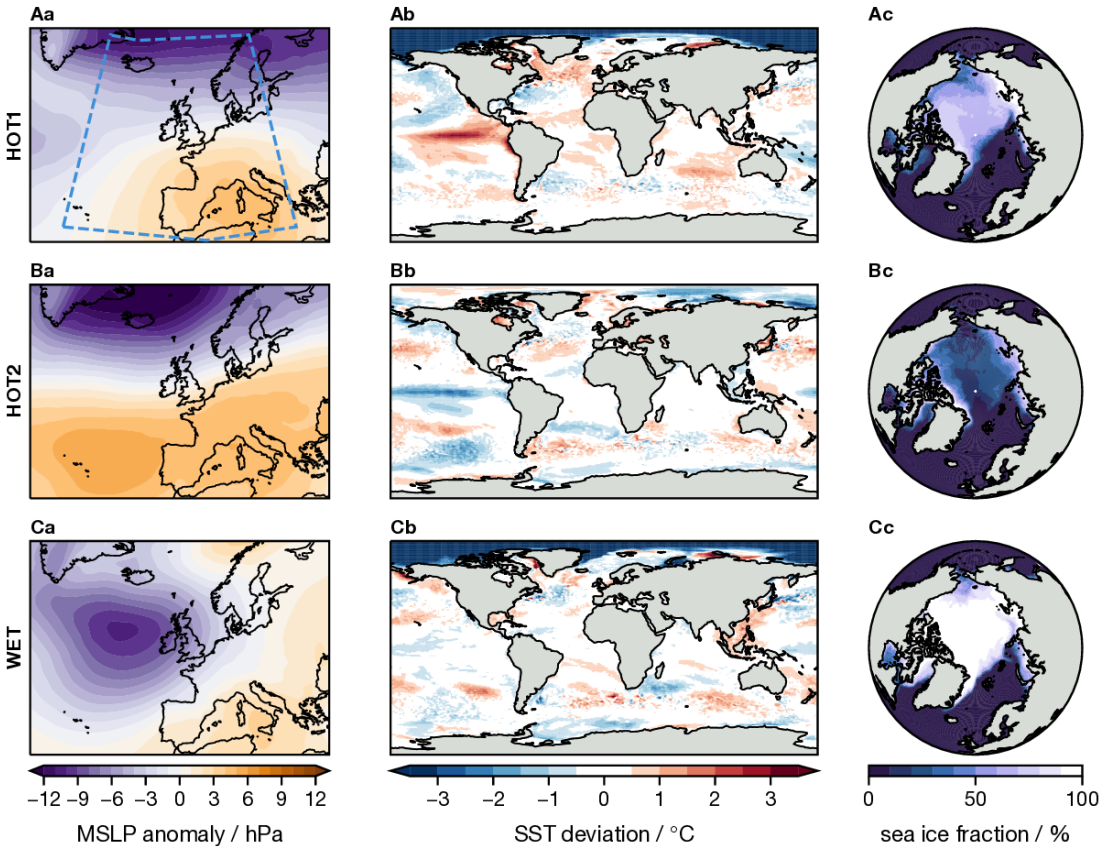

Figure 3.2: synoptic features of the study winters within the UKCP simulations. The row titles indicate the study winter. **a**, DJF mean MSLP anomalies for each winter. **b**, DJF mean SST deviations for each winter. Deviations are calculated for each gridpoint timeseries over 2061-2080. **c**, DJF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

3.3.3 Statistical methods

Estimating distributions of extremes

We estimate distributions using the method of L-moments (*74*–*76*). We use L-moments for their computational efficiency and stability. Uncertainties in the fit distributions, their CDFs and corresponding return periods are calculated using a 10,000 resample nonparametric bootstrap. The specific distributions used for each variable analysed are described in the following paragraphs.

For mean DJF daily maximum temperatures (TXm) and mean DJF precipitation rate (PRm), we use a generalised pareto distribution (*77*, *78*) 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 (*75*) 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 (*79*, *80*). 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 (*81*), though further work is needed to determine this conclusively.

Analog construction

In order to assess the dynamical contributions to the extreme weather simulated during the study winters, we use an MSLP analog approach (*82*–*84*). For each future ExSamples ensemble (and each corresponding baseline ensemble), we create a subsample of analogs composed of ensemble members that have a root mean square error (Euclidean distance) of less than 3 hPa from the UKCP18 PPE study winter average MSLP over the domain enclosed by the dashed blue lines in Figure 2Aa (-30:20° E; 35:70° N). This domain was the best for explaining variance in UK temperatures and close to best for UK precipitation of those investigated by (*59*). We used a 3hPa threshold as this was the tightest constraint that resulted in analog ensembles large enough to infer statistics from with any degree of certainty ($>$20 members in each case). The MSLP distance based subsampling results in an ensemble of analogs in which the mean large scale flow during the winter very closely matches the study winter. We can then use these ensembles of analogs to estimate the dynamical contribution and associated uncertainty to the extreme weather.

3.4 Results

3.4.1 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 (*85*–*88*) at the 5 % significance level on the anomalies of the variable in question unless stated otherwise. We use anomalies here since our main results are presented using anomalies to account for any model mean biases (and biases between different UKCP18 PPE members), but note if there are significant differences between the two model climate means. Verifying the accuracy of these models against reality lies outside of the scope of this paper, but has already been studied for both the UKCP18 PPE (*32*) and HadAM4 (*47*, *48*).

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

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

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

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

3.4.2 Projections of future extremes

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

HOT1

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

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

In order to compare the conditioning (effectively the "sampling advantage") across the three ensembles, we examine the relative exceedance risk of three different extreme thresholds set by the following UKCP18 PPE distribution quantiles: 0.9, 0.95 and 0.99; representing 1-in-10, -20 and -100 year extremes. We do this for both the TXm and PRm variables. We first calculate the threshold values that correspond to the given extremes using the UKCP 2061-2080 deviations statistical fit (ie. the black line in Figure 3B). We then calculate the fractions of the UKCP 2061-2080 and ExSamples ensembles that lie above these thresholds. We present the results in Table 2 in terms of the relative risk of the given extreme in the ExSamples ensemble compared to the UKCP ensemble. This is calculated as the fraction of the ExSamples ensemble that exceeds the threshold divided by the corresponding fraction of the UKCP ensemble, analogous to the "risk ratios" often used in extreme event attribution studies (*89*, *90*). This relative risk provides a measure of how many more samples of extremes of a particular return period we would expect to see in the ExSamples ensembles compared to a UKCP18 PPE-style ensemble of equal size. The quantitative results in Table 2 support the picture provided by Figure 3: the HOT1 ensemble was not conditioned towards producing any more extremes than expected from the unconditioned UKCP 2061-2080 ensemble (for several thresholds it actually appears to have been marginally conditioned away from producing extremes).

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

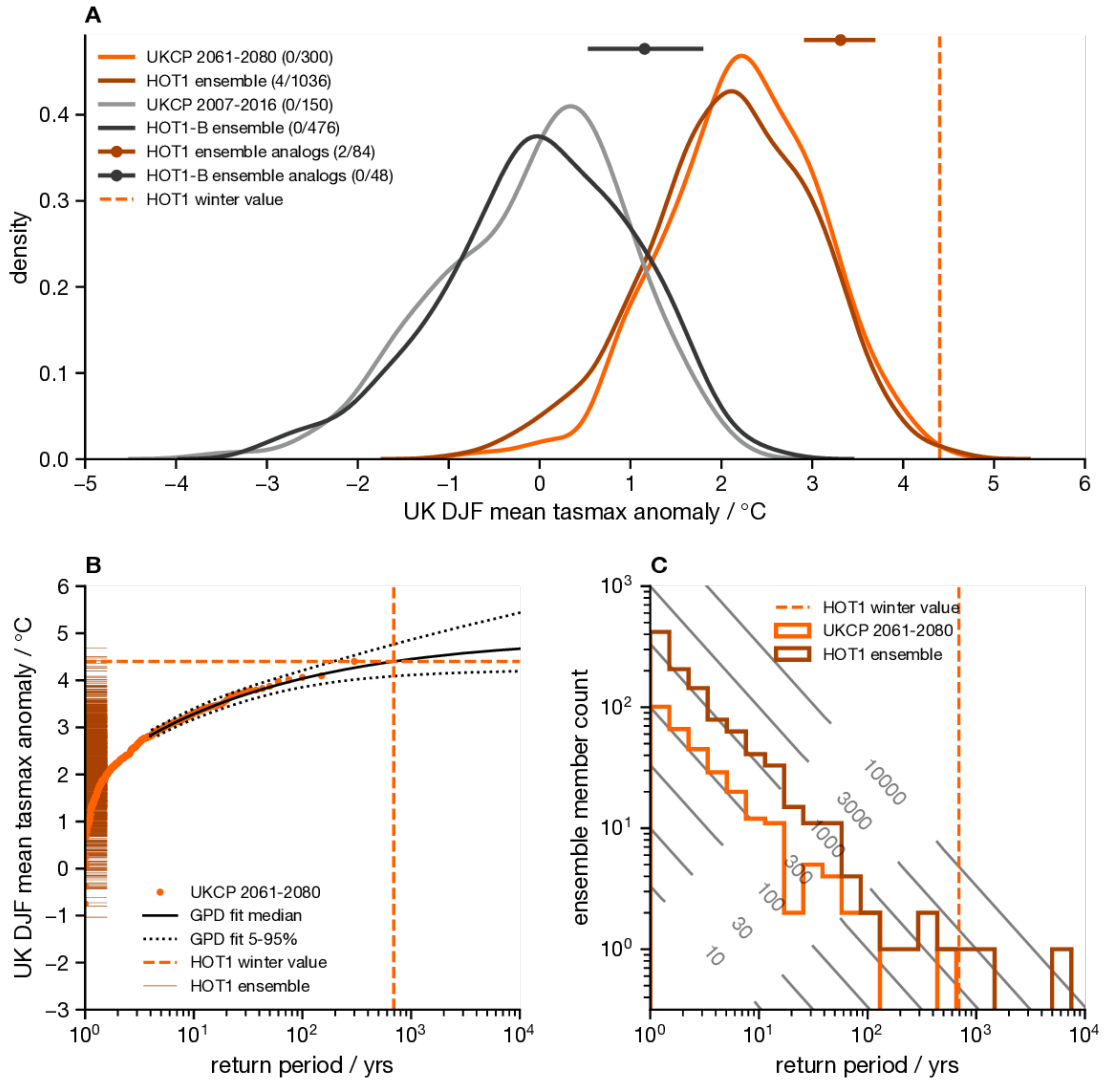

Figure 3.3: comparing statistics of DJF mean of daily maximum temperatures (TXm) averaged over the UK region for the HOT1 winter. **A**, PDFs of baseline and future ensembles. The light orange PDF shows UKCP 2061-2080 deviations, with the distribution mean set to the ensemble mean anomaly between 2007-2016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

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

Study winter	Variable	UKCP18 quantile (return period)		
		0.9 (1-in-10 year)	0.95 (1-in-20)	0.99 (1-in-100)
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 3.2: Ratio of exceedance likelihood of three extreme thresholds between the ExSamples future ensembles and the UKCP18 PPE 2061-2080 deviations. Square brackets indicate a 90 % CI.

HOT2

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

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

In addition to the SST conditioning, the dynamical situation of the study winter also made an extreme season more likely, as shown by the horizontal lines representing the likely range of the analog subsamples in Figure 4A. Based on the number of analogs sampled, the frequency of this particular large-scale flow was increased by a factor of 3.6 [2.6 , 5.4] relative to the climatological frequency estimated using the ExSamples baseline ensemble, which may be due to the prescribed boundary conditions. This would fit within the canonical picture that the negative La Nina ENSO phase is associated with positive NAO (*65*, *91*).

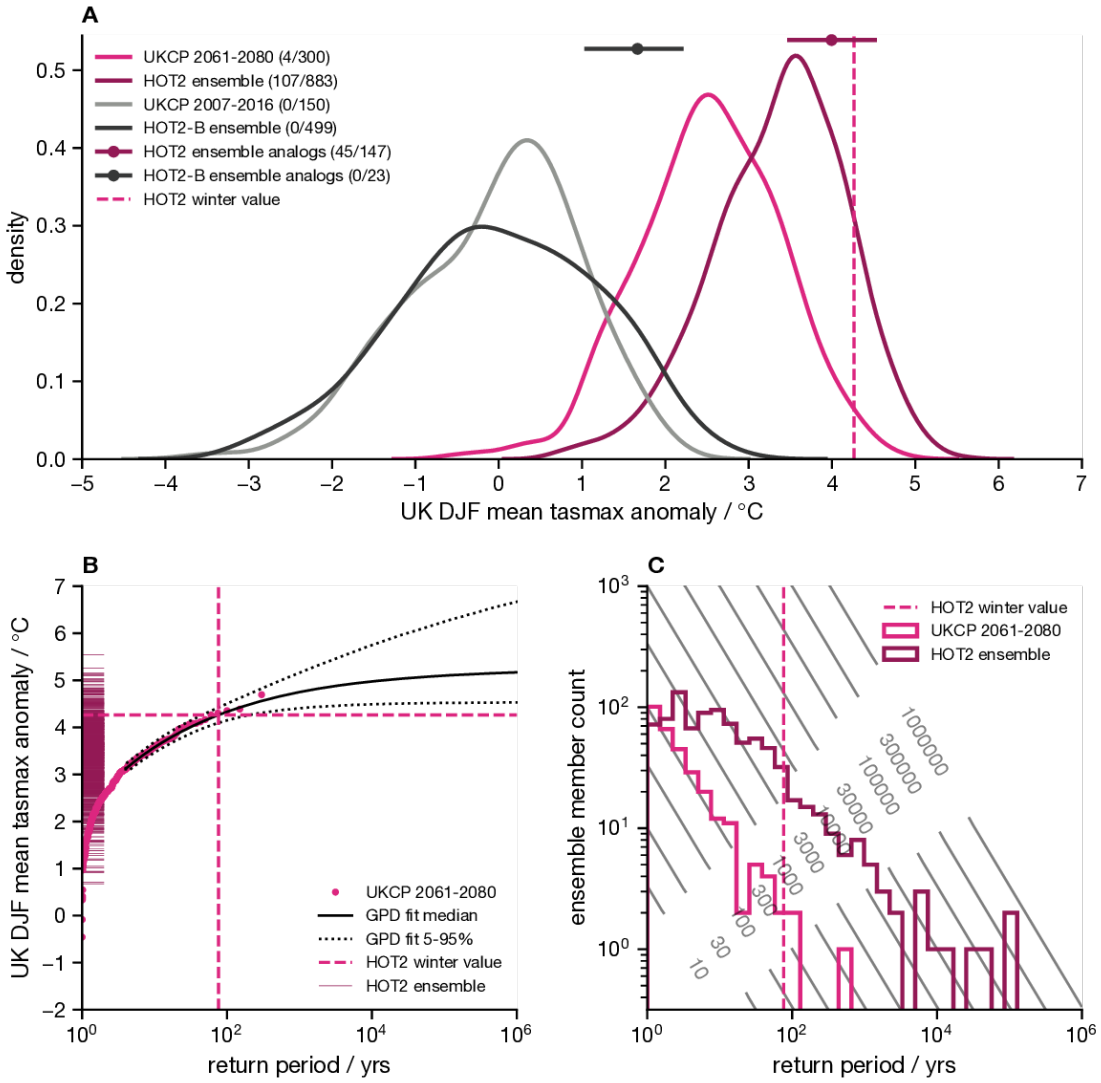

Figure 3.4: As Figure 3, but for the HOT2 winter.

WET

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

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

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

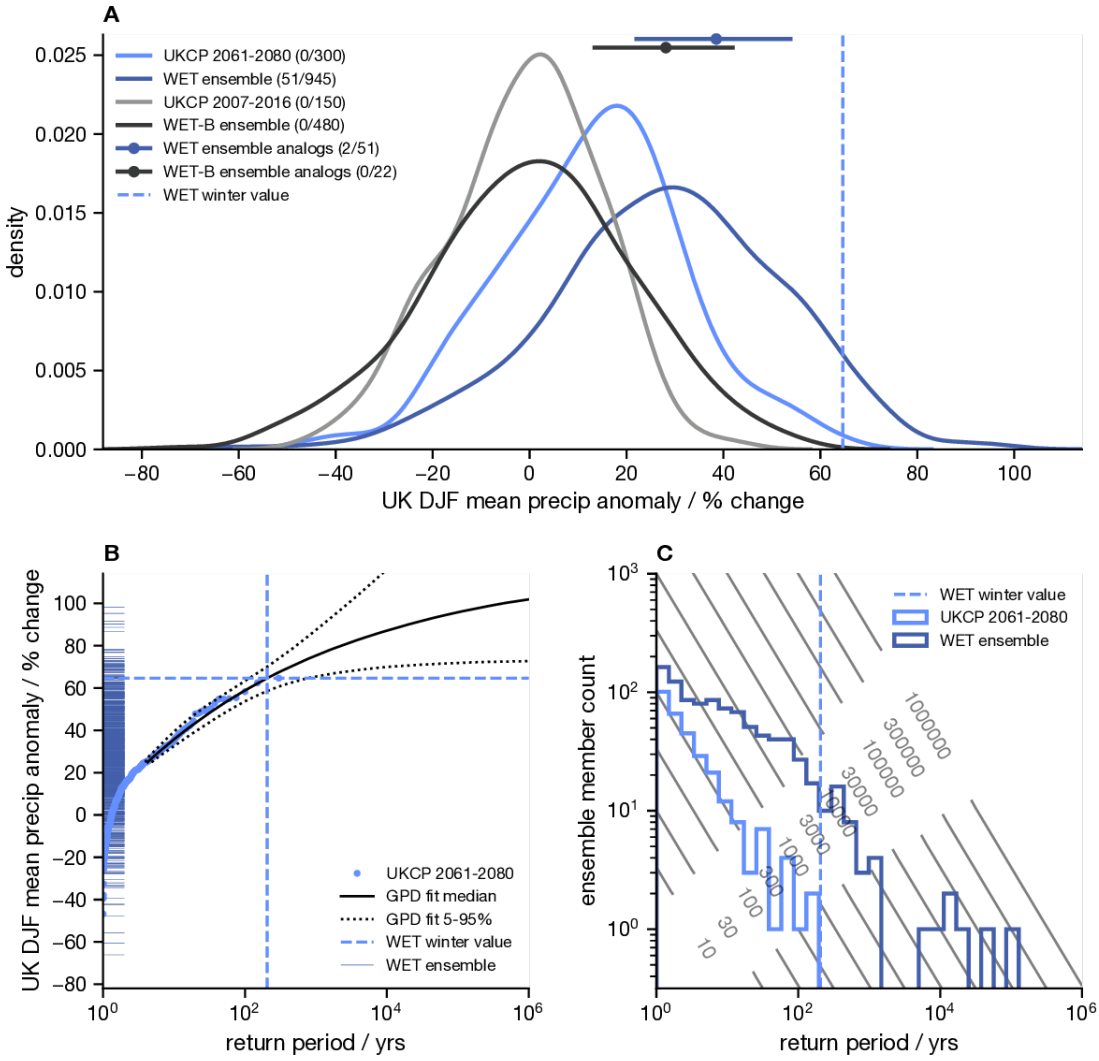

Figure 3.5: As Figure 3, but of DJF mean precipitation averaged over the UK region for the WET winter.

3.4.3 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 (*92*–*94*). 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 (TXx as previously defined). Figure 6 shows a return period diagram of UKCP 2061-2080 TXx deviations (centered on the mean anomaly for 2061-2071 over 2007-2016), plus a fitted generalised extreme value distribution (GEV) and associated uncertainty. GEVs are often used to statistically model block maxima of climate variables; and therefore infer information about the likelihood of such extreme events (*95*). 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.* (*96*). The dashed dark orange line in Figure 6 shows the TXx for HOT1 ensemble member c0qu, which lies considerably above (by 2.3 ℃) any UKCP18 PPE samples. This event is roughly 5 standard deviations above the mean of the UKCP18 deviations distribution shown in Figure 6. This is an example of a potential "record-shattering" event as discussed by Fischer et al. (2021). Since the particular GEV fitted to the UKCP18 deviations is type III (*77*), it sets an upper bound on TXx, consistent with previous studies of extreme heat events (*36*). 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 TXx 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 TXx distributions simulated by HadAM4 and HadGEM3-GC3.05; and differences in the response of those tails to climate change. We note that this exceptional TXx 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 (*60*, *97*). However, we believe that the key point to take away from this is not necessarily the specific estimated likelihood of these extreme weather events, but that the methodology used here could help to provide multivariate spatially, temporally and physically coherent examples of the kinds of H++ scenarios used to consider the limits to adaptation.

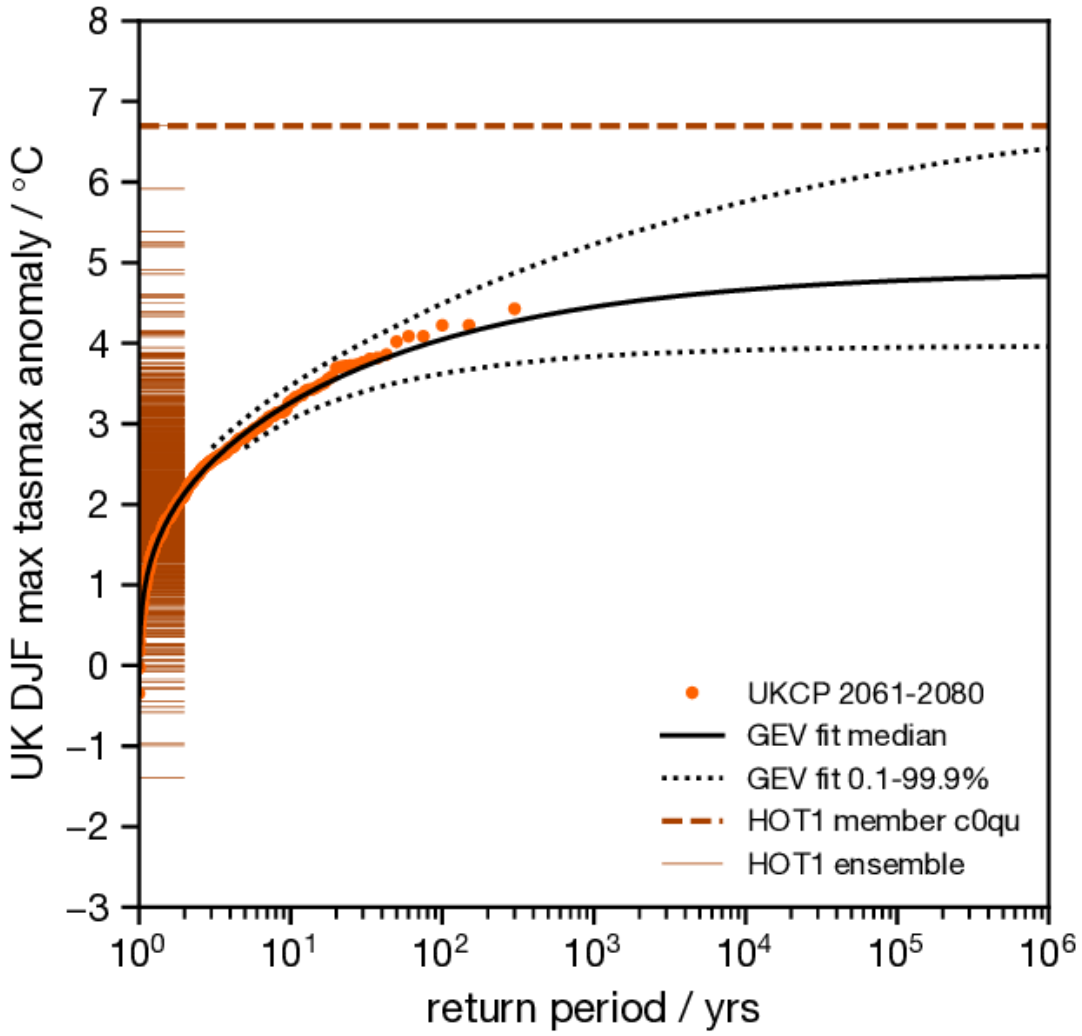

Figure 3.6: as return period diagram of Figure 3, but of DJF maximum of maximum daily temperatures averaged over the UK region for the HOT1 winter. The statistical model indicated by the solid and dotted black lines is a generalised extreme value distribution fit over the entire population of UKCP PPE 2061-2080 deviations, which are shown as light orange dots. Note the dotted lines indicate a 0.1 - 99.9 % CI in this instance. The dashed dark orange line shows the value of the most extreme member within the HOT1 ensemble.

3.5 Discussion

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

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

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

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

change, and the limits to adaptation planning (94). 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 (25, 26, 98, 99). 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 (98) and could potentially exhibit lower variability in other quantities, we have not found this to be the case for the mean UK temperature and precipitation studied here (though definitive proof of this would require us to repeat the ExSamples exercise with the coupled model). For the baseline period, the atmosphere-only model did not systematically underestimate the internal variability of the seasonal (or daily) timescale extreme variables considered here (Figures 3-5, S4-S12). Since we only have ExSamples future ensembles for three different sets of SST/SIC conditions, it is more difficult to quantify whether the projected internal variability is significantly different from the coupled model simulations, but the climate distributions of the relatively unconditioned HOT1 ensemble suggest that this is not the case.

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

Another research direction that could be taken would be to attempt to extract additional information from the existing set of events provided by the ExSamples ensembles presented here. Although the 60 km (N216) resolution of both the ExSamples ensemble and UKCP18 PPE is very competitive within the context of the current generation of climate models (100, 101), 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 (102). Hence, we suggest that the ExSamples ensembles could be statistically downscaled (or dynamically downscaled using a regional model if suitable model output was stored to drive these models) in order to provide information that is more relevant for localised climate change adaptation planning. Such downscaling could result in an extensive set of extreme local scenarios to complement the raw model output that provides a corresponding set of extreme national scenarios. For downscaling to be trustworthy, the large-scale dynamical features of the

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

3.6 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 (*32*, *48*).

3.7 Closing remarks

Quote

— author

4

Partial forecast-based attribution

This chapter contains much of the conceptual description of, and motivation for, forecast-based attribution. Using the well-predicted February 2019 heatwave as a case study, I carry out forecasts with the operational medium-range ECMWF model in which I have instantaneously perturbed the CO_2 concentration at initialisation. These perturbed forecasts allow me to estimate the direct contribution of diabatic heating due to CO_2 to the heatwave. This partial attribution provides a proof-of-concept of the forecast-based approach, and I close with a discussussion of how I could perform a more complete estimate of anthropogenic influence on a specific extreme event in following work.

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

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

Contents

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

4.1 Section

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

Quote

— author

5 Forecast-based attribution

Chapter description.

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

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

Contents

*with the author contributing as follows.

5.1 Section

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

Quote

— author

6 Discussion

Chapter description.

Contents

6.1 Section

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

6.2 Concluding remarks

Appendices

The first kind of intellectual and artistic personality belongs to the hedgehogs, the second to the foxes . . .

— Sir Isaiah Berlin (*103*)

References

1. C. Johnston, *Heatwave temperatures may top 45C in southern Europe*, Publication Title: The Guardian, Aug. 2018, (https://www.theguardian.com/world/2018/aug/04/temperatures-in-southern-europe-top-45-heatwave-spain-portugal).
2. NESDIS, *Record Summer Heat Bakes Europe*, Publication Title: NESDIS, Aug. 2018, (https://www.nesdis.noaa.gov/content/record-summer-heat-bakes-europe).
3. F. Krikken, F. Lehner, K. Haustein, I. Drobyshev, G. J. van Oldenborgh, English, *Natural Hazards and Earth System Sciences* **21**, Publisher: Copernicus GmbH, 2169–2179 (July 2021).
4. J. Watts, *Wildfires rage in Arctic Circle as Sweden calls for help*, Publication Title: The Guardian, July 2018, (https://www.theguardian.com/world/2018/jul/18/sweden-calls-for-help-as-arctic-circle-hit-by-wildfires).
5. Publico, *Nueve fallecidos por la ola de calor en España*, Publication Title: Publico, Aug. 2018, (https://www.publico.es/sociedad/nueve-fallecidos-ola-calor-espana.html).
6. C. Harris, *Heat, hardship and horrible harvests: Europe's drought explained*, Publication Title: Euronews, Aug. 2018, (https://www.euronews.com/2018/08/10/explained-europe-s-devastating-drought-and-the-countries-worst-hit).
7. World Weather Attribution, *Heatwave in northern Europe, summer 2018*, Publication Title: World Weather Attribution, July 2018, (https://www.worldweatherattribution.org/attribution-of-the-2018-heat-in-northern-europe/).
8. Press Office, *Chance of summer heatwaves now thirty times more likely*, Publication Title: Met Office News, Dec. 2018, (https://www.metoffice.gov.uk/about-us/press-office/news/weather-and-climate/2018/2018-uk-summer-heatwave).
9. M. McCarthy *et al.*, *Weather*, wea.3628 (Nov. 2019).
10. J. H. Christensen, O. B. Christensen, *Climatic Change* **81**, Publisher: Springer Netherlands, 7–30 (May 2007).
11. R. C. Cornes, G. van der Schrier, E. J. M. van den Besselaar, P. D. Jones, *Journal of Geophysical Research: Atmospheres* **123**, Publisher: John Wiley & Sons, Ltd, 9391–9409 (Sept. 2018).
12. D. D'Ippoliti *et al.*, *Environmental Health: A Global Access Science Source* **9** (2010).

13. J. Cattiaux, A. Ribes, J. Cattiaux, A. Ribes, *Bulletin of the American Meteorological Society* **99**, 1557–1568 (Aug. 2018).
14. N. Christidis *et al.*, *Journal of Climate* **26**, 2756–2783 (May 2013).
15. A. Ciavarella *et al.*, *Weather and Climate Extremes* **20**, Publisher: Elsevier, 9–32 (June 2018).
16. R. Vautard *et al.*, *Climate Dynamics* **41**, Publisher: Springer Berlin Heidelberg, 2555–2575 (Nov. 2013).
17. D. Jacob *et al.*, *Regional Environmental Change* **14**, Publisher: Springer Berlin Heidelberg, 563–578 (Apr. 2014).
18. M. Vrac, P. Vaittinada Ayar, M. Vrac, P. V. Ayar, *Journal of Applied Meteorology and Climatology* **56**, 5–26 (Jan. 2017).
19. E. E. Aalbers, G. Lenderink, E. van Meijgaard, B. J. J. M. van den Hurk, *Climate Dynamics* **50**, Publisher: Springer Berlin Heidelberg, 4745–4766 (June 2018).
20. G. Lenderink *et al.*, *Environmental Research Letters* **9**, Publisher: IOP Publishing, 115008 (Nov. 2014).
21. N. Massey *et al.*, *Quarterly Journal of the Royal Meteorological Society* **141**, Publisher: John Wiley & Sons, Ltd, 1528–1545 (July 2015).
22. N. S. Diffenbaugh *et al.*, *Proceedings of the National Academy of Sciences* **114**, Publisher: National Academy of Sciences, 4881–4886 (May 2017).
23. S. Jeon, C. J. Paciorek, M. F. Wehner, *Weather and Climate Extremes* **12**, Publisher: Elsevier, 24–32 (June 2016).
24. O. Bellprat, V. Guemas, F. Doblas-Reyes, M. G. Donat, *Nature Communications* **10**, Publisher: Nature Publishing Group, 1732 (Dec. 2019).
25. E. M. Fischer, U. Beyerle, C. F. Schleussner, A. D. King, R. Knutti, *Geophysical Research Letters* **45**, Publisher: John Wiley & Sons, Ltd, 8500–8509 (Aug. 2018).
26. J. He, B. J. Soden, EN, *Journal of Climate* **29**, Publisher: American Meteorological Society Section: Journal of Climate, 4317–4325 (June 2016).
27. WEF, "The Global Risks Report 2021", en, tech. rep. (World Economic Forum, 2021), (2021; https://www.weforum.org/reports/the-global-risks-report-2021/).
28. S. I. Seneviratne *et al.*, 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*, ed. by V. Masson-Delmotte *et al.* (Cambridge University Press, 2021), (https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_12.pdf).
29. M. R. Allen *et al.*, *Nature* **458**, ISBN: 1476-4687 (Electronic)\r0028-0836 (Linking), 1163–1166 (2009).
30. S. Rahmstorf, D. Coumou, *Proceedings of the National Academy of Sciences of the United States of America* **108**, Publisher: National Academy of Sciences, 17905–17909 (2011).

31. J. A. Lowe *et al.*, "UKCP18 Science Overview Report", tech. rep. (Met Office Hadley Centre, 2018), (2021; https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf).

32. J. Murphy *et al.*, "UKCP18 Land Projections: Science Report", tech. rep. (Met Office, Exeter, 2018), (https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf).

33. T. Kelder *et al.*, *npj Climate and Atmospheric Science* **3**, 47 (2020).

34. V. Thompson *et al.*, *Nature Communications* **8**, Publisher: Nature Publishing Group, 1–6 (2017).

35. E. M. Fischer, S. Sippel, R. Knutti, en, *Nature Climate Change*, Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate and Earth system modelling;Projection and prediction Subject_term_id: climate-and-earth-system-modelling;projection-and-prediction, 1–7 (2021).

36. C. Gessner, E. M. Fischer, U. Beyerle, R. Knutti, EN, *Journal of Climate* **34**, Publisher: American Meteorological Society Section: Journal of Climate, 6619–6634 (2021).

37. K. Riahi *et al.*, *Climatic Change* **109**, 33–57 (2011).

38. K. D. Williams *et al.*, en, *Journal of Advances in Modeling Earth Systems* **10**, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2017MS001115, 357–380 (2018).

39. A. V. Karmalkar *et al.*, en, *Climate Dynamics* **53**, 847–877 (2019).

40. D. M. H. Sexton *et al.*, en, *Climate Dynamics* **53**, 989–1022 (2019).

41. D. M. H. Sexton *et al.*, en, *Climate Dynamics* **56**, 3395–3436 (2021).

42. K. Yamazaki *et al.*, en, *Climate Dynamics* **56**, 3437–3471 (2021).

43. D. Sexton, K. Yamazaki, J. Murphy, J. Rostron, "Assessment of drifts and internal variability in UKCP projections", en, tech. rep. (Met Office, 2020), p. 20, (https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp-climate-drifts-report.pdf).

44. M. Webb, C. Senior, S. Bony, J.-J. Morcrette, en, *Climate Dynamics* **17**, 905–922 (2001).

45. K. D. Williams, M. A. Ringer, C. A. Senior, en, *Climate Dynamics* **20**, 705–721 (2003).

46. V. D. Pope, M. L. Gallani, P. R. Rowntree, R. A. Stratton, en, *Climate Dynamics* **16**, 123–146 (2000).

47. E. Bevacqua *et al.*, en, *Geophysical Research Letters* **48**, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL091990, e2020GL091990 (2021).

48. P. Watson *et al.*, “Multi-thousand member ensemble atmospheric simulations with global 60km resolution using climateprediction.net”, en, tech. rep. EGU2020-10895, Conference Name: EGU2020 (Copernicus Meetings, 2020), (2021; https://meetingorganizer.copernicus.org/EGU2020/EGU2020-10895.html).

49. M. Allen, en, *Nature* **401**, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6754 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group, 642–642 (1999).

50. D. Anderson, presented at the Fifth IEEE/ACM International Workshop on Grid Computing, ISSN: 1550-5510, pp. 4–10.

51. D. Stainforth *et al.*, *Computing in Science Engineering* **4**, Conference Name: Computing in Science Engineering, 82–89 (2002).

52. A. Brown *et al.*, EN, *Bulletin of the American Meteorological Society* **93**, Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 1865–1877 (2012).

53. D. A. Stainforth *et al.*, en, *Nature* **433**, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7024 Primary_atype: Research Publisher: Nature Publishing Group, 403–406 (2005).

54. D. Frame *et al.*, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **367**, Publisher: Royal Society, 855–870 (2009).

55. P. Pall *et al.*, *Nature* **470**, 382–385 (2011).

56. E. Kendon *et al.*, “UKCP Convection-permitting model projections: Science report”, English, tech. rep. (Met Office Hadley Centre, Exeter, 2019), (2022; https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP-Convection-permitting-model-projections-report.pdf).

57. S. Sparrow, D. Sexton, N. J. Leach, P. A. G. Watson, D. C. H. Wallom, *Scientific Data* **in prep.** (2021).

58. D. M. H. Sexton, D. P. Rowell, C. K. Folland, D. J. Karoly, en, *Climate Dynamics* **17**, 669–685 (2001).

59. R. Neal, D. Fereday, R. Crocker, R. E. Comer, *Meteorological Applications* **23**, Publisher: John Wiley and Sons Ltd, 389–400 (2016).

60. M. Kendon, D. Sexton, M. McCarthy, *Weather* **75**, Publisher: John Wiley and Sons Ltd, 318–324 (2020).

61. C. Deser, M. A. Alexander, S.-P. Xie, A. S. Phillips, *Annual Review of Marine Science* **2**, _eprint: https://doi.org/10.1146/annurev-marine-120408-151453, 115–143 (2010).

62. W. T. K. Huang *et al.*, en, *Environmental Research Letters* **15**, Publisher: IOP Publishing, 124052 (2020).

63. D. Richardson *et al.*, en, *Meteorological Applications* **27**, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/met.1931, e1931 (2020).

64. D. Richardson, H. J. Fowler, C. G. Kilsby, R. Neal, *International Journal of Climatology* **38**, Publisher: John Wiley and Sons Ltd, 630–648 (2018).

65. C. Deser, I. R. Simpson, K. A. McKinnon, A. S. Phillips, EN, *Journal of Climate* **30**, Publisher: American Meteorological Society Section: Journal of Climate, 5059–5082 (2017).

66. M. P. King, E. Yu, J. Sillmann, *Tellus A: Dynamic Meteorology and Oceanography* **72**, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/16000870.2019.1704342, 1–10 (2020).

67. M. P. King *et al.*, EN, *Bulletin of the American Meteorological Society* **99**, Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 1337–1343 (2018).

68. J. López-Parages, B. Rodríguez-Fonseca, D. Dommenget, C. Frauen, en, *Climate Dynamics* **47**, 2071–2084 (2016).

69. J. A. Francis, S. J. Vavrus, en, *Geophysical Research Letters* **39**, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012GL051000, (2021; https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012GL051000) (2012).

70. R. A. Pedersen, I. Cvijanovic, P. L. Langen, B. M. Vinther, EN, *Journal of Climate* **29**, Publisher: American Meteorological Society Section: Journal of Climate, 889–902 (2016).

71. M. Kretschmer, G. Zappa, T. G. Shepherd, English, *Weather and Climate Dynamics* **1**, Publisher: Copernicus GmbH, 715–730 (2020).

72. J. A. Screen, I. Simmonds, en, *Geophysical Research Letters* **40**, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/grl.50174, 959–964 (2013).

73. J. A. Screen, en, *Nature Communications* **8**, Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Atmospheric dynamics;Climate change;Climate sciences Subject_term_id: atmospheric-dynamics;climate-change;climate-sciences, 14603 (2017).

74. J. R. Hosking, J. R. Wallis, E. F. Wood, *Technometrics* **27**, 251–261 (1985).

75. J. R. M. Hosking, *Journal of the Royal Statistical Society: Series B (Methodological)* **52**, Publisher: [Royal Statistical Society, Wiley], 105–124 (1990).

76. J. R. M. Hosking, J. R. Wallis, *Regional Frequency Analysis*, Publication Title: Regional Frequency Analysis (Cambridge University Press, Cambridge, 1997), (https://www.cambridge.org/core/books/regional-frequency-analysis/8C59835F9361705DAAE1ADFDEA7ECD30%20https://www.cambridge.org/core/product/identifier/9780511529443/type/book).

77. S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, en (Springer-Verlag, London, 2001), (2021; https://www.springer.com/gp/book/9781852334598).

78. J. R. M. Hosking, J. R. Wallis, *Technometrics* **29**, Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1987.10488243, 339–349 (1987).

79. J. Kysely, J. Picek, English, presented at the Advances in Geosciences, vol. 12, ISSN: 1680-7340, pp. 43–50, (2021; https://adgeo.copernicus.org/articles/12/43/2007/).

80. W. Z. Wan Zin, A. A. Jemain, K. Ibrahim, en, *Theoretical and Applied Climatology* **96**, 337–344 (2009).

81. M. Marani, M. Ignaccolo, en, *Advances in Water Resources* **79**, 121–126 (2015).

82. J. Cattiaux *et al.*, *Geophysical Research Letters* **37**, Publisher: Blackwell Publishing Ltd (2010).

83. R. Vautard *et al.*, *Environmental Research Letters* **11**, Publisher: IOP Publishing, 114009 (2016).

84. P. Yiou *et al.*, *Advances in Statistical Climatology, Meteorology and Oceanography* **3**, Publisher: Copernicus GmbH, 17–31 (2017).

85. J. L. Hodges, en, *Arkiv för Matematik* **3**, 469–486 (1958).

86. A. N. Kolmogorov, *Giornale dell'Istituto Italiano degli Attuari* **4**, 83–91 (1933).

87. N. Smirnoff, *Bulletin Mathématique de L'Université de Moscow* **2**, 3–11 (1939).

88. N. Smirnoff, *Matematicheskii Sbornik* **6(48)**, 3–26 (1939).

89. D. A. Stone, M. R. Allen, *Climatic Change* **71**, Publisher: Springer, 303–318 (2005).

90. P. A. Stott, D. A. Stone, M. R. Allen, *Nature* **432**, Publisher: Nature Publishing Group, 610–614 (2004).

91. S. Brönnimann, en, *Reviews of Geophysics* **45**, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006RG000199, (2021; https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006RG000199) (2007).

92. D. King, D. Schrag, Z. Dadi, Q. Ye, A. Ghosh, "Climate change: A risk assessment", tech. rep. (UK Foreign & Commonwealth Office, London, 2015), (https://www.csap.cam.ac.uk/projects/climate-change-risk-assessment/).

93. J. A. Lowe *et al.*, "UK Climate Projections science report: Marine and coastal projections", tech. rep. (Met Office Hadley Centre, Exeter, UK, 2009), (https://webarchive.nationalarchives.gov.uk/ukgwa/20181204111026mp_/http://ukclimateprojections-ukcp09.metoffice.gov.uk/media.jsp?mediaid=87906&filetype=pdf).

94. S. Wade *et al.*, "Developing H++ climate change scenarios for heat waves, droughts, floods, windstorms and cold snaps", tech. rep., This report has been produced by the Met Office, University of Reading and CEH for the Adaptation Sub-Committee of the Committee on Climate Change and to support the second Climate Change Risk Assessment (CCRA). Freely available online via Official URL link. (Committee on Climate Change, London, 2015), p. 144, (2021; https://www.theccc.org.uk/publication/met-office-for-the-asc-developing-h-climate-change-scenarios/).

95. S. J. Brown, J. M. Murphy, D. M. H. Sexton, G. R. Harris, en, *Climate Dynamics* **43**, 2681–2705 (2014).

96. S. Sippel *et al.*, *Weather and Climate Extremes* **9**, Publisher: Elsevier B.V., 25–35 (2015).

97. M. Young, J. Galvin, *Weather* **75**, Publisher: John Wiley and Sons Ltd, 36–45 (2020).

98. J. J. Barsugli, D. S. Battisti, EN, *Journal of the Atmospheric Sciences* **55**, Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences, 477–493 (1998).

99. B. Dong, R. T. Sutton, L. Shaffrey, N. P. Klingaman, EN, *Journal of Climate* **30**, Publisher: American Meteorological Society Section: Journal of Climate, 6203–6223 (2017).

100. C. source id values, *CMIP6 source id values*, type: dataset version: 6.2.56.8, 2022, (2021; https://wcrp-cmip.github.io/CMIP6_CVs/docs/CMIP6_source_id.html).

101. V. Eyring *et al.*, *Geoscientific Model Development* **9**, Publisher: Copernicus GmbH, 1937–1958 (2016).

102. R. Charlton, R. Fealy, S. Moore, J. Sweeney, C. Murphy, en, *Climatic Change* **74**, 475–491 (2006).

103. I. Berlin, *The Hedgehog and the Fox: An Essay on Tolstoy's View of History*, English, ed. by H. Hardy (Princeton University Press, 2nd, 2013).