

Forecast-based attribution of a winter heatwave within the limit of predictability

Nicholas J. Leach^{a,1}, Antje Weisheimer^{a,b,c}, Myles R. Allen^{a,d}, and Tim Palmer^a

^aAtmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford, OX1 3PU, UK; ^bNational Centre for Atmospheric Science, Atmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK; ^cEuropean Centre for Medium-Range Weather Forecasts, Reading, UK; ^dEnvironmental Change Institute, School of Geography and the Environment, University of Oxford, South Parks Road, Oxford, OX1 3QY, UK

This manuscript was compiled on September 3, 2021

Attribution of extreme weather events has expanded rapidly as a field over the past decade. However, deficiencies in climate model representation of key dynamical drivers of extreme events have led to some concerns over the robustness of climate model-based attribution studies. It has also been suggested that the unconditioned risk-based approach to event attribution may result in false negative results due to dynamical noise overwhelming any climate change signal. The “storyline” attribution framework, in which the impact of climate change on individual drivers of an extreme event is examined, aims to mitigate these concerns. Here we propose a novel methodology for attribution of extreme weather events using the operational ECMWF medium-range forecast model that successfully predicted the event. The use of a successful forecast ensures not only that the model is able to accurately represent the event in question; but also that the analysis is unequivocally an attribution of this specific event, rather than a mixture of multiple different events that share some characteristic. Since this attribution methodology is conditioned on the component of the event that was predictable at forecast initialisation, we show how adjusting the lead time of the forecast can flexibly set the level of conditioning desired. This flexible adjustment of the conditioning allows us to synthesise between a storyline (highly-conditioned) and a risk-based (relatively unconditioned) approach. We demonstrate this forecast-based methodology through a partial attribution of the direct radiative effect of increased CO₂ concentrations on the exceptional European winter heatwave of February 2019.

climate change | event attribution | numerical weather prediction

Attribution of extreme weather events is a relatively young field of research within climate science. However, it has expanded rapidly from its conceptual introduction (1) over the past twenty years; it now has an annual special issue in The Bulletin of the American Meteorological Society (2). Extreme event attribution is of particular importance for communicating the impacts of climate change to the public (3, 4), since the changing frequency of extreme weather events due to climate change is an impact that is physically experienced by society. As a result of this rapid expansion, there now exists a large number of different methodologies for carrying out an event attribution (5). Many of these rely on large ensembles of climate model simulations, the credibility of which has been questioned by recent studies (6–8). A particular issue is the dynamical response of the atmosphere to external forcing, which is highly uncertain within these models (9). As attribution studies try to provide quicker results, with an operational system a clear aim, it is vital that any such system provides trustworthy results. In this study we propose a “forecast-based” attribution methodology using medium-range weather forecasts which could provide several key advantages over traditional climate model-based approaches. Firstly, if an

event is predictable within a forecasting system, we know that that system is capable of accurately representing the event. Secondly, we know that any attribution performed is unequivocally an attribution of the specific event that occurred; unlike in unconditioned climate model simulations. Finally, weather forecasts are run routinely by many different national and research centers. The models used are generally state-of-the-art and extensively verified. We propose that the attribution community could and should take advantage of the massive amount of resources that are put into these forecasts by developing methodologies that use the same type of simulation. Ideally, the experiments required for attribution with forecast models would be able to be run with little additional effort on top of the routine weather forecasts; in this way they might provide a rapid operational attribution system. We discuss these ideas further throughout the text.

There have been several studies that propose or perform methodologies related to the forecast-based attribution demonstrated here. Hoerling et al. (10) used two seasonal forecast ensembles to examine the predictability of the 2011 Texas drought/heatwave within a comprehensive attribution analysis involving several different types of types of climate simulation. Meredith et al. (11) used a triply nested convection-permitting regional forecast model to investigate the role of historical SST warming within an extreme precipitation event. They conditioned their analysis on the large-scale dynamics of the event through nudging in the outermost domain. More recently, Van

Significance Statement

The question of how humans have influenced individual extreme weather events is both scientifically and socially important. However, deficiencies in climate models’ representations of key mechanisms within the process chains that drive weather reduce our confidence in estimates of the human influence on extreme events. We propose that using forecast models that successfully predicted the event in question could increase the robustness of such estimates. Using a successful forecast means we can be confident that the model is able to faithfully represent the characteristics of the specific extreme event. We use this forecast-based methodology to estimate the direct radiative impact of increased CO₂ concentrations (one component, but not the entirety, of human influence) on the European heatwave of February 2019.

TP, MRA, AW and NJL conceived the study. NJL ran the simulations, carried out the analysis and produced the figures. NJL and AW wrote the text. AW and MRA provided supervision.

We declare that we have no competing interests.

¹To whom correspondence should be addressed. E-mail: nicholas.leach@stx.ox.ac.uk

Garderen et al. (12) employed spectrally nudged simulations to assess the contribution of human influence on the climate over the 20th century on the 2003 European and 2010 Russian heatwaves. Possibly the most similar studies to the one presented here are a series of studies by Hope and colleagues (13–15). They used a seasonal forecast model to assess anthropogenic CO₂ contributions to record-breaking heat and fire weather in Australia. Two more similar studies carried out forecast-based hurricane attribution studies (16, 17). Tropical cyclones are a natural candidate for forecast-based methodologies due to the high model resolution required to represent them accurately, if at all. A final distinct, but related study is Hannart et al. (18), which proposes the use of Data Assimilation for Detection and Attribution (DADA). They suggest that operational causal attribution statements could be made in a computationally efficient manner using the kind of data assimilation procedure carried out by weather centers (to initialise forecasts) to compute the likelihood of a particular weather event under different forcings (these would be observed and estimated pre-industrial forcings for conventional attribution). Our forecast-based framework differs from these other studies in several regards. Firstly, we use a state-of-the-art forecast model to perform the attribution analysis of the event in question; rather than to solely assess the predictability of the event. We use free-running coupled global integrations here, allowing the predictable component at initialisation to dynamically condition the ensemble; as opposed to nudging our simulations towards the dynamics of the event, using nested regional simulations, or using the highly observationally constrained output of data assimilation procedures. A final key difference is that here we present an attribution of the direct radiative effect of CO₂ in isolation, though we hope that our approach could be extended in the future to provide an estimate of the full anthropogenic contribution to extreme weather events as in these other studies. We argue that the relative simplicity in the validation, setup and conditioning of our simulations is desirable from an operational attribution perspective; and flexible across many different types of extreme event.

We begin by introducing the chosen case study, the 2019 February heatwave in Europe, describing its synoptic characteristics and formally defining the event quantitatively. We then demonstrate the predictability of the event within the ECMWF ensemble prediction system, showing that this operational weather forecast was able to capture both the dynamical and thermodynamical features of the event. In [Perturbed CO₂ forecasts](#), we outline the experiments we have performed in order to quantitatively determine the direct CO₂ contribution to the heatwave. We then provide quantitative results from these experiments, and finally conclude with a discussion of the strengths and potential issues of our forecast-based attribution methodology, including our proposed directions for further work.

The 2019 February heatwave in Europe

Between the 21st and 27th February 2019, climatologically exceptional warm temperature anomalies of 10–15 °C were experienced throughout Northern and Western Europe (19), as shown in Fig. 1A. In particular, the 25th – 27th February saw record-breaking temperatures measured at many weather stations and over wide areas of Iberia, France, the British Isles, the Netherlands, Germany and Southern Sweden, as

shown in Fig. 1C (20). Fig. 1D, comparing the regional mean maximum temperatures during the 2019 heatwave with timeseries of winter mean maximum temperatures between 1950 and 2018, illustrates just how unusual and widespread the event was. This heat was associated with a characteristic flow pattern: a narrow tilted ridge extending from north-west Africa out to the southern tip of Scandinavia, advecting warm subtropical air north-east (21), as shown in the geopotential height field in Fig. 1A. This dynamical driver was accompanied by another synoptic feature that further enhanced the warming: widespread clear skies between the 25th – 27th, shown in Fig. 1B. These clear skies resulted in a widespread and persistent strong diurnal cycle, reaching 20 °C in some locations. Further details of the meteorological mechanisms and historical context of the heatwave are provided in refs. (19, 22, 23).

In order to quantify the direct impact of CO₂ on the heatwave in question within this study, we need to characterise the heatwave in an “event definition”. The choice of event definition is subjective but can impact on the quantitative results of an attribution study significantly (24–26). The most remarkable feature of the February 2019 heatwave were the maximum temperatures observed, which peaked between the 25th and 27th for the majority of the affected area. Focusing on this relatively short time-period ensures that the synoptic situation driving the heat is coherent throughout the event definition window. For the spatial extent of the event, we use the eight European sub-areas described in ref. (27). The use of regions previously defined in the literature aims to avoid selection bias. Our resulting event definition is as follows: the hottest temperature observed between 2019-02-25 and 2019-02-27, then averaged over the land points within each region (the temporal maximum is calculated before the spatial averaging). Although we carry out our calculations for all sub-areas, several regions were characteristically very similar in terms of both the event itself, and the forecasts of the event. We therefore focus on three of the eight regions: the British Isles (BI), which experienced exceptional heat and was well predicted; France (FR), which experienced exceptional heat but where the magnitude of the heat was less well forecast; and the Mediterranean (MD), which experienced well-predicted but climatologically average heat.

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

This successful forecast is a crucial part of our study as it means that we are not only confident that the model used is able to simulate the event in question; but that we are unequivocally performing an attribution analysis of the specific winter heatwave that occurred in Europe during February 2019. This is an important distinction to the framework used in “conventional” or “risk-based” (9) attribution studies (24, 29–31), which in general reduce the event to some impact-relevant

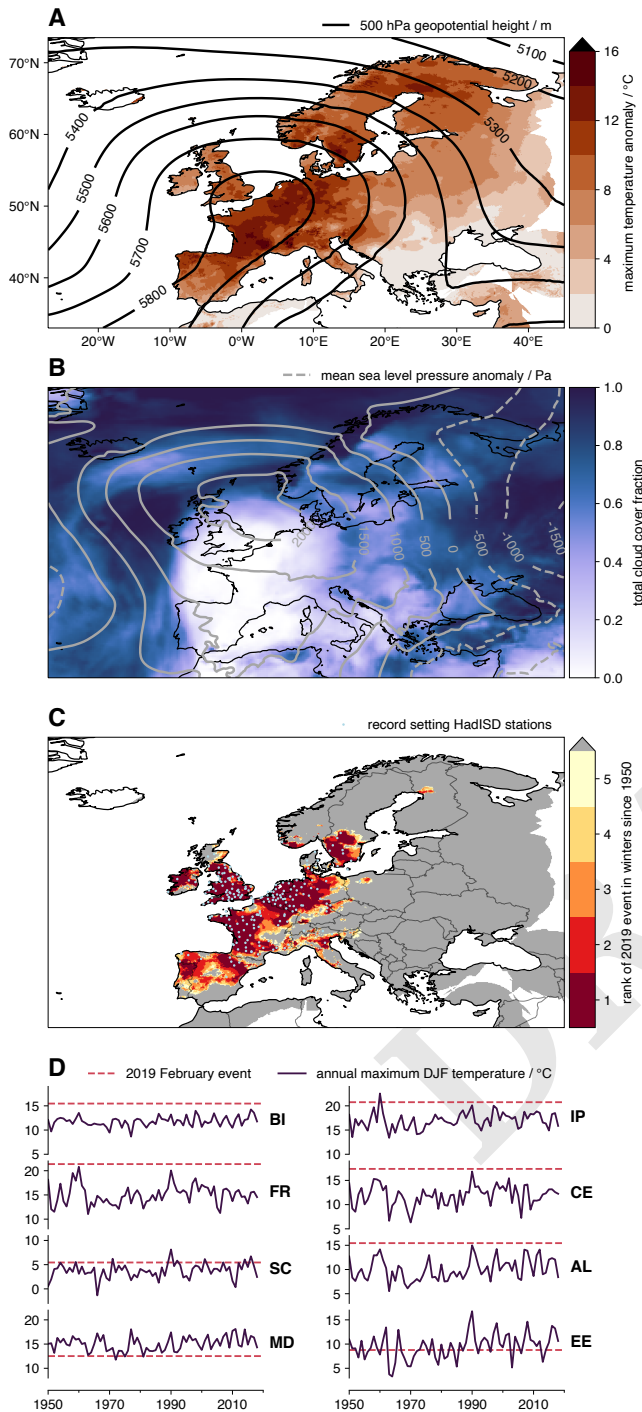


Fig. 1. Observations and historical context of the February 2019 heatwave. **A**, maximum temperature anomaly in E-OBS with overlying contours of mean Z500 anomaly from ERA5 (28) over 25–27 February 2019. **B**, mean total cloud cover with overlying contours of mean sea level pressure anomaly averaged over 25–27 February 2019. **C**, Rank of the maximum temperature in E-OBS over 25–27 February 2019 out of all winter temperature maxima since 1950 and light-blue scatterplot of 216 HadISD stations (with > 30 winters of measurements) which recorded their highest recorded value over the same three days. **D**, historical winter maximum regional mean daily maximum temperatures in E-OBS. Solid purple line shows timeseries of winter maxima for 1950–2018; dashed pink line indicates maximum value observed over the 25–27 February 2019. Regions as ref. (27).

quantitative index, then estimate the increase in likelihood of events that exceed the magnitude of the event in question. For example, a heatwave attribution study may choose to define the event as the hottest observed temperature during the heatwave, and then compute the attributable change in likelihood of temperatures hotter than this recorded maximum (eg. using models or historical records). While this does provide useful information, it does not answer the question of how much more likely anthropogenic activities have made the *specific* heatwave that occurred, rather the question of how much more likely anthropogenic activities have made a mixture of events that share one or more characteristics. Studies have attempted to provide a more satisfactory answer to this first question by including a level of conditioning on the set of events considered by using circulation analogues (32), or by nudging model simulations towards the specific dynamical situation that occurred during the event in question (11, 12). Here we are evidently performing an attribution study of the specific record-breaking heatwave that occurred in February 2019 due to the use of these successful forecasts, that not only captured the heat experienced at the surface, but also the dynamical drivers behind the heat.

As well as enabling us to answer the attribution question for a single specific heatwave, the use of a numerical weather prediction model provides additional benefits. Since large model ensembles are required to properly capture the statistics of extreme events, many previous attribution studies, especially in the context of heatwaves, have used relatively coarse, atmosphere-only climate models (33–35), which may not fully capture all the physical processes required to credibly simulate the extreme in question (36). In particular, the use of atmosphere-only simulations may result in the full space of climate variability being under-sampled due to the lack of atmosphere-ocean interaction (37). This can lead to studies overestimating the impact of anthropogenic activity on weather extremes (6, 38). More generally, Bellprat et al., and Palmer and Weisheimer (8, 39) have shown the importance of initial-value reliability in model ensembles underlying robust attribution statements. Model evaluation is therefore a key part of any robust model-based attribution study. Here, the demonstrably successful forecast enables us to be confident that the model used is providing credible realisations of the event.

A clear distinction between the typical climate model simulations used for attribution (33, 35) and the forecasts used here is that the climate model simulations are usually allowed to spin out for a sufficient length of time such that they have no memory of their initial conditions; an ensemble constructed in this way will therefore be representative of the climatology of the model. If such simulations use prescribed-SST boundary conditions, then the ensemble will be representative of the climatology conditioned on the prescribed SST pattern (34). Unlike climatological simulations, a successful forecast is conditioned upon the component of the weather that is predictable at initialisation. In general, the level of conditioning imposed upon the ensemble by the initial conditions reduces as the model integrates forwards from the initialisation date. Hence a forecast ensemble initialised only a few days before an event will be much more heavily conditioned (and therefore much less spread) than one initialised weeks before. As the lead time increases, a forecast ensemble will tend towards the

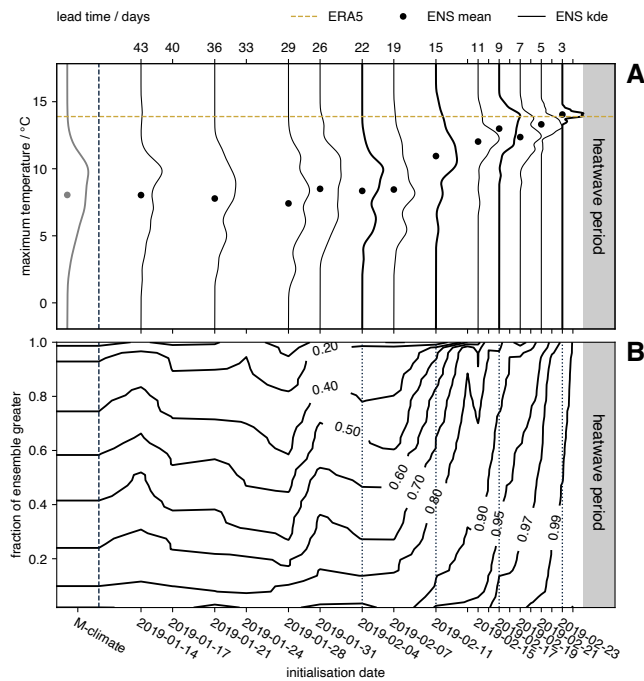


Fig. 2. Medium-range forecasts of the heatwave. **A**, ensemble distribution of heatwave as event definition against forecast initialisation date for the British Isles region. Grey line shows model climatology, bold black lines show lead times selected for the perturbed CO₂ experiments, dashed gold line shows heatwave in ERA5. Dots show ensemble mean. **B**, forecasts of Z500 over Europe during the heatwave period compared to ERA5. Y axis shows the fraction of the forecast ensemble with a pattern correlation at least as great as the levels indicated by the contour lines; against forecast initialisation date. Thin dotted lines show lead times selected for the perturbed CO₂ experiments.

model climatology, analogous to the climate model simulations discussed above. We can relate these situations to the two broad attribution frameworks discussed in (9): very long lead times, where the forecast simulates model climatology, are analogous to “conventional” attribution; while short lead times, in which the forecast ensemble is heavily dependent on the initial conditions and therefore conditional on the actual dynamical drivers that lead to the extreme event, are analogous to the “storyline” approach in (12, 40). In order to synthesize between these two frameworks, here we have chosen 4 initialisation times (3-, 9-, 15-, and 22-day leads) for our experiments that span the range from a near-unconditioned climatological forecast to a short-term forecast that is tightly conditioned on the actual dynamical drivers of the heatwave.

Perturbed CO₂ forecasts

In this study we choose to only change one feature of the operational forecast in our experiments: the CO₂ concentration. This means that the analysis we carry out is limited to attributing the impact of diabatic heating due to increased CO₂ concentrations above pre-industrial levels just over the days between the model initialisation date and the event. Although this results in a counterfactual that does not correspond to any “real” world (since it is one with approximately present-day temperatures but pre-industrial CO₂ concentrations), and thus reduces the relevance of our analysis to stakeholders or policy-makers; it does significantly increase the interpretability of our

results, and remove a major source of uncertainty associated with a “complete” attribution to human influence: the estimation of the pre-industrial ocean and sea-ice state vector used to initialise the model (41). Here we define a complete attribution as an estimate of the total impact of human influence on the climate arising from anthropogenic emissions of greenhouse gases and aerosols since the pre-industrial period. For each lead time chosen, in addition to the operational forecast (ENS) we run two experiments using operational initial conditions and identical to the operational forecast in every way except the experiments have specified fixed CO₂ concentrations. One experiment has CO₂ concentrations fixed at pre-industrial levels of 285 ppm (PI-CO₂), while in the other they are increased to 600 ppm (INCR-CO₂). These represent approximately equal and opposite perturbations on global radiative forcing (42). We carry out these two experiments for each lead time, perturbing the CO₂ concentration in opposite directions, to ensure that any changes to the likelihood of the event can be confidently attributed to the changed CO₂ concentrations. It is possible that, due to the chaotic nature of the weather, the operational conditions were ideal for generating the observed extreme, and any perturbation to the dynamical system would reduce the likelihood of its occurrence (9). If this were the case we would see a reduction in event probability regardless of whether we increased or reduced the CO₂ concentration.

Some previous work has been done on the impact of reduced CO₂ concentrations in the absence of changes to global sea surface temperatures. Baker et al. (43) explored how temperature and precipitation extremes were affected by the direct effect of CO₂ concentrations (defined there as all the effects of CO₂ on climate beside those occurring through ocean warming), finding the direct effect of CO₂ increases risk of temperature extremes, especially within the Northern hemisphere summer. Our experimental design is also reminiscent of some of the earliest work done on investigating the impact of CO₂ on climate in global circulation models (44, 45). This work found that, in the absence of changes to sea surface temperatures or sea ice concentrations, a doubling of CO₂ concentrations would change global mean surface temperatures over land by ~0.4 °C. These early studies indicate that changes in global land temperatures are approximately linear with the logarithm of CO₂ concentration.

We find that the best-estimate global mean change in land surface temperatures attributable to the additional diabatic heating due to CO₂ over pre-industrial levels (henceforth the “CO₂ signal”, calculated as half the difference between the two experiments for a particular variable) at a lead time of two weeks (over the final 5 days of the forecasts initialised on 2019-02-11) is 0.22 [0.20, 0.25] °C (square brackets indicate a 90 % confidence interval throughout). In general, the further away from the initialisation date, the slower the rate of change of the globally-averaged ensemble mean CO₂ signal, and the larger the ensemble spread (Fig. 3A). While in experiments with prescribed SSTs, we might expect the CO₂ signal in surface temperatures to approach a maximum value within timescales on the order of months, in our experiments the CO₂ signal will likely continue to increase in magnitude for centuries due to the ocean-coupling, as is the case in the abrupt-4xCO₂ experiment carried out in CMIP (46–48). The zonal-mean patterns of surface temperature CO₂ signal are qualitatively similar to those exhibited by CMIP5 and CMIP6 models during the

319 abrupt-4xCO₂ experiment (49, 50), despite the considerably
 320 shorter timescales involved: small and very confident changes
 321 in the tropics become larger but much less confident changes
 322 at the poles. This heterogeneity in the zonal distribution of
 323 warming appears to originate in the zonal distribution of the
 324 lapse-rate feedback; the weekly timescales of these experiments
 325 is insufficient for the surface-albedo feedbacks to have any
 326 significant impact (51).

327 We also examine the impact on the specific event dynamics
 328 over our region of interest; since these were crucial in develop-
 329 ing the extremes observed. Fig. 3B shows the growth in
 330 Z500 errors (measured as the mean absolute distance from
 331 ERA5 over the European domain) for each of the experiments.
 332 This figure illustrates that there are no clear differences in the
 333 ability of each experimental ensemble to predict the dynamical
 334 characteristics of the event. In other words, we have not
 335 made the synoptic event any more or less likely as a result
 336 of our perturbations. This is crucial as it means that we can
 337 consider any changes to the magnitude of the temperatures
 338 observed to be entirely due to the thermodynamic effect of
 339 changed diabatic CO₂ heating, and not due to the attractor
 340 of the dynamical system having changed as a result of the
 341 perturbations we have made.

342 Figs. 3C and 3D show analogous plots to 3B, but for inter-
 343 experimental and intra-ensemble errors respectively. These
 344 indicate a couple of important features. Firstly, no two exper-
 345 iments are more similar than any other two; the magnitude
 346 of Z500 distances in Fig. 3C are near identical for all lead
 347 and validation times. Secondly, the error growth due to the
 348 CO₂ perturbation is slower than due to the initial condition
 349 perturbations; the errors in Fig. 3C increase slower than in 3D.
 350 However, by the end of the longest lead forecast, we can see
 351 that the intra-ensemble errors have saturated, and the inter-
 352 experimental errors have grown to be the same magnitude.
 353 The saturation of intra-ensemble errors by the end of this lead
 354 time reinforces our assertion that at this lead the forecast is a
 355 good approximation of a climatological simulation.

356 Attributing the heatwave to diabatic CO₂ heating

357 First, we examine the geographical pattern of the CO₂ signal
 358 in the heatwave in Fig. 4A – D. These indicate several key
 359 features of the attributable direct CO₂ effect on the heat-
 360 wave. The CO₂ effect tends to grow with lead time, consistent
 361 with its impact on global mean temperatures. It is generally
 362 stronger over land than ocean, also consistent with global
 363 mean temperatures. Finally, the ensemble tends to become
 364 less confident in its effect as the lead time increases and the
 365 ensemble members diverge. The CO₂ signal magnitude in the
 366 heatwave generally exceeds the signal in the global mean sur-
 367 face temperature (Fig. 3A), in particular in Central Europe;
 368 possibly due to the high contribution of diabatic heating to
 369 the heatwave arising from ideal dynamical conditions. Fig.
 370 4E shows boxplots of the heatwave CO₂ signal for the three
 371 regions of interest plus the full European land area. Although
 372 there is some region-specific variability, these reinforce the
 373 main messages illustrated by the maps: the CO₂ signal grows
 374 and decreases in confidence as the lead time increases.

375 In addition to the absolute impact of the direct CO₂ effect
 376 on the heatwave, we also carry out a probabilistic assessment
 377 of its impact, consistent with conventional “risk-based” at-
 378 tribution studies (9, 52). Due to the novel approach we are

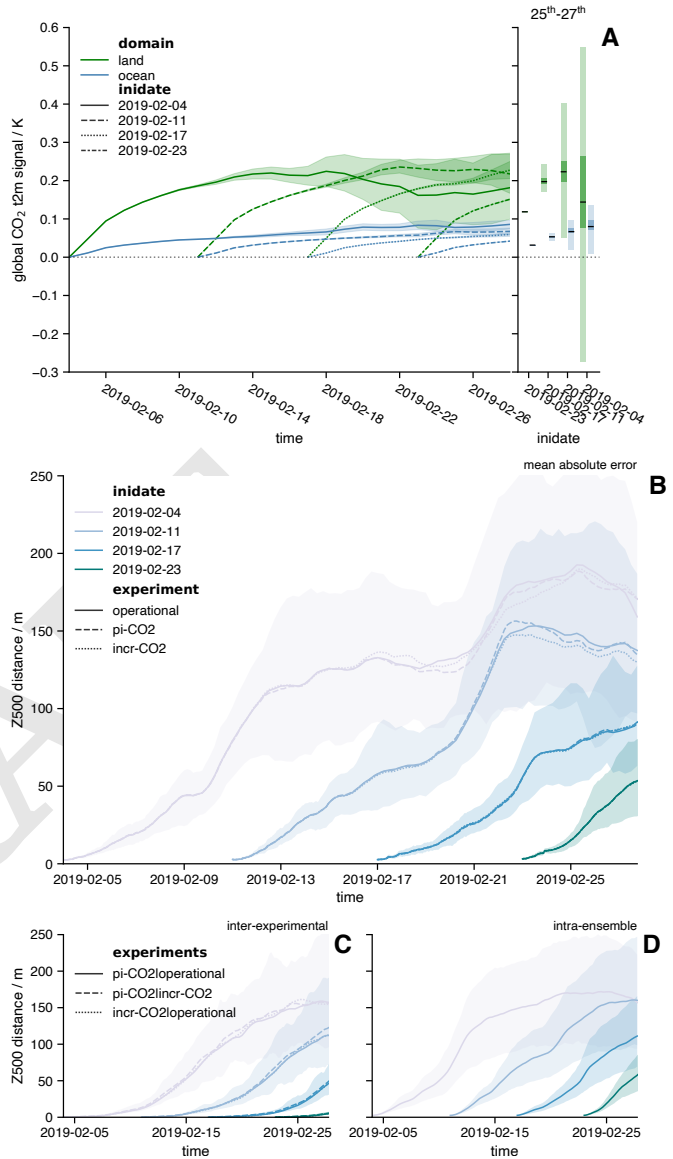


Fig. 3. Perturbed CO₂ forecast experiments. **A**, CO₂ signal in global mean surface temperatures. Green features show mean land temperatures and blue features show ocean. Line styles indicate initialisation date of the experiments. In the boxplot of the temporal mean over 2019-02-25 to 2019-02-27, the black line shows the ensemble mean, dark shading indicates 90 % confidence in the mean, and light shading indicates 90 % confidence in the ensemble. **B**, Mean absolute error in Z500 between experiments and ERA5. Color indicates initialisation date and line style indicates experiment. Solid line indicates ensemble mean. The shading shows the 5-95 % range from the operational ensemble. **C**, as **B** but for mean absolute error between corresponding ensemble members of different experiments. Line style here indicates the two differenced experiments. The shading shows the 5-95 % range of the differences between the PI-CO₂ and INCR-CO₂ experiments. **D**, as **B** but for intra-ensemble errors. Errors here are only shown for the operational ensembles.

taking within this study, it is worth clarifying exactly what question we are answering with this probabilistic analysis. The specific question is: “given the forecast initial conditions, how did the direct impact of increased CO₂ concentrations compared to pre-industrial levels just over the days between initialisation and the heatwave itself change the probability of temperatures at least as hot as were observed?”. Using conventional attribution terminology, we call the operational forecast ensemble of the event as our “factual” ensemble, and the pre-industrial CO₂ experiment as our “counter-factual” ensemble. We calculate the probability of simulating an event at least as extreme as observed in the factual ensemble, P_1 , and in the counterfactual ensemble, P_0 . These probabilities are estimated by fitting a generalised extreme value distribution to the 51-member ensemble in each case. We then express the change in event probability as a risk-ratio, $RR = P_1/P_0$, which represents the fractional increase in the likelihood of an event at least as extreme as observed in the factual ensemble over the counterfactual ensemble (29, 53). Uncertainties are estimated with a 100,000 member bootstrap with replacement, rejecting samples for which the probability of the event in the factual ensemble is zero. The resulting risk-ratios are shown in Fig. 4F. There are several key factors that contribute to the best-estimate and confidence in the risk ratios: the CO₂ signal growth with lead time; the ensemble spread growth with lead time; how extreme the event was; and how well-forecast the event was. The larger the CO₂ signal, the greater the increase in risk; the larger the ensemble spread, the lesser the increase in risk and the lower the confidence; the more extreme the event, the greater the increase in risk; and the better the forecast (ie. the closer the event to the ensemble centre), the greater the confidence.

We find that on the shortest lead time, the direct CO₂ effect increases the probability of the event over all European regions (significant at the 5 % level based on a one-sided test). For the well-forecast event experienced over the British Isles, the direct CO₂ effect increases the probability of the extreme heat by 42 [30 , 60] %. For the France heatwave, which was well-forecast given its exceptional nature, but for which the ensemble did not quite reach the total magnitude of the heat experienced, the event probability increased by at least 100 % (5th percentile), but with a very wide uncertainty range. Finally, for the least remarkable but relatively well-forecast event over the Mediterranean, the direct impact of CO₂ increased the event probability by 6.7 [4.6 , 9.7] %. These results from the very short lead experiments represent very highly conditioned statements: in both ensembles the dynamical evolution of the event was near-identical (pattern correlation of > 0.99 for all ensemble members, Fig. 2B).

Moving out to the longer lead times, we find that the confidence in the change in event probability decreases almost ubiquitously. This is as expected, since the further we move away from the event, the less highly conditioned our ensemble is, and the more dynamical noise we are adding to the system (9). However, for the 9-day lead forecast, the uncertainty is low enough to have confidence in the results for the majority of study regions. In particular, the British Isles heatwave, for which the 9-day lead forecast was better than several of the regional 3-day lead forecasts (as measured by the Continuous Ranked Probability Skill Score), increases in probability by 52 [29 , 94] % due to the direct CO₂ effect. However, for

France the uncertainty range is so large that based on these results alone we would have no confidence in the direction of the CO₂ effect. Moving further out to the 15- and 22-day lead forecasts, this loss in confidence becomes more pronounced, especially for the British Isles region. For this region, we can get virtually no useful information out of these probabilistic results for the two longest lead experiments. This drop-off in confidence arises due to the increasing ensemble spread from dynamical noise, and large reduction in the number of factual ensemble members able to simulate an event as hot as occurred in reality between the 9- and 15-day leads. A similar, though generally less pronounced drop-off in confidence is found in all other regions.

We can make use of our INCR-CO₂ experiment to increase our confidence that the positive results we obtained in the probabilistic analysis above are in fact due to the direct CO₂ effect, and not just random variability. If CO₂ were driving the changes in event probability between the PI-CO₂ and operational forecasts, then we would expect to see an even more dramatic increase in event probability between the PI-CO₂ and INCR-CO₂ forecasts. This is indeed what we find. For all regions and lead times, our best-estimate change in event probability is above zero when CO₂ concentration is increased from pre-industrial levels of 285 ppm to 600 ppm. This therefore increases our confidence further that the positive attribution to CO₂ under high conditioning is genuinely significant. From these results, it also appears that there is a general trend of change in event probability increasing as the forecast lead increases, similar to the absolute impact of the direct CO₂ effect trend; though it is still somewhat masked by uncertainty.

An important caveat on all of these results, probabilistic and absolute, is that they represent a lower bound on the estimate of the direct CO₂ effect. As is clear from the development of the CO₂ signal estimates with lead time, the model is still adjusting to the sudden change in CO₂ concentration (and would continue to do so for centuries due to the very long deep ocean equilibration timescales). Hence we would expect the “full” effect of CO₂ to be greater than the estimates we present here. This is consistent with a recent study that used unconditioned climate model simulations to carry out an attribution of the complete anthropogenic contribution to the same event, which produced much higher estimates of the risk-ratio (23).

Discussion

Here we have presented a partial, forecast-based attribution of the European 2019 winter heatwave. Taking advantage of successful medium-range forecasts from ECMWF, we used a state-of-the-art numerical weather prediction model that was demonstrably able to predict the event to attribute the direct impact of CO₂ through diabatic heating over pre-industrial levels and just over the days immediately preceding the event on the high temperatures experienced in several regions of Europe. We explored how the level of dynamical conditioning imposed can be specified by changing the lead time of the forecasts. Finally, we presented our quantitative results using two different approaches: measuring the attributable absolute and probabilistic impacts of CO₂; inspired by the “storyline” and “risk-based” attribution frameworks (9, 29, 52, 54).

There are several advantages associated with this novel

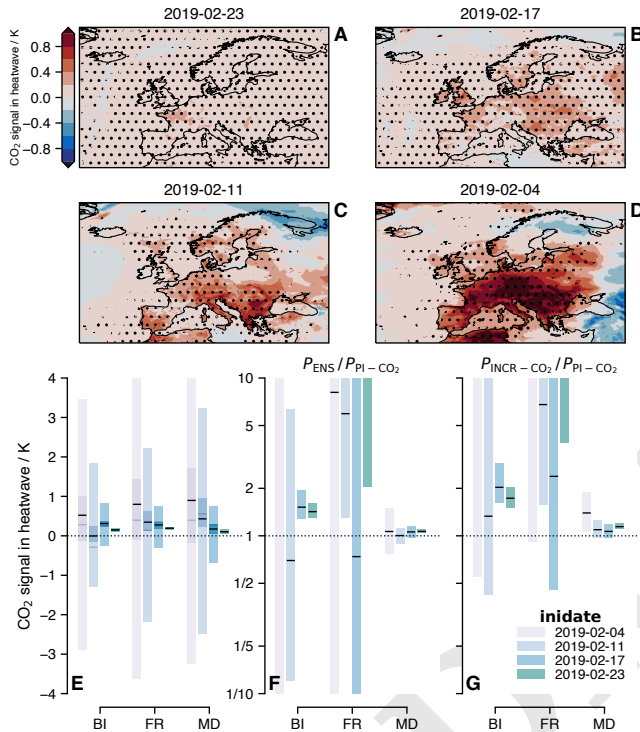


Fig. 4. Attribution of the direct CO₂ influence on the heatwave. **A–D**, maps of the ensemble mean attributable CO₂ signal in the heatwave for the four forecast lead times, which are indicated by the subplot titles. Stippling indicates a significant positive signal at the 90 % level. **E**, boxplot of the absolute CO₂ signal for the three regions of interest, and over the four forecast lead dates. Black line indicates ensemble median. Dark shading indicates 90 % confidence in the median, and light shading indicates 90 % confidence in the ensemble. Grey line indicates median difference between the operational forecast and PI-CO₂ experiment. **F**, as **E**, but showing risk ratios using the operational forecast as a factual, and PI-CO₂ experiment as a counterfactual ensemble. **G**, as **F**, but using the INCR-CO₂ experiment as a factual, and PI-CO₂ experiment as a counterfactual ensemble.

forecast-based attribution methodology, compared to conventional climate model based attribution. One simple advantage is that forecast models generally represent the technological peak within the spectrum of General Circulation Models. They tend to have a higher resolution than the models used for climate simulation. In addition, the forecast model used here is coupled, while the large climate model ensembles used for attribution tend to use prescribed sea surface temperatures (34). The use of prescribed SSTs can lead to model biases that project strongly onto attribution results (37). A final advantage arising from the use of an operational forecast model is the wealth of literature and model analysis that will already be available before an attribution study is initiated. As well as these advantages associated with the type of model there is the crucial advantage associated with using successful forecasts: the specific and intrinsic model verification. Due to the difficulty in fully quantifying how well climate models can represent an individual specific event (in particular, the very large ensembles required to have a large enough sample of characteristically similar events), climate model based attribution studies tend to perform statistical model evaluations; or/and account for this uncertainty through multi-model ensembles (55). On the other hand, if a forecast model that demonstrably predicted the event as it occurred is used, no further model verification or evaluation is required to test whether the model is capable of producing a faithful representation of the specific event.

Related to this intrinsic verification is an important point on the framing of forecast-based attribution studies. Climate model based attribution studies tend to characterise an event in terms of some quantitative index closely related to the impact of the event (such as the maximum temperature observed during a heatwave). They then use climate model simulations to determine how climate change has affected the probability of observing an event at least as extreme as the actual event. This is often done without imposing any dynamical conditioning on the simulations, though this is an area of active research (32, 56). This unconditional approach means that the specific question being answered is not “how has anthropogenic climate change affected the probability of event X?”, but “how has anthropogenic climate change affected the probability of all events that are at least as extreme as event X in terms of the index used to define X?”. This second question does not fully answer the question of how climate change has affected the actual event that the study is concerned with. In contrast, the use of a forecast model that predicted the event ensures that any attribution analysis is unequivocally an attribution of that specific event (15).

In addition to its advantages, this novel forecast-based attribution methodology also has associated issues that must be overcome. Firstly, the forecast model must have produced a “good” forecast of the event. If the model is unable to represent the event as it happened, then we cannot have confidence in any estimates of the impact of climate change on that event. Issues can arise even in qualitatively “good” forecasts, such as the forecast of the heatwave over France in this study. As very few ensemble members, if any, exceeded the observed magnitude of the event for this region, the confidence in our estimates of the probabilistic impact of CO₂ on the event is extremely low (since we are extrapolating the distribution shape outside of the range of our data). Although the estimates of the absolute

impact of CO₂ do not share this lack of confidence, this is still a problem. It is possible that applying some bias correction procedure (e.g. 57–59) based on the model climatology to the model output before analysis might alleviate these issues to some extent, but not if the model is simply unable to predict the event in question (ie. a forecast bust). Secondly, the short timescales involved in these medium-range forecasts mean that the interpretation of any results becomes more difficult as the model is still adjusting to the perturbations imposed (13), at least in the case of the CO₂ perturbations applied here. This adjustment is clear on a global scale in Fig. 3A. Due to this incomplete adjustment, any quantitative statements of attribution represent a lower bound on the “true” value.

We have shown that the direct effect of CO₂ concentrations over pre-industrial levels on the February heatwave is significant, even on timescales as short as a few days. Based on the very good 9-day lead forecast of the heatwave over the British Isles, the region that saw the most climatologically exceptional event, the direct effect of CO₂ was to increase the magnitude of the heatwave by 0.31 [0.24 , 0.37] K, and the conditional probability of the heatwave by 52 [29 , 94] %. It is very important to bear in mind that this statement of risk is highly dynamically conditioned (Fig. 2B). These estimates of the impact of CO₂ on the heatwave follow the storyline attribution framework, since we have effectively removed the dynamical uncertainty from our simulations with this strong conditioning imposed by the short lead time (9, 54, 60). Our longer, 22-day lead experiments can contrast this storyline analysis with relatively unconditioned results much closer to the climatological simulations typically used in the conventional “Risk-based” attribution framework (29, 55). At this lead, we find that although over all regions the best-estimate impact of the direct CO₂ effect is to enhance the heatwave by approximately 0.5 K, in none of the regions is this impact significantly positive at the 90 % level (based on the bootstrapped confidence in the median value). Corresponding estimates of the risk ratio have so low confidence that they provide virtually no useful information. Increasing the forecast ensemble size, which is small compared to the climate model ensembles used in most attribution studies, would increase the confidence, potentially resulting in useful quantitative estimates of the risk ratio even at these longer lead times. Our results illustrate some of the concerns voiced recently over the conventional risk-based approach to attribution (9, 52). Due to the dynamical noise present in unconditioned ensembles, it is possible to obtain an inconclusive attribution within a conventional risk-based framework, and at the same time obtain a confident positive attribution if the dynamical uncertainty is removed through conditioning (in our case achieved by reducing the forecast lead).

While this study provides a demonstration of the potential use for forecast models within attribution science, it remains a partial attribution to the direct CO₂ effect only. For forecast-based attribution to provide results that are fully comparable to conventional climate model-based attribution, we will need to demonstrate how the complete anthropogenic contribution to an extreme event could be estimated with successful forecasts. The next step to progress forecast-based attribution further will be to remove an estimate of the anthropogenic contribution to ocean temperatures from the model initial conditions (e.g. 41). If performed in addition to reducing other

greenhouse gas concentrations and aerosol climatology down to their pre-industrial levels, this should allow us to run pre-industrial forecasts of an event. This has been done previously for a seasonal forecast model by Hope et al. (13–15). They removed the anthropogenic signal from 1960 onwards from the initial conditions, but we could in principle remove the signal from pre-industrial times onwards in order to estimate the complete anthropogenic contribution to an event. Although it is highly likely that there will be methodology specific issues that arise in this direction, we suggest that being able to estimate the complete anthropogenic contribution to an extreme event using a forecast model that was able to predict the event in question would be extremely valuable. Developing a methodology to allow us to do so might also provide a pathway to operational attribution being able to be carried out by weather prediction centres, due to the routine frequency at which they produce forecasts. In addition to attempting a “complete” forecast-based attribution of an extreme event, we would like to explore how increasing the ensemble size may allow us to provide confident forecast-based attribution analyses within the unconditioned risk-based framework (ie. at long forecast lead times). One potential avenue to allow us to do this efficiently might be to reduce the resolution of the forecasts, though this would not be appropriate if it reduced the ability of the model to represent the event in question. On a similar note, we would also like to extend our experiments out to seasonal timescales. This would reduce the issues with the interpretation of our medium-range results that occurred due to the model adjustment to the sudden changes to the CO₂ concentration. It is possible that seasonal forecasts have the greatest potential to target for an operational forecast-based attribution methodology.

Materials and Methods

The ECMWF Integrated Forecasting System (IFS). In this study, we use IFS model cycle CY45R1, the operational cycle at the time of the event. The 51-member ensemble prediction system comprises a 91 layer, TCo639 resolution atmospheric model coupled to the 75 level, 0.25 degree resolution NEMO 3.4 ocean model (61). Once the model integration reaches the extended-range (day 15 onwards), the atmospheric model resolution is reduced to TCo319.

The IFS model climatology. We define the IFS model climatology, used to compute model anomalies and the CRPSS, in an identical manner as is done operationally (for example to calculate the Extreme Forecast Index product) (62). This climatology is defined using 9 consecutive re-forecast sets, spanning 5 weeks centered on the forecast initialisation date (re-forecast sets are run twice a week, every Monday and Thursday), of 11 members per re-forecast. These sets are created by initialising the re-forecast ensemble on the same calendar date over the previous 20 years. This procedure results in a model climatology of $9 \times 11 \times 20 = 1980$ members covering the 1999–2018 period. Throughout this article, we use the model climatology defined for the forecast initialised on 2019-02-11. Climatologies defined for other initialisation dates are virtually identical.

A. Statistical methods.

A.1. Significance testing. For the significance stippling displayed on the maps, we use a non-parametric (binomial) pairwise sign test at a 90 % confidence level.

A.2. Distribution fitting. When fitting statistical distributions to the ensembles during the risk-ratio calculation, we employ the method of L-moments (63), due to its numerical stability under small sample sizes.

683 Datasets used.

684 **ERA5.** We use the latest generation climate reanalysis from ECMWF,
685 ERA5 (28).

686 **E-OBS.** We use the E-OBS analysis v20.0e at 0.1 degree resolution
687 (20).

688 **HadISD.** We use HadISD v3.1.1.202003p (64–67).

689 **ACKNOWLEDGMENTS.** We thank Paul Dando, Mat Chantry,
690 and Robin Hogan for their help with setting up IFS for our
691 experiments. We acknowledge the E-OBS dataset from the EU-FP6
692 project UERRA (<http://www.uerra.eu>) and the data providers in the
693 ECA&D project (<https://www.ecad.eu>). We acknowledge the use
694 of ECMWF's computing and archiving facilities in this research
695 through the special project *spgbleac*.

697 NJL was supported by the Natural Environment Research
698 Council under grant number NE/L002612/1.

- 699 1. M Allen, Liability for climate change. *Nature* **421**, 891–892 (2003).
- 700 2. TC Peterson, PA Stott, S Herring, Explaining Extreme Events of 2011 from a Climate Per-
701 spective. *Bull. Am. Meteorol. Soc.* **93**, 1041–1067 (2012).
- 702 3. M Hulme, Attributing weather extremes to 'climate change': A review. *Prog. Phys. Geogr.* **38**,
703 499–511 (2014).
- 704 4. SJ Hassol, S Torok, S Lewis, P Luganda, (Un)Natural Disasters: Communicating Linkages
705 Between Extreme Events and Climate Change. *WMO Bull.* **65**, 2–9 (2016).
- 706 5. SC Herring, N Christidis, A Hoell, MP Hoerling, PA Stott, Explaining Extreme Events of 2019
707 from a Climate Perspective. *Bull. Am. Meteorol. Soc.* **102**, S1–S116 (2021).
- 708 6. O Bellprat, F Doblas-Reyes, Attribution of extreme weather and climate events overestimated
709 by unreliable climate simulations. *Geophys. Res. Lett.* **43**, 2158–2164 (2016).
- 710 7. O Bellprat, V Guemas, F Doblas-Reyes, MG Donat, Towards reliable extreme weather and
711 climate event attribution. *Nat. Commun.* **10**, 1732 (2019).
- 712 8. TN Palmer, A Weisheimer, A SIMPLE PEDAGOGICAL MODEL LINKING INITIAL-VALUE
713 RELIABILITY WITH TRUSTWORTHINESS IN THE FORCED CLIMATE RESPONSE. *Bull.*
714 *Am. Meteorol. Soc.* **99**, 605–614 (2018).
- 715 9. TG Shepherd, A Common Framework for Approaches to Extreme Event Attribution. *Curr*
716 *Clim. Chang. Rep.* **2**, 28–38 (2016).
- 717 10. M Hoerling, et al., Anatomy of an extreme event. *J. Clim.* **26**, 2811–2832 (2013).
- 718 11. EP Meredith, VA Semenov, D Maraun, W Park, AV Chernokulsky, Crucial role of Black Sea
719 warming in amplifying the 2012 Krymsk precipitation extreme. *Nat. Geosci.* **8**, 615–619
720 (2015).
- 721 12. L Van Garderen, F Feser, TG Shepherd, A methodology for attributing the role of climate
722 change in extreme events: A global spectrally nudged storyline. *Nat. Hazards Earth Syst.*
723 *Sci.* **21**, 171–186 (2021).
- 724 13. P Hope, EP Lim, G Wang, HH Hendon, JM Arblaster, Contributors to the Record High Tem-
725 peratures Across Australia in Late Spring 2014. *Bull. Am. Meteorol. Soc.* **96**, S149–S153
726 (2015).
- 727 14. P Hope, G Wang, EP Lim, HH Hendon, JM Arblaster, What Caused the Record-Breaking
728 Heat Across Australia in October 2015? *Bull. Am. Meteorol. Soc.* **97**, S122–S126 (2016).
- 729 15. P Hope, et al., On Determining the Impact of Increasing Atmospheric CO₂ on the Record Fire
730 Weather in Eastern Australia in February 2017. *Bull. Am. Meteorol. Soc.* **100**, S111–S117
731 (2019).
- 732 16. KA Reed, AM Stansfield, MF Wehner, CM Zarzycki, Forecasted attribution of the human
733 influence on Hurricane Florence. *Sci. Adv.* **6**, eaaw9253 (2020).
- 734 17. GM Lackmann, Hurricane Sandy before 1900 and after 2100. *Bull. Am. Meteorol. Soc.* **96**,
735 S47–S60 (2015).
- 736 18. A Hannart, et al., DADA: data assimilation for the detection and attribution of weather and
737 climate-related events. *Clim. Chang.* **136**, 155–174 (2016).
- 738 19. M Young, J Galvin, The record-breaking warm spell of February 2019 in Britain, the Channel
739 Islands, France and the Netherlands. *Weather* **75**, 36–45 (2020).
- 740 20. RC Cornes, G van der Schrier, EJM van den Besselaar, PD Jones, An Ensemble Version of
741 the E-OBS Temperature and Precipitation Data Sets. *J. Geophys. Res. Atmospheres* **123**,
742 9391–9409 (2018).
- 743 21. PM Sousa, RM Trigo, D Barriopedro, PMM Soares, JA Santos, European temperature re-
744 sponses to blocking and ridge regional patterns. *Clim. Dyn.* **50**, 457–477 (2018).
- 745 22. M Kendon, D Sexton, M McCarthy, A temperature of 20°C in the UK winter: a sign of the
746 future? *Weather* **75**, 318–324 (2020).
- 747 23. N Christidis, PA Stott, Extremely Warm Days in the United Kingdom in Winter 2018/19. *Bull.*
748 *Am. Meteorol. Soc.* **102**, S39–S44 (2021).
- 749 24. NJ Leach, et al., Anthropogenic Influence on the 2018 Summer Warm Spell in Europe: The
750 Impact of Different Spatio-Temporal Scales. *Bull. Am. Meteorol. Soc.* **101**, S41–S46 (2020).
- 751 25. P Uhe, et al., Comparison of methods: Attributing the 2014 record European temperatures to
752 human influences. *Geophys. Res. Lett.* **43**, 8685–8693 (2016).
- 753 26. MC Kirchmeier-Young, H Wan, X Zhang, SI Senéviratne, Importance of Framing for Extreme
754 Event Attribution: The Role of Spatial and Temporal Scales. *Earth's Futur.* **7**, 1192–1204
755 (2019).
- 756 27. JH Christensen, OB Christensen, A summary of the PRUDENCE model projections of
757 changes in European climate by the end of this century. *Clim. Chang.* **81**, 7–30 (2007).
- 758 28. H Hersbach, et al., The ERA5 global reanalysis. *Q. J. Royal Meteorol. Soc.* **146**, 1999–2049
759 (2020).

29. PA Stott, DA Stone, MR Allen, Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
30. P Pall, et al., Anthropogenic greenhouse gas contribution to flood risk in England and Wales
in autumn 2000. *Nature* **470**, 382–385 (2011).
31. S Sparrow, et al., Attributing human influence on the July 2017 Chinese heatwave: the influ-
ence of sea-surface temperatures. *Environ. Res. Lett.* **13**, 114004 (2018).
32. P Yiou, et al., A statistical framework for conditional extreme event attribution. *Adv. Stat.*
Climatol. Meteorol. Oceanogr. **3**, 17–31 (2017).
33. N Massey, et al., weather@home-development and validation of a very large ensemble mod-
elling system for probabilistic event attribution. *Q. J. Royal Meteorol. Soc.* **141**, 1528–1545
(2015).
34. A Ciavarella, et al., Upgrade of the HadGEM3-A based attribution system to high resolution
and a new validation framework for probabilistic event attribution. *Weather. Clim. Extrem.* **20**,
9–32 (2018).
35. N Christidis, et al., A New HadGEM3-A-Based System for Attribution of Weather- and Climate-
Related Extreme Events. *J. Clim.* **26**, 2756–2783 (2013).
36. J Sillmann, et al., Understanding, modeling and predicting weather and climate extremes:
Challenges and opportunities (2017).
37. EM Fischer, U Beyerle, CF Schleussner, AD King, R Knutti, Biased Estimates of Changes
in Climate Extremes From Prescribed SST Simulations. *Geophys. Res. Lett.* **45**, 8500–8509
(2018).
38. N Leach, et al., FaIRv2.0.0: a generalised impulse-response model for climate uncertainty
and future scenario exploration (2020).
39. O Bellprat, V Guemas, FJ Doblas-Reyes, MG Donat, Towards reliable attribution of extreme
weather and climate events. *Nat. communications* **10**, 1732 (2019).
40. W Hazeleger, et al., Tales of future weather. *NATURE CLIMATE CHANGE* **5** (2015).
41. DA Stone, P Pall, Benchmark estimate of the effect of anthropogenic emissions on the ocean
surface. *Int. J. Climatol.* **41**, 3010–3026 (2021).
42. M Etmann, G Myhre, EJ Highwood, KP Shine, Radiative forcing of carbon dioxide, methane,
and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.*
43, 12.614–12.623 (2016).
43. HS Baker, et al., Higher CO₂ concentrations increase extreme event risk in a 1.5 °C world.
Nat. Clim. Chang. **8**, 604–608 (2018).
44. WL Gates, KH Cook, ME Schlesinger, Preliminary analysis of experiments on the climatic
effects of increased CO₂ with an atmospheric general circulation model and a climatological
ocean. *J. Geophys. Res.* **86**, 6385 (1981).
45. JF Mitchell, The seasonal response of a general circulation model to changes in CO₂ and sea
surface temperatures. *Q. J. Royal Meteorol. Soc.* **109**, 113–152 (1983).
46. KE Taylor, RJ Stouffer, GA Meehl, An Overview of CMIP5 and the Experiment Design. *Bull.*
Am. Meteorol. Soc. **93**, 485–498 (2012).
47. V Eyring, et al., Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)
experimental design and organization. *Geosci. Model. Dev.* **9**, 1937–1958 (2016).
48. M Rugenstein, et al., Equilibrium Climate Sensitivity Estimated by Equilibrating Climate Mod-
els. *Geophys. Res. Lett.* **47** (2020).
49. CM Flynn, T Mauritsen, On the climate sensitivity and historical warming evolution in recent
coupled model ensembles. *Atmospheric Chem. Phys.* **20**, 7829–7842 (2020).
50. T Andrews, JM Gregory, MJ Webb, The dependence of radiative forcing and feedback on
evolving patterns of surface temperature change in climate models. *J. Clim.* **28**, 1630–1648
(2015).
51. DM Smith, et al., The Polar Amplification Model Intercomparison Project (PAMIP) contribution
to CMIP6: investigating the causes and consequences of polar amplification. *Geosci. Model.*
Dev. **12**, 1139–1164 (2019).
52. E Winsberg, N Oreskes, E Lloyd, Severe weather event attribution: Why values won't go
away. *Stud. Hist. Philos. Sci. Part A* **84**, 142–149 (2020).
53. DA Stone, MR Allen, The end-to-end attribution problem: From emissions to impacts. *Clim.*
Chang. **71**, 303–318 (2005).
54. A Jézéquel, et al., Behind the veil of extreme event attribution. *Clim. Chang.* **149**, 367–383
(2018).
55. S Philip, et al., A protocol for probabilistic extreme event attribution analyses. *Adv. Stat.*
Climatol. Meteorol. Oceanogr. **6**, 177–203 (2020).
56. P Pall, et al., Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall
in September 2013. *Weather. Clim. Extrem.* **17**, 1–6 (2017).
57. S Sippel, et al., A novel bias correction methodology for climate impact simulations. *Earth*
Syst. Dyn. **7**, 71–88 (2016).
58. S Jeon, CJ Paciorek, MF Wehner, Quantile-based bias correction and uncertainty quantifica-
tion of extreme event attribution statements. *Weather. Clim. Extrem.* **12**, 24–32 (2016).
59. S Li, et al., Reducing climate model biases by exploring parameter space with large ensem-
bles of climate model simulations and statistical emulation. *Geosci. Model. Dev.* **12**, 3017–
3043 (2019).
60. TG Shepherd, et al., Storylines: an alternative approach to representing uncertainty in physi-
cal aspects of climate change. *Clim. Chang.* **151**, 555–571 (2018).
61. ECMWF, IFS evolution - summary of cycle 45R1 (2018).
62. ECMWF, ECMWF Confluence- M-climate, the ENS Model Climate (2018).
63. JRM Hosking, L-Moments: Analysis and Estimation of Distributions Using Linear Combina-
tions of Order Statistics. *J. Royal Stat. Soc. Ser. B (Methodological)* **52**, 105–124 (1990).
64. RJ Dunn, et al., HadISD: A quality-controlled global synoptic report database for selected
variables at long-term stations from 1973–2011. *Clim. Past* **8**, 1649–1679 (2012).
65. RJ Dunn, KM Willett, CP Morice, DE Parker, Pairwise homogeneity assessment of HadISD.
Clim. Past **10**, 1501–1522 (2014).
66. RJH Dunn, KM Willett, DE Parker, L Mitchell, Expanding HadISD: quality-controlled, sub-daily
station data from 1931. *Geosci. Instrumentation, Methods Data Syst.* **5**, 473–491 (2016).
67. A Smith, N Lott, R Vose, The integrated surface database: Recent developments and part-
nerships. *Bull. Am. Meteorol. Soc.* **92**, 704–708 (2011).