Generating samples of extreme winters to support climate adaptation

Nicholas J. Leach[1•], Peter A. G. Watson[2], Sarah N. Sparrow[3], David C. H. Wallom[3], David M. H. Sexton[4]

[1] Atmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

[2] School of Geographical Sciences, University of Bristol, Bristol, UK

[3] Oxford e-Research Centre, Engineering Science, University of Oxford, Oxford, UK

[4] Met Office Hadley Centre, Exeter, UK

• Corresponding author. *Email:* nicholas.leach@stx.ox.ac.uk

Key Points

- We create 3x1000-member ensembles in an atmosphere-only model, taking boundary conditions from coupled simulations of future extreme winters
- The boundary conditions in two of the three winters favour extreme winters and we find examples that are beyond 1-in-10000 year events
- These high-end extremes could be used for more robust adaptation planning and climate scenario development

Abstract

Recent extreme weather in the UK 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 use 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 return periods over 100 years, 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 future extreme winters within the UKCP18 ensemble. We find that every UKCP18 extreme winter 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 simulated lie above those for UKCP18 by 0.85K for daily maximum temperature and 37% of the present-day average for precipitation (UK winter means).

Plain language summary

The risk from extreme weather events is important to understand due to the damage that they can cause. This study is concerned with understanding extremes both in the present-day, and how they may change with global warming. Extreme weather events can be difficult to study with climate models due to their rarity: sampling the most extreme events requires very long simulations. In a climate model containing a representation of both the atmosphere and ocean such long simulations would be prohibitively expensive. In this study

we explore a new approach for efficiently sampling extremes over the UK (though it is applicable to other regions). Starting with a small sample of coupled climate model simulations, we use surface temperature and sea ice conditions from the most extreme future winters within these coupled simulations to drive an atmosphere-only model. We run this model on volunteers' personal computers, which allows us to create very large ensembles of simulated winter weather. Our ensembles contain many samples of very extreme events; some sea surface temperature patterns condition the ensemble towards producing extremes. Our approach could be used to explore the uncertainty surrounding extreme weather events at the present-day, or create climate scenarios to inform adaptation planning.

Key words

climate modelling | climate change projection | understanding extreme weather events | large ensembles | climate change adaptation

1. Introduction

Weather extremes are one of the most damaging hazards that society faces at the present-day (*The Global Risks Report 2021*, 2021). 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 (Seneviratne et al., 2021). 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 (M. R. Allen et al., 2009; Diffenbaugh et al., 2017; Rahmstorf & Coumou, 2011). 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 (Lowe et al., 2018; Murphy et al., 2018) 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 set of 15 PPE simulations, and our analysis and results build upon the information provided by these global 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. 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. 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 (Fischer et al., 2021; Gessner et al., 2021). 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.

2. Study design and methods

2.1 Models

2.1.1 HadGEM3-GC3.05 global climate model

In addition to the ExSamples ensembles, we also analyse UKCP18 global PPE simulations of the RCP8.5 emission scenario (Riahi et al., 2011). This PPE is based on the global HadGEM3-GC3.05 coupled ocean atmosphere model (Murphy et al., 2018; Williams et al., 2018). 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 (Karmalkar et al., 2019; D. M. H. Sexton et al., 2019, 2021; Yamazaki et al., 2021). 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 (Murphy et al., 2018), minus two members that displayed a significant weakening of the Atlantic Meridional Overturning Circulation (D. Sexton et al., 2020). 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.

2.1.2 HadAM4 N216 atmospheric model

The novel simulations presented here are performed by the global HadAM4 atmosphere and land surface model (Webb et al., 2001; Williams et al., 2003). Like its predecessor, HadAM3 (Pope et al., 2000), 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 (Bevacqua et al., 2021; Watson et al., 2020), 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 (M. Allen, 1999; Anderson, 2004; D. Stainforth et al., 2002). This system has been used previously to run a range of Hadley Center Unified Model variants (A. Brown et al., 2012), including a coupled atmosphere-slab ocean model (D. A. Stainforth et al., 2005), a fully coupled model (Frame et al., 2009) and an atmosphere-only model (Pall et al., 2011) 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”.

2.2 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”.

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 (M. Allen, 1999; D. A. Stainforth et al., 2005), our target final ensemble size is 1500 members, 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 (Sparrow et al., 2021).

3. We then 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 study winter.

2.2.1 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 (Sexton et al., 2001). 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 a priori 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.

2.2.2 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. Deviations of the UKCP18 PPE also provide the closest simple comparison to the atmosphere-only ExSamples ensembles which only sample atmospheric internal variability.

2.2.3 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.

We aimed to select two “hot” winters and one “wet” winter. As shown in Figure 1, there is one clear candidate for each type of extreme: UKCP18 PPE member 02868 (ID numbers as Sexton et al., 2021) 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.

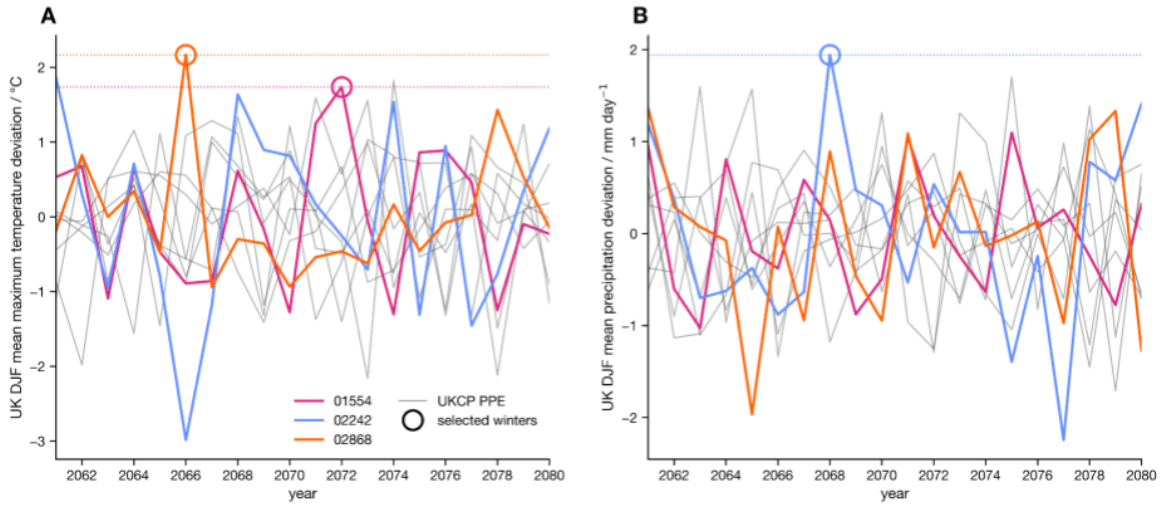

FIG 1: UKCP PPE 2061-2080 deviations. **A**, DJF mean of daily maximum temperatures averaged over the UK region. Coloured lines indicate the three UKCP runs from which the study winters were chosen. The study winters are circled and dotted horizontal lines indicate the deviation of each study winter. 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) info			Abbreviation used
	UKCP18 member	Year	Extreme type	
Future projections	02868	2066	HOT	HOT1
	01554	2072	HOT	HOT2
	02242	2068	WET	WET1
Baseline ensembles	02868	2007-2016	-	HOT1-B
	01554	2007-2016	-	HOT2-B
	02242	2007-2016	-	WET1-B

Table 1: Summary of experiments performed for ExSamples project.

2.2.4 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. This shares similarities with several weather patterns, including 20 and 23. There is a strong positive ENSO phase during this winter, alongside moderately positive Atlantic Multidecadal Variability and negative phase Pacific Decadal Oscillation. This extreme winter places between the other two in terms of SIC - mean Arctic sea ice fraction is approximately 70 %.

The HOT2 winter displays a similar MSLP pattern to the first hot winter. In terms of the 30 weather patterns developed by the Met Office (Neal et al., 2016), the mean large scale flow over the whole winter is closest to weather pattern 20. This weather pattern is associated with warm and wet weather over the UK (Huang et al., 2020; Richardson et al., 2018, 2020), and has also been shown to be conducive to producing record temperature extremes on daily timescales (Kendon et al., 2020). During this winter, the El Nino Southern Oscillation (ENSO) pattern of global SST variability was in a weak La Nina (negative) phase (Deser et al., 2010); this phase has previously been linked to an increased likelihood of positive NAO (Deser et al., 2017; M. P. King et al., 2018, 2020; López-Parages et al., 2016). No other modes of SST variability are present. With regards to SIC, this particular PPE member has virtually entirely 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 (Francis & Vavrus, 2012; Pedersen et al., 2016), though this is still a subject of active scientific interest (Kretschmer et al., 2020; Screen, 2017; Screen & Simmonds, 2013).

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.

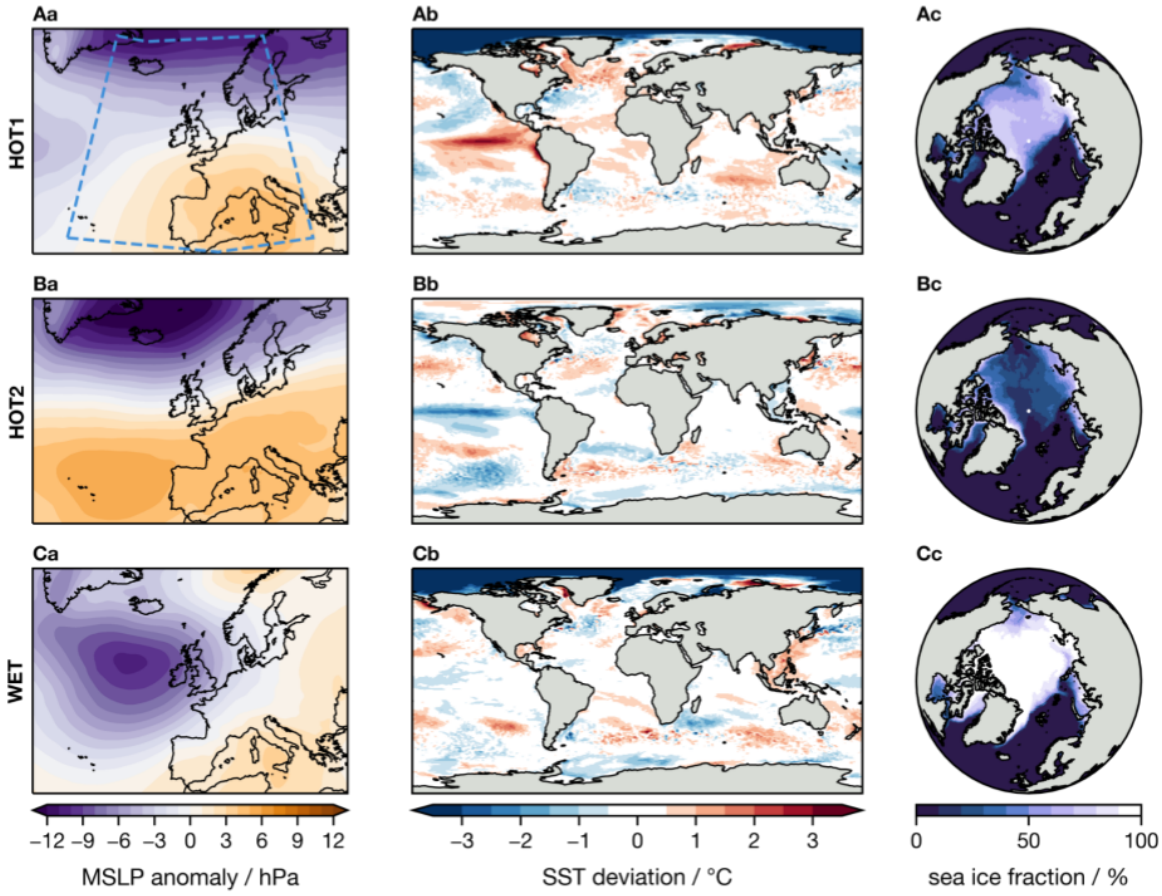

FIG 2: synoptic features of the study winters within the UKCP simulations. The row titles indicate the study winter. **a**, DJF mean MSLP anomalies for each winter. **b**, DJF mean SST deviations for each winter. Deviations are calculated for each gridpoint timeseries over 2061-2080. **c**, DJF mean Arctic sea ice fraction for each winter. The blue dashed line in Aa indicates the area used for analog subsampling.

2.3 Statistical methods

2.3.1 Estimating distributions of extremes

We estimate distributions using the method of L-moments (Hosking, 1990; Hosking et al., 1985; Hosking & Wallis, 1997). 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 as follows:

Mean DJF daily maximum temperatures (TXm) & daily mean precipitation rate (PRm)

We use a generalised pareto distribution (Coles, 2001; Hosking & Wallis, 1987) fit to the upper quantile 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.

Maximum DJF daily maximum temperatures (TXx)

We use a generalised extreme value distribution fit to the sample population.

Maximum DJF daily mean precipitation rate (PRx)

We use a generalised logistic distribution (Hosking, 1990) 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 (Kysely & Picek, 2007; Wan Zin et al., 2009). This issue is not a feature of the L-Moments estimator used: a maximum likelihood estimator yields near-identical results. The apparent discrepancy with the generalised extreme value distribution may arise as the number of independent precipitation events per season may not be 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 (Marani & Ignaccolo, 2015).

2.3.2 Analog construction

In order to assess the dynamical contributions to the extreme weather simulated during the study winters, we use an MSLP analog approach (Cattiaux et al., 2010; Vautard et al., 2016; Yiou et al., 2017). 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 the top left subplot of Figure 2 (-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 (Neal et al., 2016). 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. Results

3.1 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 (150 members total) with each of the three 2007-2016 ExSamples baseline ensembles (~500 members each) in turn, and their aggregate ensemble. Here we quantify whether the simulated climates differ using a two-sample Kolmogorov-Smirnoff (K-S) test (Hodges, 1958; Kolmogorov, 1933; Smirnoff, 1939a, 1939b) 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 (Murphy et al., 2018) and HadAM4 (Bevacqua et al., 2021; Watson et al., 2020).

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.

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 Analysis in the Supplementary Information. 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 percentages, the differences

in the spread of the distributions becomes relatively larger (Figure 5) and the distributions of percentage anomalies are statistically significantly different.

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.

3.2 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.

3.2.1 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 (Stone & Allen, 2005; Stott et al., 2004). 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).

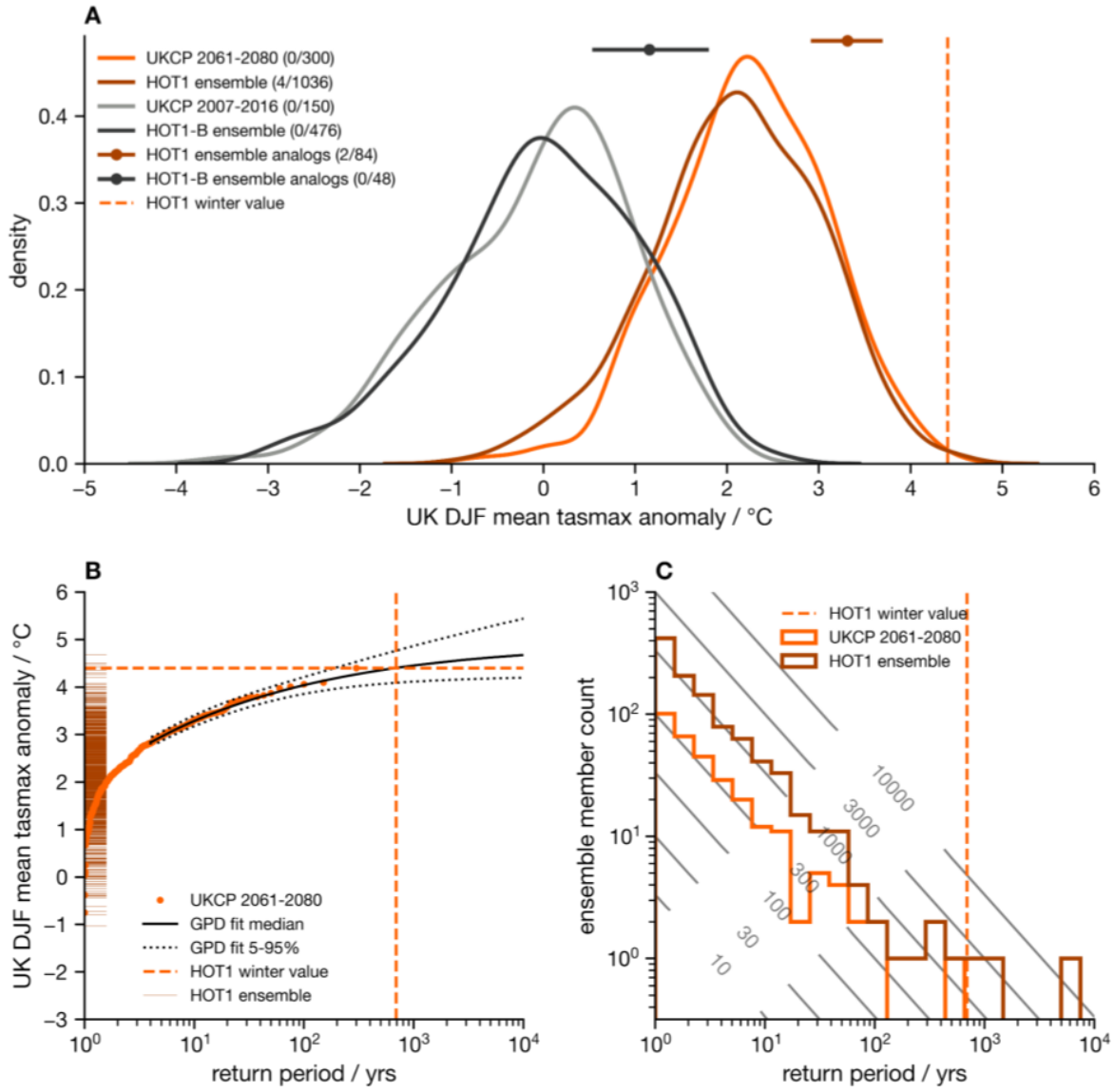

FIG 3: comparing statistics of DJF mean of daily maximum temperatures averaged over the UK region for the HOT1 winter. **A**, PDFs of baseline and future ensembles. The light orange PDF shows UKCP 2061-2080 deviations, with the distribution mean set to the ensemble mean anomaly between 2007-2016 and 2061-2071. The dark orange PDF shows HOT1 ensemble anomalies. The light grey PDF shows UKCP 2007-2016 anomalies. The dashed horizontal light orange line indicates HOT1 winter deviation. The dark orange and black dotted bars indicate 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. Black 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**.

		UKCP18 quantile (return period)		
Study winter	**Variable**	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]
WET1	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 2: Relative risk of three extreme thresholds in ExSamples future ensembles compared to UKCP18 PPE 2061-2080 deviations. Square brackets indicate a 90 % CI.

3.2.2 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 preconditioned 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 preconditioning, 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 (Brönnimann, 2007; Deser et al., 2017).

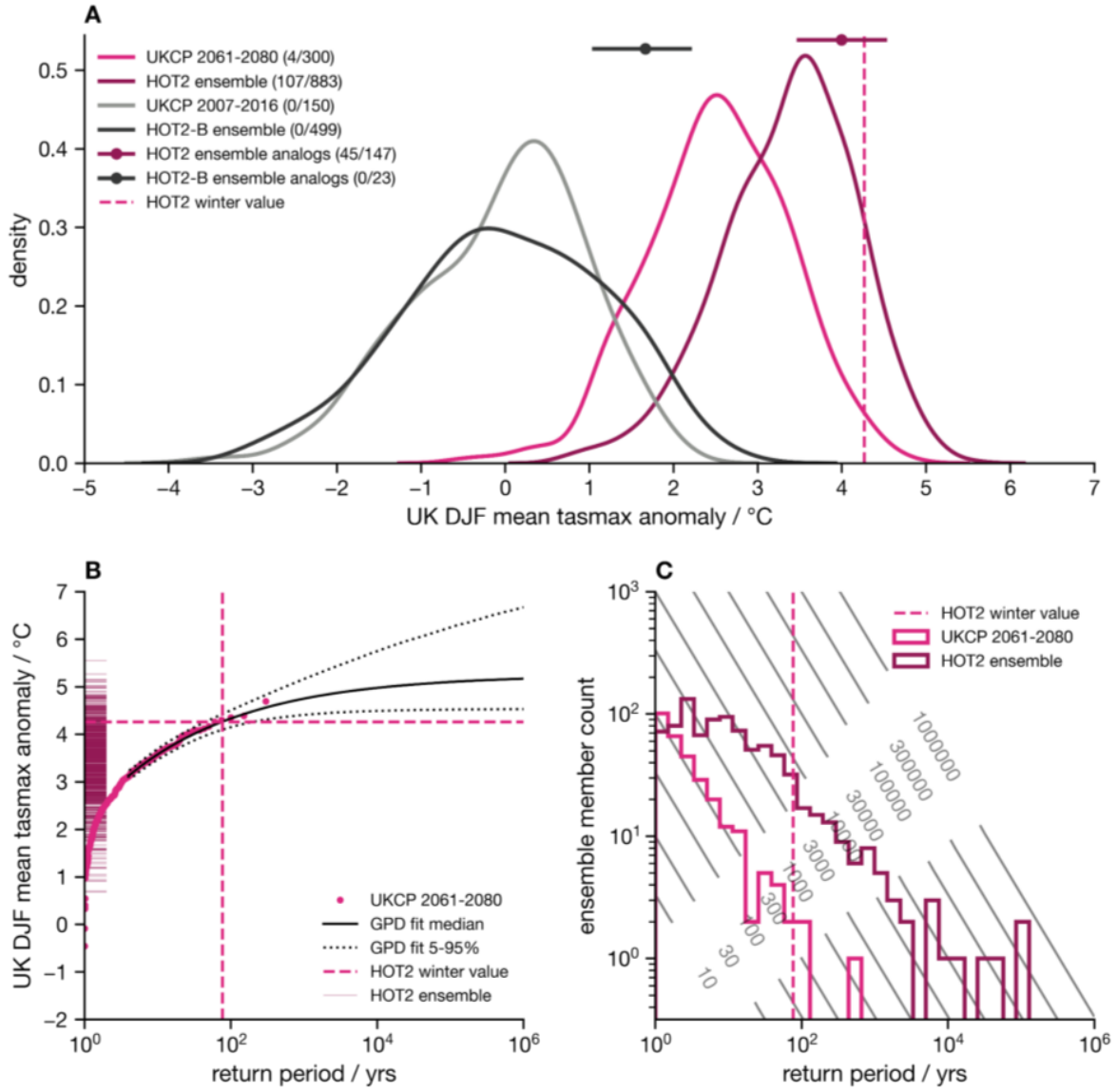

FIG 4: as FIG 4, but for the HOT2 winter.

3.2.3 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 preconditioning present in the 2072 ensemble.

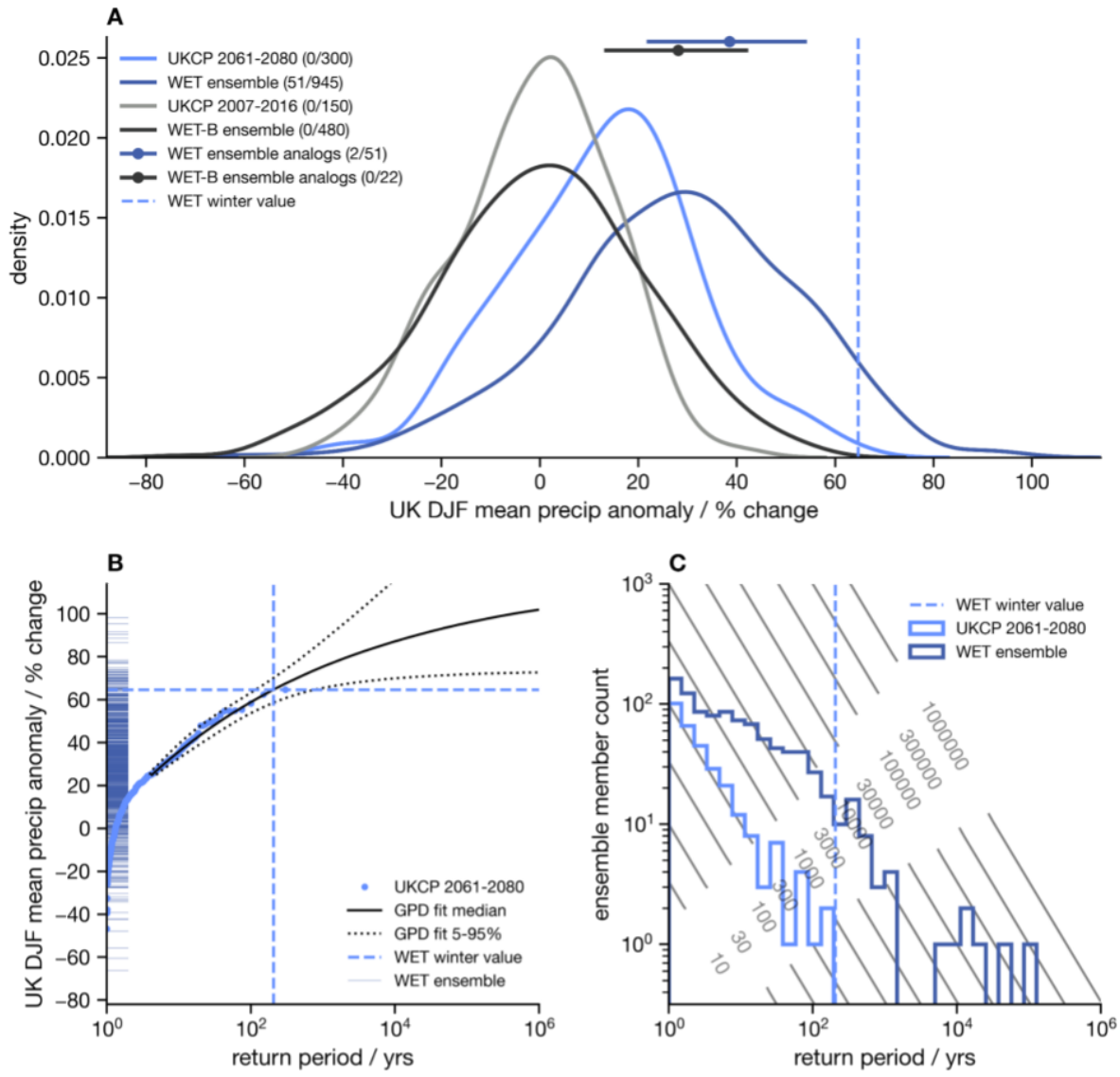

FIG 5: as FIG 4, but of DJF mean precipitation averaged over the UK region for the WET winter.

3.3 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 (D. King et al., 2015; Lowe et al., 2009; Wade

et al., 2015). 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 (S. J. Brown et al., 2014). 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. (2015). The dashed dark orange line in Figure 6 shows the TXx for HOT1 ensemble member c0qu, which lies considerably above (by 2.3 °C) any UKCP18 PPE samples. This event is roughly 5 standard deviations above the mean of the UKCP18 deviations distribution shown in Figure 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 (Coles, 2001), it sets a theoretical upper bound on TXx, consistent with the physical laws governing energy transfer in the climate system. 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 is not due to a mean bias between the two models: they display near-identical climatological distributions of TXx over the baseline period. 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 (Kendon et al., 2020; Young & Galvin, 2020). Therefore, the methodology used here could help to provide examples of the kinds of H++ scenarios used to consider the limits to adaptation.

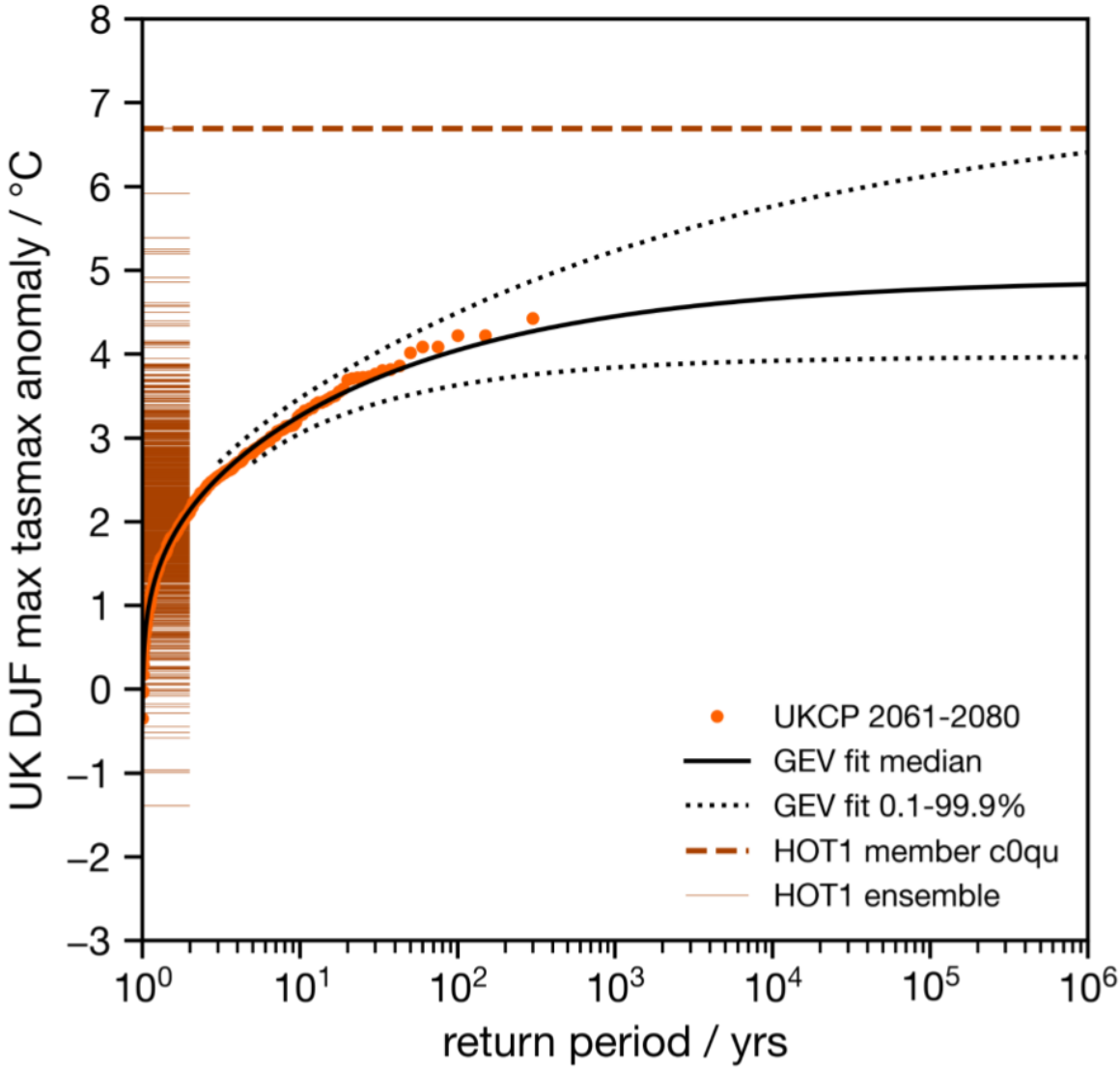

FIG 6: as return period diagram of FIG 4, but of DJF maximum of maximum daily temperatures averaged over the UK region for the HOT1 winter. The statistical model fit indicated by the solid and dotted black lines is a generalised extreme value distribution over the entire population. Note the dotted lines 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.

4. 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 (Wade et al., 2015). 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. 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). It has previously been found that atmosphere-only simulations can underestimate the internal variability of the climate system (Fischer et al., 2018; He & Soden, 2016), thereby producing

biases in the estimated frequency of extreme events and their changes in frequency with climate change, though this is an area of ongoing research (Barsugli & Battisti, 1998; Copsey et al., 2006; Dong et al., 2017; Saravanan, 1998). However, for the UK region studied here, we have not found that this is the case. 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 (*CMIP6 Source_id Values*, n.d.; Eyring et al., 2016), 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 (Charlton et al., 2006). Hence, we suggest that the ExSamples ensembles could be statistically downscaled (or dynamically downscaled using a regional model if suitable model output was stored to drive these models) in order to provide information that is more relevant for localised climate change adaptation planning. Such downscaling could result in an extensive set of extreme local scenarios to complement the raw model output that provides a corresponding set of extreme national scenarios. For downscaling to be trustworthy, the large-scale dynamical features of the input simulations must be an accurate representation of reality. The analysis that we have performed here suggests this is the case: as demonstrated in the Supplementary Information, the large-scale dynamics over the Euro-Atlantic sector within HadAM4 very closely replicates those within HadGEM3-GC3.05.

5. 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 physically plausible simulations 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 (Murphy et al., 2018; Watson et al., 2020).

Acknowledgements

NJL was supported by the Natural Environment Research Council (grant no. NE/L002612/1). PAGW was supported by a Natural Environmental Research Council Independent Research Fellowship (grant no. NE/S014713/1).

We thank Myles R. Allen for his input to the initial discussions of this project, and linking the members of this authorship team up. We thank Jason Lowe for his helpful comments and suggestions regarding the text.

We thank all of the volunteers who have donated their computing time to climateprediction.net to perform the HadAM4 simulations. The UK Climate Resilience programme is supported by the UKRI Strategic Priorities Fund. The programme is co-delivered by the Met Office and NERC on behalf of UKRI partners AHRC, EPSRC, ESRC.

Code and data availability

The code used to carry out the analysis and produce the figures within this study are available at https://github.com/njleach/ExSamples-analysis-notebooks.

The novel ExSamples data used in this study are available from MASS via the CEDA archive (https://help.jasmin.ac.uk/category/227-mass). They are located at /adhoc/projects/qump_hadgem3/ExSamples/netcdf/product/20210622. The UKCP data used in this study are available from the CEDA archive (http://data.ceda.ac.uk/badc/ukcp18/data/land-gcm/global/60km/rcp85).

References

Allen, M. (1999). Do-it-yourself climate prediction. *Nature*, *401*(6754), 642–642. https://doi.org/10.1038/44266

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., & Meinshausen, N. (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, *458*(7242), 1163–1166. https://doi.org/10.1038/nature08019

Anderson, D. P. (2004). BOINC: A system for public-resource computing and storage. *Fifth IEEE/ACM International Workshop on Grid Computing*, 4–10. https://doi.org/10.1109/GRID.2004.14

Baker, H. S., Woollings, T., Forest, C. E., & Allen, M. R. (2019). The Linear Sensitivity of the North Atlantic Oscillation and Eddy-Driven Jet to SSTs. *Journal of Climate*, *32*(19), 6491–6511. https://doi.org/10.1175/JCLI-D-19-0038.1

Barsugli, J. J., & Battisti, D. S. (1998). The Basic Effects of Atmosphere–Ocean Thermal Coupling on Midlatitude Variability. *Journal of the Atmospheric Sciences*, *55*(4), 477–493. https://doi.org/10.1175/1520-0469(1998)055<0477:TBEOAO>2.0.CO;2

Bevacqua, E., Shepherd, T. G., Watson, P. A. G., Sparrow, S., Wallom, D., & Mitchell, D. (2021). Larger Spatial Footprint of Wintertime Total Precipitation Extremes in a Warmer Climate. *Geophysical Research Letters*, *48*(8), e2020GL091990. https://doi.org/10.1029/2020GL091990

Brönnimann, S. (2007). Impact of El Niño–Southern Oscillation on European climate. *Reviews of Geophysics*, *45*(3). https://doi.org/10.1029/2006RG000199

Brown, A., Milton, S., Cullen, M., Golding, B., Mitchell, J., & Shelly, A. (2012). Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey. *Bulletin of the American Meteorological Society*, *93*(12), 1865–1877. https://doi.org/10.1175/BAMS-D-12-00018.1

Brown, S. J., Murphy, J. M., Sexton, D. M. H., & Harris, G. R. (2014). Climate projections of future extreme events accounting for modelling uncertainties and historical simulation biases. *Climate Dynamics*, *43*(9), 2681–2705. https://doi.org/10.1007/s00382-014-2080-1

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., & Codron, F. (2010). Winter 2010 in Europe: A cold extreme in a warming climate. *Geophysical Research Letters*, *37*(20). https://doi.org/10.1029/2010GL044613

Charlton, R., Fealy, R., Moore, S., Sweeney, J., & Murphy, C. (2006). Assessing the Impact of Climate Change on Water Supply and Flood Hazard in Ireland Using Statistical Downscaling and Hydrological Modelling Techniques. *Climatic Change*, *74*(4),

475–491. https://doi.org/10.1007/s10584-006-0472-x

CMIP6 source_id values. (n.d.). Retrieved 10 August 2021, from https://wcrp-cmip.github.io/CMIP6_CVs/docs/CMIP6_source_id.html

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag. https://doi.org/10.1007/978-1-4471-3675-0

Copsey, D., Sutton, R., & Knight, J. R. (2006). Recent trends in sea level pressure in the Indian Ocean region. *Geophysical Research Letters*, *33*(19). https://doi.org/10.1029/2006GL027175

Deser, C., Alexander, M. A., Xie, S.-P., & Phillips, A. S. (2010). Sea Surface Temperature Variability: Patterns and Mechanisms. *Annual Review of Marine Science*, *2*(1), 115–143. https://doi.org/10.1146/annurev-marine-120408-151453

Deser, C., Simpson, I. R., McKinnon, K. A., & Phillips, A. S. (2017). The Northern Hemisphere Extratropical Atmospheric Circulation Response to ENSO: How Well Do We Know It and How Do We Evaluate Models Accordingly? *Journal of Climate*, *30*(13), 5059–5082. https://doi.org/10.1175/JCLI-D-16-0844.1

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., Charland, A., Liu, Y., Haugen, M., Tsiang, M., & Rajaratnam, B. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Sciences*, *114*(19), 4881–4886. https://doi.org/10.1073/pnas.1618082114

Dong, B., Sutton, R. T., Shaffrey, L., & Klingaman, N. P. (2017). Attribution of Forced Decadal Climate Change in Coupled and Uncoupled Ocean–Atmosphere Model Experiments. *Journal of Climate*, *30*(16), 6203–6223. https://doi.org/10.1175/JCLI-D-16-0578.1

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5),

1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fischer, E. M., Beyerle, U., Schleussner, C. F., King, A. D., & Knutti, R. (2018). Biased Estimates of Changes in Climate Extremes From Prescribed SST Simulations. *Geophysical Research Letters*, *45*(16), 8500–8509. https://doi.org/10.1029/2018GL079176

Fischer, E. M., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, 1–7. https://doi.org/10.1038/s41558-021-01092-9

Frame, D. j, Aina, T., Christensen, C. m, Faull, N. e, Knight, S. h. e, Piani, C., Rosier, S. m, Yamazaki, K., Yamazaki, Y., & Allen, M. r. (2009). The climateprediction.net BBC climate change experiment: Design of the coupled model ensemble. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1890), 855–870. https://doi.org/10.1098/rsta.2008.0240

Francis, J. A., & Vavrus, S. J. (2012). Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, *39*(6). https://doi.org/10.1029/2012GL051000

Gessner, C., Fischer, E. M., Beyerle, U., & Knutti, R. (2021). Very Rare Heat Extremes: Quantifying and Understanding Using Ensemble Reinitialization. *Journal of Climate*, *34*(16), 6619–6634. https://doi.org/10.1175/JCLI-D-20-0916.1

He, J., & Soden, B. J. (2016). Does the Lack of Coupling in SST-Forced Atmosphere-Only Models Limit Their Usefulness for Climate Change Studies? *Journal of Climate*, *29*(12), 4317–4325. https://doi.org/10.1175/JCLI-D-14-00597.1

Hodges, J. L. (1958). The significance probability of the smirnov two-sample test. *Arkiv För Matematik*, *3*(5), 469–486. https://doi.org/10.1007/BF02589501

Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(1), 105–124. https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, *29*(3), 339–349. https://doi.org/10.1080/00401706.1987.10488243

Hosking, J. R. M., & Wallis, J. R. (1997). Regional Frequency Analysis. In *Regional Frequency Analysis*. Cambridge University Press. https://doi.org/10.1017/cbo9780511529443

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, *27*(3), 251–261. https://doi.org/10.1080/00401706.1985.10488049

Huang, W. T. K., Charlton-Perez, A., Lee, R. W., Neal, R., Sarran, C., & Sun, T. (2020). Weather regimes and patterns associated with temperature-related excess mortality in the UK: A pathway to sub-seasonal risk forecasting. *Environmental Research Letters*, *15*(12), 124052. https://doi.org/10.1088/1748-9326/abcbba

Karmalkar, A. V., Sexton, D. M. H., Murphy, J. M., Booth, B. B. B., Rostron, J. W., & McNeall, D. J. (2019). Finding plausible and diverse variants of a climate model. Part II: Development and validation of methodology. *Climate Dynamics*, *53*(1), 847–877. https://doi.org/10.1007/s00382-019-04617-3

Kendon, M., Sexton, D., & McCarthy, M. (2020). A temperature of 20°C in the UK winter: A sign of the future? *Weather*, *75*(10), 318–324. https://doi.org/10.1002/wea.3811

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

King, M. P., Herceg-Bulić, I., Bladé, I., García-Serrano, J., Keenlyside, N., Kucharski, F., Li, C., & Sobolowski, S. (2018). Importance of Late Fall ENSO Teleconnection in the Euro-Atlantic Sector. *Bulletin of the American Meteorological Society*, *99*(7), 1337–1343. https://doi.org/10.1175/BAMS-D-17-0020.1

King, M. P., Yu, E., & Sillmann, J. (2020). Impact of strong and extreme El Niños on European

hydroclimate. *Tellus A: Dynamic Meteorology and Oceanography*, *72*(1), 1–10. https://doi.org/10.1080/16000870.2019.1704342

Kolmogorov, A. N. (1933). Sulla Determinazione Empirica di Una Legge di Distribuzione. *Giornale Dell'Istituto Italiano Degli Attuari*, *4*, 83–91.

Kretschmer, M., Zappa, G., & Shepherd, T. G. (2020). The role of Barents–Kara sea ice loss in projected polar vortex changes. *Weather and Climate Dynamics*, *1*(2), 715–730. https://doi.org/10.5194/wcd-1-715-2020

Kysely, J., & Picek, J. (2007). Probability estimates of heavy precipitation events in a flood-prone central-European region with enhanced influence of Mediterranean cyclones. *Advances in Geosciences*, *12*, 43–50. https://doi.org/10.5194/adgeo-12-43-2007

López-Parages, J., Rodríguez-Fonseca, B., Dommenget, D., & Frauen, C. (2016). ENSO influence on the North Atlantic European climate: A non-linear and non-stationary approach. *Climate Dynamics*, *47*(7), 2071–2084. https://doi.org/10.1007/s00382-015-2951-0

Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, T., Fosser, G., Fung, F., Gohar, L., Good, P., Gregory, J., Harris, G., Howard, T., Kaye, N., Kendon, E., Krijnen, J., … Belcher, S. (2018). *UKCP18 Science Overview Report*. Met Office Hadley Centre. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf

Lowe, J. A., Howard, T. P., Pardaens, A., Tinker, J., Holt, J., Wakelin, S., Milne, G., Leake, J., Wolf, J., Horsburgh, K., Reeder, T., Jenkins, G., Ridley, J., Dye, S., & Bradley, S. (2009). *UK Climate Projections science report: Marine and coastal projections*. Met Office Hadley Centre. https://webarchive.nationalarchives.gov.uk/ukgwa/20181204111026mp_/http://ukclimat

eprojections-ukcp09.metoffice.gov.uk/media.jsp?mediaid=87906&filetype=pdf

Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, *79*, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Murphy, J. M., Harris, G. R., Sexton, D. M. H., Kendon, E. J., Bett, P. E., Clark, R. T., Eagle, K. E., Fosser, G., Fung, F., Lowe, J. A., McDonald, R. E., McInnes, R. N., McSweeney, C. F., Mitchell, J. F. B., Rostron, J. W., Thornton, H. E., Tucker, S., & Yamazaki, K. (2018). *UKCP18 Land Projections: Science Report*. Met Office. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf

Neal, R., Fereday, D., Crocker, R., & Comer, R. E. (2016). A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, *23*(3), 389–400. https://doi.org/10.1002/met.1563

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., Lohmann, D., & Allen, M. R. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, *470*(7334), 382–385. https://doi.org/10.1038/nature09762

Pedersen, R. A., Cvijanovic, I., Langen, P. L., & Vinther, B. M. (2016). The Impact of Regional Arctic Sea Ice Loss on Atmospheric Circulation and the NAO. *Journal of Climate*, *29*(2), 889–902. https://doi.org/10.1175/JCLI-D-15-0315.1

Pope, V. D., Gallani, M. L., Rowntree, P. R., & Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model: HadAM3. *Climate Dynamics*, *16*(2), 123–146. https://doi.org/10.1007/s003820050009

Rahmstorf, S., & Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(44), 17905–17909. https://doi.org/10.1073/pnas.1101766108

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., &

Rafaj, P. (2011). RCP 8.5—A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1–2), 33–57. https://doi.org/10.1007/s10584-011-0149-y

Richardson, D., Fowler, H. J., Kilsby, C. G., & Neal, R. (2018). A new precipitation and drought climatology based on weather patterns. *International Journal of Climatology*, *38*(2), 630–648. https://doi.org/10.1002/joc.5199

Richardson, D., Neal, R., Dankers, R., Mylne, K., Cowling, R., Clements, H., & Millard, J. (2020). Linking weather patterns to regional extreme precipitation for highlighting potential flood events in medium- to long-range forecasts. *Meteorological Applications*, *27*(4), e1931. https://doi.org/10.1002/met.1931

Saravanan, R. (1998). Atmospheric Low-Frequency Variability and Its Relationship to Midlatitude SST Variability: Studies Using the NCAR Climate System Model. *Journal of Climate*, *11*(6), 1386–1404. https://doi.org/10.1175/1520-0442(1998)011<1386:ALFVAI>2.0.CO;2

Screen, J. A. (2017). The missing Northern European winter cooling response to Arctic sea ice loss. *Nature Communications*, *8*(1), 14603. https://doi.org/10.1038/ncomms14603

Screen, J. A., & Simmonds, I. (2013). Exploring links between Arctic amplification and mid-latitude weather. *Geophysical Research Letters*, *40*(5), 959–964. https://doi.org/10.1002/grl.50174

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vincente-Serrano, S. M., Wehner, M., & Zhou, B. (2021). Weather and Climate Extreme Events in a Changing Climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the*

Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_12.pdf

Sexton, D. M. H., Karmalkar, A. V., Murphy, J. M., Williams, K. D., Boutle, I. A., Morcrette, C. J., Stirling, A. J., & Vosper, S. B. (2019). Finding plausible and diverse variants of a climate model. Part 1: Establishing the relationship between errors at weather and climate time scales. *Climate Dynamics*, *53*(1), 989–1022. https://doi.org/10.1007/s00382-019-04625-3

Sexton, D. M. H., McSweeney, C. F., Rostron, J. W., Yamazaki, K., Booth, B. B. B., Murphy, J. M., Regayre, L., Johnson, J. S., & Karmalkar, A. V. (2021). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 1: selecting the parameter combinations. *Climate Dynamics*, *56*(11), 3395–3436. https://doi.org/10.1007/s00382-021-05709-9

Sexton, D. M. H., Rowell, D. P., Folland, C. K., & Karoly, D. J. (2001). Detection of anthropogenic climate change using an atmospheric GCM. *Climate Dynamics*, *17*(9), 669–685. https://doi.org/10.1007/s003820000141

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

Sippel, S., Mitchell, D., Black, M. T., Dittus, A. J., Harrington, L., Schaller, N., & Otto, F. E. L. (2015). Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Weather and Climate Extremes*, *9*, 25–35. https://doi.org/10.1016/j.wace.2015.06.004

Smirnoff, N. (1939a). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Mathématique de L´Université de*

Moscow, *2*(2), 3–11.

Smirnoff, N. (1939b). Sur les écarts de la courbe de distribution empirique. *Matematicheskii Sbornik*, *6(48)*(1), 3–26.

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

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., & Allen, M. R. (2005). Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, *433*(7024), 403–406. https://doi.org/10.1038/nature03301

Stainforth, D., Kettleborough, J., Allen, M., Collins, M., Heaps, A., & Murphy, J. (2002). Distributed computing for public-interest climate modeling research. *Computing in Science Engineering*, *4*(3), 82–89. https://doi.org/10.1109/5992.998644

Stone, D. A., & Allen, M. R. (2005). The end-to-end attribution problem: From emissions to impacts. *Climatic Change*, *71*(3), 303–318. https://doi.org/10.1007/s10584-005-6778-2

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, *432*(7017), 610–614. https://doi.org/10.1038/nature03089

The Global Risks Report 2021. (2021). World Economic Forum. https://www.weforum.org/reports/the-global-risks-report-2021/

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., van Oldenborgh, G. J., & Schaller, N. (2016). Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events. *Environmental Research Letters*, *11*(11), 114009. https://doi.org/10.1088/1748-9326/11/11/114009

Wade, S., Sanderson, M., Golding, N., Lowe, J. A., Betts, R. A., Reynard, N., Kay, A. L., Stewart, L., Prudhomme, C., Shaffrey, L., Lloyd-Hughes, B., & Harvey, B. (2015). *Developing H++*

climate change scenarios for heat waves, droughts, floods, windstorms and cold snaps (p. 144). Committee on Climate Change. https://www.theccc.org.uk/publication/met-office-for-the-asc-developing-h-climate-change-scenarios/

Wan Zin, W. Z., Jemain, A. A., & Ibrahim, K. (2009). The best fitting distribution of annual maximum rainfall in Peninsular Malaysia based on methods of L-moment and LQ-moment. *Theoretical and Applied Climatology*, *96*(3), 337–344. https://doi.org/10.1007/s00704-008-0044-2

Watson, P., Sparrow, S., Ingram, W., Wilson, S., Marie, D., Zappa, G., Jones, R., Mitchell, D., Woollings, T., & Allen, M. (2020). *Multi-thousand member ensemble atmospheric simulations with global 60km resolution using climateprediction.net* (No. EGU2020-10895). EGU2020. Copernicus Meetings. https://doi.org/10.5194/egusphere-egu2020-10895

Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dynamics*, *17*(12), 905–922. https://doi.org/10.1007/s003820100157

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., … Xavier, P. K. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380. https://doi.org/10.1002/2017MS001115

Williams, K. D., Ringer, M. A., & Senior, C. A. (2003). Evaluating the cloud response to climate change and current climate variability. *Climate Dynamics*, *20*(7), 705–721. https://doi.org/10.1007/s00382-002-0303-3

Yamazaki, K., Sexton, D. M. H., Rostron, J. W., McSweeney, C. F., Murphy, J. M., & Harris, G. R.

(2021). A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: Part 2: global performance and future changes. *Climate Dynamics*, *56*(11), 3437–3471. https://doi.org/10.1007/s00382-020-05608-5

Yiou, P., Jézéquel, A., Naveau, P., Otto, F. E. L., Vautard, R., & Vrac, M. (2017). A statistical framework for conditional extreme event attribution. *Advances in Statistical Climatology, Meteorology and Oceanography*, *3*(1), 17–31. https://doi.org/10.5194/ascmo-3-17-2017

Young, M., & Galvin, J. (2020). The record-breaking warm spell of February 2019 in Britain, the Channel Islands, France and the Netherlands. *Weather*, *75*(2), 36–45. https://doi.org/10.1002/wea.3664