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]

[a]Atmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford OX1 3PU, United Kingdom; [b]National Centre for Atmospheric Science, Atmospheric, Oceanic, and Planetary Physics, Department of Physics, University of Oxford, Oxford OX1 3PU, United Kingdom; [c]Earth System Predictability Section, European Centre for Medium-Range Weather Forecasts, Reading RG2 9AX, United Kingdom; and [d]Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford OX1 3QY, United Kingdom

Attribution of extreme weather events has expanded rapidly as a field over the past decade. However, deficiencies in climate model representation of key dynamical drivers of extreme events have led to some concerns over the robustness of climate model–based attribution studies. It has also been suggested that the unconditioned risk-based approach to event attribution may result in false negative results due to dynamical noise overwhelming any climate change signal. The "storyline" attribution framework, in which the impact of climate change on individual drivers of an extreme event is examined, aims to mitigate these concerns. Here we propose a methodology for attribution of extreme weather events using the operational European Centre for Medium-Range Weather Forecasts (ECMWF) medium-range forecast model that successfully predicted the event. The use of a successful forecast ensures not only that the model is able to accurately represent the event in question, but also that the analysis is unequivocally an attribution of this specific event, rather than a mixture of multiple different events that share some characteristic. Since this attribution methodology is conditioned on the component of the event that was predictable at forecast initialization, we show how adjusting the lead time of the forecast can flexibly set the level of conditioning desired. This flexible adjustment of the conditioning allows us to synthesize between a storyline (highly conditioned) and a risk-based (relatively unconditioned) approach. We demonstrate this forecast-based methodology through a partial attribution of the direct radiative effect of increased CO_2 concentrations on the exceptional European winter heatwave of February 2019.

climate change | extreme event attribution | numerical weather prediction

A ttribution 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 20 y; 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 exist 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 that could provide several key advantages over traditional climate model-based approaches. First, if an event is predictable within a forecasting system, we know that that system is capable of accurately representing the event. Second, 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 climate simulation. Meredith et al. (11) used a triply nested convection-permitting regional forecast model to investigate the role of historical sea surface temperature (SST) warming within an extreme precipitation event. They conditioned their analysis on the large-scale dynamics of the event through nudging in the outermost domain. More recently, Van Garderen et al. (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 et al. (13–15). They used a seasonal forecast model to assess anthropogenic CO_2 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 initialize forecasts) to compute the likelihood of a particular weather event under different forcings (these would be observed and estimated preindustrial forcings for conventional attribution). Our forecast-based framework differs from these other studies in several regards. First, we use a state-of-the-art forecast model to perform the attribution analysis of the event in question, rather than to solely assess the predictability of the event. We use free-running coupled ocean-atmosphere global integrations here, allowing the predictable component at initialization to dynamically condition the ensemble, as opposed to nudging our simulations toward the dynamics of the event, using nested regional simulations or using the highly observationally constrained output of data assimilation procedures. A final key difference is that here we present an attribution of the direct radiative effect of CO_2 in isolation, although 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 European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble prediction system, showing that this operational weather forecast was able to capture both the dynamical and thermodynamical features of the event. In perturbed CO_2 forecasts, we outline the experiments we have performed to quantitatively determine the direct CO_2 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 21 and 27 February 2019, climatologically exceptional warm temperature anomalies of 10 to 15 °C were experienced throughout Northern and Western Europe (19), as shown in Fig. 1*A*. In particular, 25 to 27 February showed 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. 1*C* (20). Fig. 1*D*, 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 titled ridge extending from northwest Africa out to the southern tip of Scandinavia, advecting warm subtropical air northeast (21), as shown in the geopotential height field in Fig. 1*A*. This dynamical driver was accompanied by another synoptic feature that further enhanced the warming: widespread clear skies between 25 and 27 February, shown in Fig. 1*B*. These clear skies resulted in a widespread and
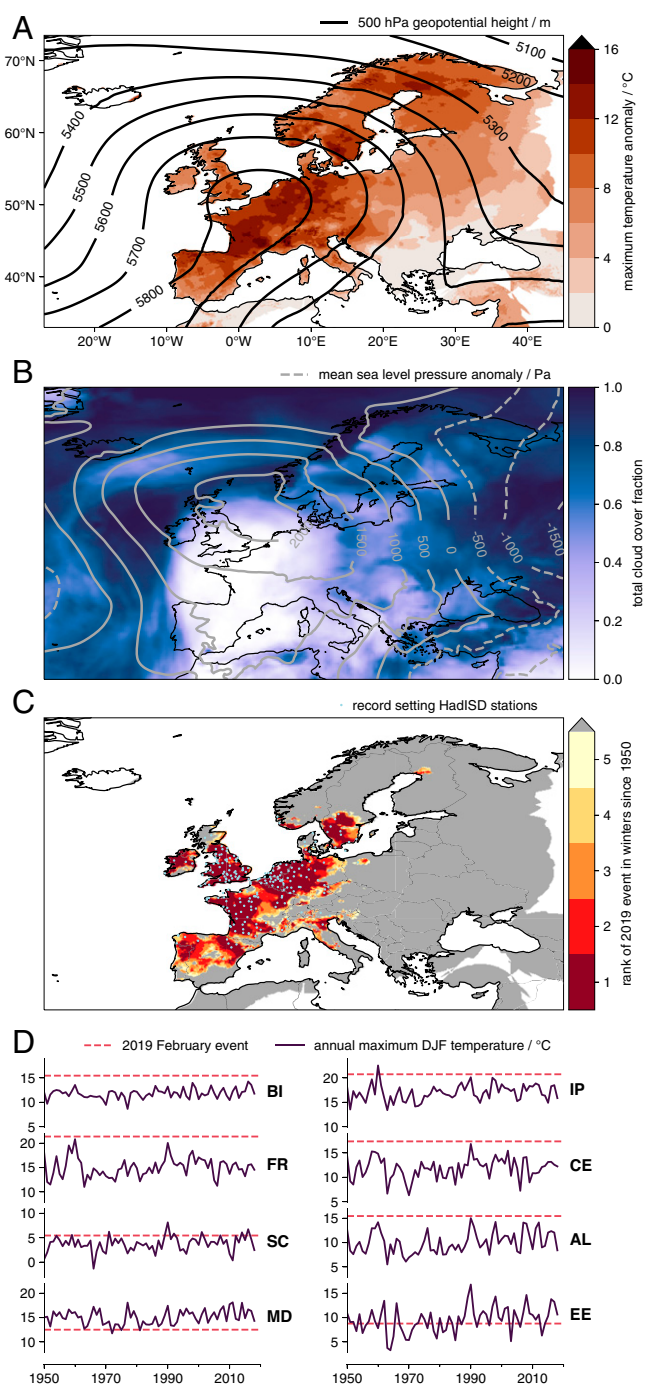
Fig. 1. Observations and historical context of the February 2019 heatwave. (*A*) Maximum temperature anomaly in E-OBS with overlying contours of mean Z500 anomaly from ERA5 (61) over 25 to 27 February 2019. (*B*) Mean total cloud cover with overlying contours of mean sea level pressure anomaly averaged over 25 to 27 February 2019. (*C*) Rank of the maximum temperature in E-OBS over 25 to 27 February 2019 out of all winter temperature maxima since 1950 and light-blue scatterplot of 216 HadISD stations (with > 30 winters of measurements) that recorded their highest recorded value over the same 3 d. (*D*) Historical winter maximum regional mean daily maximum temperatures in E-OBS. Solid purple line shows timeseries of winter maxima for 1950 to 2018; dashed pink line indicates maximum value observed over 25 to 27 February 2019. Regions are as in ref. 27.

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, and 23.

To quantify the direct impact of CO_2 on the heatwave in question within this study, we need to characterize 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 was the maximum temperatures observed, which peaked between 25 and 27 February 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 subareas 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 25 February 2019 and 27 February 2019, 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 subareas, 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 that 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 ECMWF ensemble prediction system. Their coupled ocean-atmosphere forecasts indicated "extreme" heat was possible at a lead time of around 2 wk and probable at a lead time of around 10 d (Fig. 2A), despite the exceptional nature of the heatwave in both the model climatology and the real world. As expected, the forecasts' performance in predicting the extreme heat at the surface is reflected in variables more closely linked to the dynamic drivers of the heat, such as 500 hPa geopotential height (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 also unequivocally performing an attribution analysis of the specific winter heatwave that occurred in Europe during February 2019. This is an important distinction from the framework used in "conventional" or "risk-based" (9) attribution studies (24, 28–30), which in general reduce the event to some impact-relevant quantitative index and then estimate the increase in likelihood of events that exceed the magnitude of the event in question. For example, a heatwave attribution study may choose to define the event as the hottest observed temperature during the heatwave and then compute the attributable change in likelihood of temperatures hotter than this recorded maximum (e.g., using models or historical records). While this does provide useful information, it does not answer the question of how much more likely anthropogenic activities have contributed to the specific heatwave that occurred, but rather the question of how much more likely anthropogenic activities have contributed to 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 analogs (31) or by nudging model simulations toward 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 captured not only 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
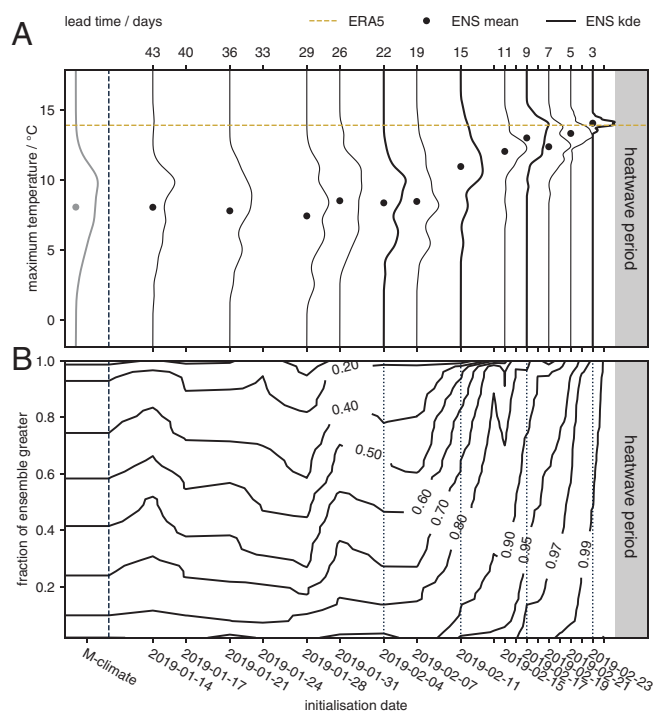
Fig. 2. Medium- to extended-range forecasts of the heatwave. (A) Ensemble distribution of heatwave as event definition against forecast initialization date for the British Isles region. Gray line on far left shows model climatology, thick black lines show lead times selected for the perturbed CO_2 experiments, dashed gold line shows heatwave magnitude in ERA5. Dots show ensemble mean. (B) Forecasts of Z500 over Europe during the heatwave period compared to ERA5. y axis shows the fraction of the forecast ensemble with a pattern correlation at least as great as the levels indicated by the contour lines, against forecast initialization date. Thin dotted lines show lead times selected for the perturbed CO_2 experiments.

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 (32–34), which may not fully capture all the physical processes required to credibly simulate the extreme in question (35). In particular, the use of atmosphere-only simulations may result in the full space of climate variability being undersampled due to the lack of atmosphere–ocean interaction (36). This can lead to studies overestimating the impact of anthropogenic activity on weather extremes (6, 24). More generally, Bellprat et al. (7) and Palmer and Weisheimer (8) 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 realizations of the event.

A clear distinction between the typical climate model simulations used for attribution (32, 34) 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 (33). Unlike climatological simulations, a successful forecast is conditioned upon the component of the weather that is predictable at initialization. In general, the level of conditioning imposed upon the ensemble by the initial conditions reduces as the model integrates forward from

Leach et al.
Forecast-based attribution of a winter heatwave within the limit of predictability

PNAS | 3 of 9
https://doi.org/10.1073/pnas.2112087118

the initialization date. Hence a forecast ensemble initialized only a few days before an event will be much more heavily conditioned (and therefore much less spread) than one initialized weeks before. As the lead time increases, a forecast ensemble will tend toward the model climatology, analogous to the climate model simulations discussed above. We can relate these situations to the two broad attribution frameworks discussed in ref. 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 refs. 12 and 37. To synthesize between these two frameworks, here we have chosen four initialization times (3-, 9-, 15-, and 22-d 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 CO$_2$ Forecasts

In this study we choose to change only one feature of the operational forecast in our experiments: the CO$_2$ concentration. This means that the analysis we carry out is limited to attributing the impact of diabatic heating due to increased CO$_2$ concentrations above preindustrial levels just over the days between the model initialization 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 preindustrial CO$_2$ concentrations) and thus reduces the relevance of our analysis to stakeholders or policymakers, it does significantly increase the interpretability of our results and removes a major source of uncertainty associated with a "complete" attribution to human influence: the estimation of the preindustrial ocean and sea-ice state vector used to initialize the model (38). 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 preindustrial period. For each lead time chosen, in addition to the operational forecast (indicated by "ENS" in the figures) we run two experiments using operational initial conditions and identical to the operational forecast in every way except the experiments have specified fixed CO$_2$ concentrations. One experiment has CO$_2$ concentrations fixed at preindustrial levels of 285 ppm (PI-CO$_2$), while in the other one they are increased to 600 ppm (INCR-CO$_2$). These represent approximately equal and opposite perturbations on global radiative forcing (39). We carry out these two experiments for each lead time, perturbing the CO$_2$ concentration in opposite directions, to ensure that any changes to the likelihood of the event can be confidently attributed to the changed CO$_2$ 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 CO$_2$ concentration.

Some previous work has been done on the impact of reduced CO$_2$ concentrations in the absence of changes to global sea surface temperatures. Baker et al. (40) explored how temperature and precipitation extremes were affected by the direct effect of CO$_2$ concentrations (defined there as all the effects of CO$_2$ on climate besides those occurring through ocean warming), finding the direct effect of CO$_2$ increases risk of temperature extremes, especially within the Northern Hemisphere summer. Our experimental design is also reminiscent of some of the earliest work done on investigating the impact of CO$_2$ on climate in global circulation models (41, 42). This work found that, in the absence of changes to sea surface temperatures

or sea ice concentrations, a doubling of CO$_2$ concentrations would change global mean surface temperatures over land by $\sim 0.4\,°$C. These early studies indicate that changes in global land temperatures are approximately linear with the logarithm of CO$_2$ concentration.
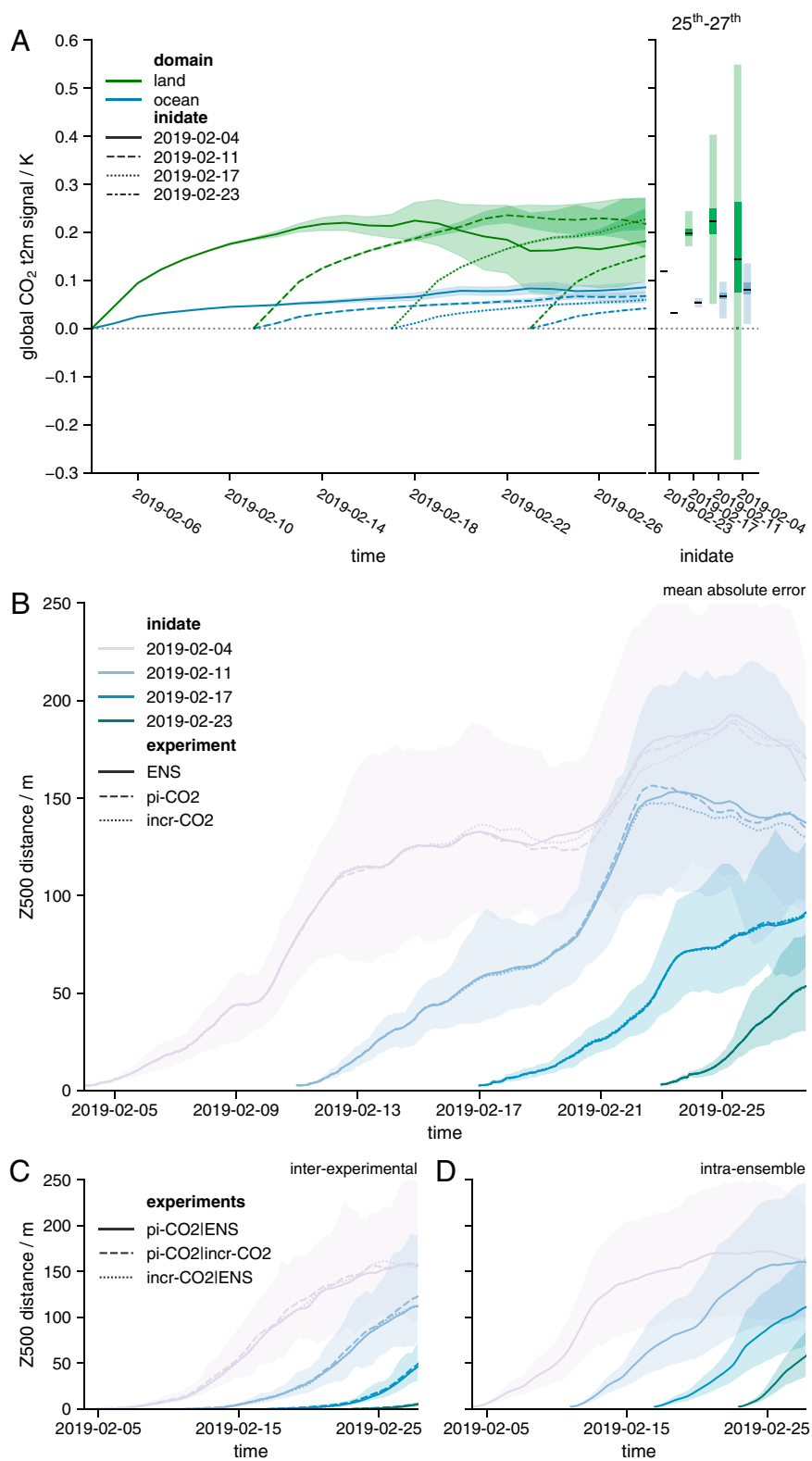
We find that the best-estimate global mean change in land surface temperatures attributable to the additional diabatic heating due to CO$_2$ over preindustrial levels (henceforth the "CO$_2$ signal," calculated as half the difference between the two experiments for a particular variable) at a lead time of 2 wk (over the final 5 d of the forecasts initialized on 11 February 2019) is 0.22 [0.20, 0.25] $°$C (square brackets indicate a 90% confidence interval throughout). In general, the farther away from the initialization date, the slower is the rate of change of the globally averaged ensemble mean CO$_2$ signal and the larger is the ensemble spread (Fig. 3A). While in experiments with prescribed SSTs, we might expect the CO$_2$ signal in surface temperatures to approach a maximum value within timescales on the order of months, in our experiments the CO$_2$ signal will likely continue to increase in magnitude for centuries due to the ocean coupling, as is the case in the abrupt-4×CO$_2$ experiment carried out in the Coupled Model Intercomparison Project (CMIP) (43–45). The zonal-mean patterns of surface temperature CO$_2$ signal are qualitatively similar to those exhibited by CMIP5 and CMIP6 models during the abrupt-4×CO$_2$ experiment (46, 47), despite the considerably shorter timescales involved: Small and very confident changes in the tropics become larger but much less confident changes at the poles. This heterogeneity in the zonal distribution of warming appears to originate in the zonal distribution of the lapse-rate feedback; the weekly timescales of these experiments are insufficient for the surface-albedo feedbacks to have any significant impact (48).

We also examine the impact on the specific event dynamics over our region of interest, since these were crucial in developing the extremes observed. Fig. 3B shows the growth in 500 hPa geopotential height (Z500) errors (measured as the mean absolute distance from ERA5 over the European domain) for each of the experiments. Fig. 3B illustrates that there are no clear differences in the ability of each experimental ensemble to predict the dynamical characteristics of the event. In other words, we have not made the synoptic event any more or less likely as a result of our perturbations. This is crucial as it means that we can consider any changes to the magnitude of the temperatures observed to be entirely due to the thermodynamic effect of changed diabatic CO$_2$ heating and not due to the attractor of the dynamical system having changed as a result of the perturbations we have made.
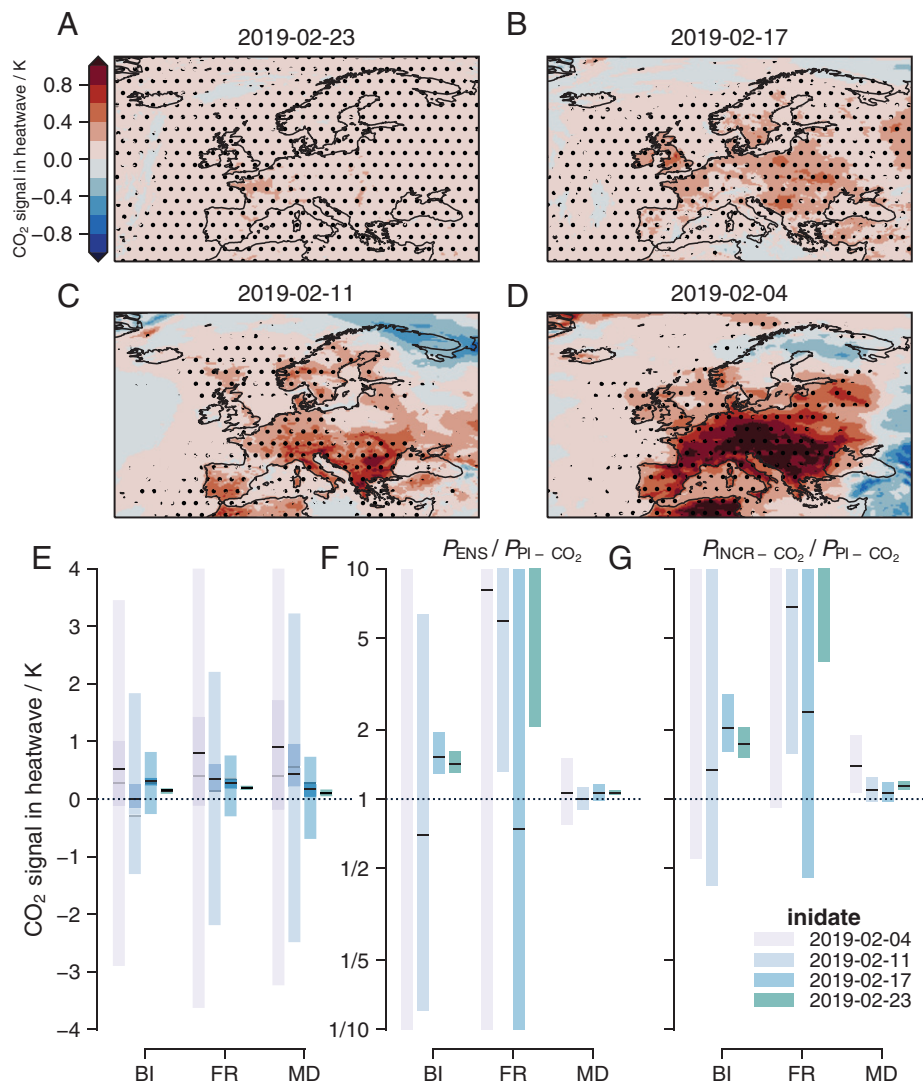
Fig. 3 C and D shows analogous plots to Fig. 3B, but for interexperimental and intraensemble errors, respectively. These indicate a couple of important features. First, no two experiments are more similar than any other two; the magnitudes of Z500 distances in Fig. 3C are nearly identical for all lead and validation times. Second, the error growth due to the CO$_2$ perturbation is slower than due to the initial condition perturbations; the errors in Fig. 3C increase slower than in Fig. 3D. However, by the end of the longest lead forecast, we can see that the intraensemble errors have saturated, and the interexperimental errors have grown to be the same magnitude. The saturation of intraensemble errors by the end of this lead time reinforces our assertion that at this lead the forecast is a good approximation of a climatological simulation.

Attributing the Heatwave to Diabatic CO$_2$ Heating

First, we examine the geographical pattern of the CO$_2$ signal in the heatwave in Fig. 4 A–D. These maps indicate several key features of the attributable direct CO$_2$ effect on the heatwave. The CO$_2$ effect tends to grow with lead time, consistent with its impact on global mean temperatures. It is generally stronger over

4 of 9 | PNAS
https://doi.org/10.1073/pnas.2112087118

Leach et al.
Forecast-based attribution of a winter heatwave within the limit of predictability

Fig. 3. Perturbed CO_2 forecast experiments. (*A*) CO_2 signal in global mean surface temperatures. Green features show mean land temperatures and blue features show ocean. Line styles indicate initialization date of the experiments. In the boxplot of the temporal mean over 25 February 2019 to 27 February 2019, 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 initialization date and line style indicates experiment. Solid line indicates ensemble mean. The shading shows the 5 to 95% range from the operational ensemble (ENS). (*C*) As in *B* but for mean absolute error between corresponding ensemble members of different experiments. Line style here indicates the two different experiments. The shading shows the 5 to 95% range of the differences between the PI-CO_2 and INCR-CO_2 experiments. (*D*) As in *B* but for intraensemble errors. Errors here are shown only for the operational ensembles.

Leach et al.
Forecast-based attribution of a winter heatwave within the limit of predictability

PNAS | 5 of 9
https://doi.org/10.1073/pnas.2112087118

Fig. 4. Attribution of the direct CO_2 influence on the heatwave. (*A–D*) Maps of the ensemble mean attributable CO_2 signal in the heatwave for the four forecast lead times, which are indicated by the subplot titles. Stippling indicates a significant positive signal at the 90% level. (*E*) Boxplot of the absolute CO_2 signal for the three regions of interest and over the four forecast lead dates. Black line indicates ensemble median. Dark shading indicates 90% confidence in the median, and light shading indicates 90% confidence in the ensemble. Gray line indicates median difference between the operational forecast and PI-CO_2 experiment. (*F*) As in *E*, but showing risk ratios using the operational forecast as a factual and PI-CO_2 experiment as a counterfactual ensemble. (*G*) As in *F*, but using the INCR-CO_2 experiment as a factual and PI-CO_2 experiment as a counterfactual ensemble.

land than ocean, also consistent with global mean temperatures. Finally, the ensemble tends to become less confident in its effect as the lead time increases and the ensemble members diverge. The CO_2 signal magnitude in the heatwave generally exceeds the signal in the global mean surface temperature (Fig. 3*A*), in particular in Central Europe, possibly due to the high contribution of diabatic heating to the heatwave arising from ideal dynamical conditions. Fig. 4*E* shows boxplots of the heatwave CO_2 signal for the three regions of interest. Although there is some region-specific variability, these reinforce the main messages illustrated by the maps: The CO_2 signal grows and decreases in confidence as the lead time increases.

In addition to the absolute impact of the direct CO_2 effect on the heatwave, we also carry out a probabilistic assessment of its impact, consistent with conventional risk-based attribution studies (9, 49). Due to the distinct approach we are taking within this study, it is worth clarifying exactly what question we are answering with this probabilistic analysis. The specific question is as follows: Given the forecast initial conditions, how did the direct impact of increased CO_2 concentrations compared to

preindustrial levels just over the days between initialization 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 our "factual" ensemble and the preindustrial CO_2 experiment our "counterfactual" 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 generalized 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 (28, 50). 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. 4*F*. There are several key factors that contribute to the best estimate and confidence in the risk ratios: the CO_2 signal growth with lead time, the ensemble spread growth with lead time, how

extreme the event was, and how well forecast the event was. The larger the CO_2 signal is, the greater the increase in risk; the larger the ensemble spread is, the lesser the increase in risk and the lower the confidence; the more extreme the event is, the greater the increase in risk; and the better the forecast is (i.e., the closer the event to the ensemble center), the greater the confidence.

We find that on the shortest lead time, the direct CO_2 effect increases the probability of the event over all European regions (significant at the 5% level based on a one-sided test). For the well-forecast event experienced over the British Isles, the direct CO_2 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% (fifth percentile), but with a very wide uncertainty range. Finally, for the least remarkable but relatively well-forecast event over the Mediterranean, the direct impact of CO_2 increased the event probability by 6.7 [4.6, 9.7]%. These results from the very short lead experiments represent very highly conditioned statements: In both ensembles the dynamical evolution of the event was nearly 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 farther 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-d 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-d lead forecast was better than several of the regional 3-d lead forecasts (as measured by the continuous ranked probability skill score), increases in probability by 52 [29, 94]% due to the direct CO_2 effect. However, for France the uncertainty range is so large that based on these results alone we would have no confidence in the direction of the CO_2 effect. Moving farther out to the 15- and 22-d 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-d leads. A similar, although generally less pronounced drop-off in confidence is found in all other regions.

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

An important caveat in all of these results, probabilistic and absolute, is that they represent a lower bound on the estimate of the direct CO_2 effect. As is clear from the development of the CO_2 signal estimates with lead time, the model is still adjusting

to the sudden change in CO_2 concentration (and would continue to do so for centuries due to the very long deep ocean equilibration timescales). Hence, we would expect the "full" effect of CO_2 to be greater than the estimates we present here. This is consistent with a recent study that used unconditioned climate model simulations to carry out an attribution of the complete anthropogenic contribution to the same event, which produced much higher estimates of the risk ratio (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 CO_2 through diabatic heating over preindustrial levels and just over the days immediately preceding the event on the high temperatures experienced in several regions of Europe. We explored how the level of dynamical conditioning imposed can be specified by changing the lead time of the forecasts. Finally, we presented our quantitative results using two different approaches: measuring the attributable absolute and probabilistic impacts of CO_2, inspired by the storyline and risk-based attribution frameworks (9, 28, 49, 51).

There are several advantages associated with this forecast-based attribution methodology, compared to conventional climate model-based attribution. One simple advantage is that forecast models generally represent the technological peak within the spectrum of general circulation models. They almost always have a higher resolution than the models used for climate simulation. In addition, the forecast model used here is coupled ocean-atmosphere, while the large climate model ensembles used for attribution often use prescribed sea surface temperatures (33). The use of prescribed SSTs can lead to model biases that project strongly onto attribution results (36). 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 multimodel ensembles (52). 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 characterize 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, although this is an area of active research (31, 53). 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 forecast-based attribution methodology also has associated issues that must be overcome. First, the forecast model must have produced a "good" forecast of the event. If the model is unable to represent the event as it happened, then we cannot have confidence in any estimates of the impact of climate change on that event. Issues can arise even in qualitatively good forecasts, such as the forecast of the heatwave over France in this study. As very few ensemble members, if any, exceeded the observed magnitude of the event for this region, the confidence in our estimates of the probabilistic impact of CO_2 on the event is extremely low (since we are extrapolating the distribution shape outside of the range of our data). Although the estimates of the absolute impact of CO_2 do not share this lack of confidence, this is still a problem. It is possible that applying some bias correction procedure (54–56) based on the model climatology to the model output before analysis might alleviate these issues to some extent, but not if the model is simply unable to predict the event in question (i.e., a forecast bust). Second, 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 CO_2 perturbations applied here. This adjustment is clear on a global scale in Fig. 3*A*. Due to this incomplete adjustment, any quantitative statements of attribution represent a lower bound on the "true" value.

We have shown that the direct effect of CO_2 concentrations over preindustrial levels on the February heatwave is significant, even on timescales as short as a few days. Based on the very good 9-d lead forecast of the heatwave over the British Isles, the region that saw the most climatologically exceptional event, the direct effect of CO_2 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. 2*B*). These estimates of the impact of CO_2 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, 51, 57). Our longer, 22-d 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 (28, 52). At this lead, we find that although over all regions the best-estimate impact of the direct CO_2 effect is to enhance the heatwave by ∼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 such 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, 49). Due to the dynamical noise present in unconditioned ensembles, it is possible to obtain an inconclusive attribution within a conventional risk-based framework and at the same time obtain a confident positive attribution if the dynamical uncertainty is removed through conditioning (in our case achieved by reducing the forecast lead).

While this study provides a demonstration of the potential use for forecast models within attribution science, it remains a partial attribution to the direct CO_2 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 farther will be to remove an estimate of the anthropogenic contribution to ocean temperatures from the model initial conditions (38). If performed in addition to reducing other greenhouse gas concentrations and aerosol climatology down to their preindustrial levels, this should allow us to run preindustrial 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 onward from the initial conditions, but we could in principle remove the signal from preindustrial times onward 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 centers, due to the routine frequency at which they produce forecasts. In addition to attempting a complete forecast-based attribution of an extreme event, we want to explore how increasing the ensemble size may allow us to provide confident forecast-based attribution analyses within the unconditioned risk-based framework (i.e., at long forecast lead times). One potential avenue to allow us to do this efficiently might be to reduce the resolution of the forecasts, although this would not be appropriate if it reduced the ability of the model to represent the event in question. On a similar note, we also want to extend our experiments out to seasonal timescales. This would reduce the issues with the interpretation of our medium-range results that occurred due to the model adjustment to the sudden changes to the CO_2 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. In this study, we use Integrated Forecasting System (IFS) model cycle CY45R1, the operational cycle at the time of the event. The 51-member ensemble prediction system comprises a 91-layer, TCo639 resolution atmospheric model coupled to the 75-level, 0.25° resolution Nucleus for European Modelling of the Ocean ocean engine v3.4 (58). Once the model integration reaches the extended range (day 15 onward), the atmospheric model resolution is reduced to TCo319.

The IFS model climatology. We define the IFS model climatology, used to compute model anomalies and the continuous ranked probability skill score, in an identical manner as is done operationally (for example, to calculate the Extreme Forecast Index product) (59). This climatology is defined using nine consecutive reforecast sets, spanning 5 wk centered on the forecast initialization date (reforecast sets are run twice a week, every Monday and Thursday), of 11 members per reforecast. These sets are created by initializing the reforecast ensemble on the same calendar date over the previous 20 y. This procedure results in a model climatology of $9 \times 11 \times 20 = 1,980$ members covering the 1999 to 2018 period. Throughout this article, we use the model climatology defined for the forecast initialized on 11 February 2019. Climatologies defined for other initialization dates are virtually identical.

Statistical Methods.

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

Distribution fitting. When fitting statistical distributions to the ensembles during the risk-ratio calculation, we employ the method of L-moments (60), due to its numerical stability under small sample sizes.

Datasets Used.

ERA5. We use the latest generation climate reanalysis from ECMWF, ERA5 (61).

E-OBS. We use the E-OBS analysis v20.0e at 0.1° resolution (20).

HadISD. We use HadISD v3.1.1.202003p (62–65).

1. M. Allen, Liability for climate change. *Nature* **421**, 891–892 (2003).
2. T. C. Peterson, P. A. Stott, S. Herring, Explaining extreme events of 2011 from a climate perspective. *Bull. Am. Meteorol. Soc.* **93**, 1041–1067 (2012).
3. M. Hulme, Attributing weather extremes to 'climate change': A review. *Prog. Phys. Geogr.* **38**, 499–511 (2014).
4. S. J. Hassol, S. Torok, S. Lewis, P. Luganda, Natural disasters: Communicating linkages between extreme events and climate change. *WMO Bull.* **65**, 2–9 (2016).
5. S. C. Herring, N. Christidis, A. Hoell, M. P. Hoerling, P. A. Stott, Explaining extreme events of 2019 from a climate perspective. *Bull. Am. Meteorol. Soc.* **102**, S1–S116 (2021).
6. O. Bellprat, F. Doblas-Reyes, Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys. Res. Lett.* **43**, 2158–2164 (2016).
7. O. Bellprat, V. Guemas, F. Doblas-Reyes, M. G. Donat, Towards reliable extreme weather and climate event attribution. *Nat. Commun.* **10**, 1–7 (2019).
8. T. N. Palmer, A. Weisheimer, A simple pedagogical model linking initial-value reliability with trustworthiness in the forced climate response. *Bull. Am. Meteorol. Soc.* **99**, 605–614 (2018).
9. T. G. Shepherd, A common framework for approaches to extreme event attribution. *Curr. Clim. Change Rep.* **2**, 28–38 (2016).
10. M. Hoerling et al., Anatomy of an extreme event. *J. Clim.* **26**, 2811–2832 (2013).
11. E. P. Meredith, V. A. Semenov, D. Maraun, W. Park, A. V. Chernokulsky, Crucial role of Black Sea warming in amplifying the 2012 Krymsk precipitation extreme. *Nat. Geosci.* **8**, 615–619 (2015).
12. L. Van Garderen, F. Feser, T. G. Shepherd, A methodology for attributing the role of climate change in extreme events: A global spectrally nudged storyline. *Nat. Hazards Earth Syst. Sci.* **21**, 171–186 (2021).
13. P. Hope, E. P. Lim, G. Wang, H. H. Hendon, J. M. Arblaster, Contributors to the record high temperatures across Australia in late spring 2014. *Bull. Am. Meteorol. Soc.* **96**, S149–S153 (2015).
14. P. Hope, G. Wang, E. P. Lim, H. H. Hendon, J. M. Arblaster, What caused the record-breaking heat across Australia in October 2015? *Bull. Am. Meteorol. Soc.* **97**, S122–S126 (2016).
15. P. Hope et al., On determining the impact of increasing atmospheric CO_2 on the record fire weather in eastern Australia in February 2017. *Bull. Am. Meteorol. Soc.* **100**, S111–S117 (2019).
16. K. A. Reed, A. M. Stansfield, M. F. Wehner, C. M. Zarzycki, Forecasted attribution of the human influence on Hurricane Florence. *Sci. Adv.* **6**, eaaw9253 (2020).
17. G. M. Lackmann, Hurricane Sandy before 1900 and after 2100. *Bull. Am. Meteorol. Soc.* **96**, 547–560 (2015).
18. A. Hannart et al., DADA: Data assimilation for the detection and attribution of weather and climate-related events. *Clim. Change* **136**, 155–174 (2016).
19. M. Young, J. Galvin, The record-breaking warm spell of February 2019 in Britain, the Channel Islands, France and the Netherlands. *Weather* **75**, 36–45 (2020).
20. R. C. Cornes, G. van der Schrier, E. J. M. van den Besselaar, P. D. Jones, An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmos.* **123**, 9391–9409 (2018).
21. P. M. Sousa, R. M. Trigo, D. Barriopedro, P. M. M. Soares, J. A. Santos, European temperature responses to blocking and ridge regional patterns. *Clim. Dyn.* **50**, 457–477 (2018).
22. M. Kendon, D. Sexton, M. McCarthy, A temperature of 20°C in the UK winter: A sign of the future? *Weather* **75**, 318–324 (2020).
23. N. Christidis, P. A. Stott, Extremely warm days in the United Kingdom in winter 2018/19. *Bull. Am. Meteorol. Soc.* **102**, S39–S44 (2021).
24. N. J. Leach et al., Anthropogenic influence on the 2018 summer warm spell in Europe: The impact of different spatio-temporal scales. *Bull. Am. Meteorol. Soc.* **101**, S41–S46 (2020).
25. P. Uhe et al., Comparison of methods: Attributing the 2014 record European temperatures to human influences. *Geophys. Res. Lett.* **43**, 8685–8693 (2016).
26. M. C. Kirchmeier-Young, H. Wan, X. Zhang, S. I. Seneviratne, Importance of framing for extreme event attribution: The role of spatial and temporal scales. *Earths Futur.* **7**, 1192–1204 (2019).
27. J. H. Christensen, O. B. Christensen, A summary of the PRUDENCE model projections of changes in European climate by the end of this century. *Clim. Change* **81**, 7–30 (2007).
28. P. A. Stott, D. A. Stone, M. R. Allen, Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
29. P. Pall et al., Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* **470**, 382–385 (2011).
30. S. Sparrow et al., Attributing human influence on the July 2017 Chinese heatwave: The influence of sea-surface temperatures. *Environ. Res. Lett.* **13**, 114004 (2018).
31. P. Yiou et al., A statistical framework for conditional extreme event attribution. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **3**, 17–31 (2017).
32. N. Massey et al., weather@home-development and validation of a very large ensemble modelling system for probabilistic event attribution. *Q. J. R. Meteorol. Soc.* **141**, 1528–1545 (2015).
33. 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).
34. N. Christidis et al., A new HadGEM3-A-based system for attribution of weather- and climate-related events. *J. Clim.* **26**, 2756–2783 (2013).
35. J. Sillmann et al., Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather Clim. Extrem.* **18**, 65–74 (2017).
36. E. M. Fischer, U. Beyerle, C. F. Schleussner, A. D. King, R. Knutti, Biased estimates of changes in climate extremes from prescribed SST simulations. *Geophys. Res. Lett.* **45**, 8500–8509 (2018).
37. W. Hazeleger et al., Tales of future weather. *Nat. Clim. Chang.* **5**, 107–113 (2015).
38. D. A. Stone, P. Pall, Benchmark estimate of the effect of anthropogenic emissions on the ocean surface. *Int. J. Climatol.* **41**, 3010–3026 (2021).
39. M. Etminan, G. Myhre, E. J. Highwood, K. P. Shine, Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophys. Res. Lett.* **43**, 12614–12623 (2016).
40. H. S. Baker et al., Higher CO_2 concentrations increase extreme event risk in a 1.5°C world. *Nat. Clim. Chang.* **8**, 604–608 (2018).
41. W. L. Gates, K. H. Cook, M. E. Schlesinger, Preliminary analysis of experiments on the climatic effects of increased CO_2 with an atmospheric general circulation model and a climatological ocean. *J. Geophys. Res.* **86**, 6385–6393 (1981).
42. J. F. Mitchell, The seasonal response of a general circulation model to changes in CO_2 and sea temperatures. *Q. J. R. Meteorol. Soc.* **109**, 113–152 (1983).
43. K. E. Taylor, R. J. Stouffer, G. A. Meehl, An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
44. 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).
45. M. Rugenstein et al., Equilibrium climate sensitivity estimated by equilibrating climate models. *Geophys. Res. Lett.* **47**, e2019GL083898 (2020).
46. C. M. Flynn, T. Mauritsen, On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmos. Chem. Phys.* **20**, 7829–7842 (2020).
47. T. Andrews, J. M. Gregory, M. J. Webb, The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *J. Clim.* **28**, 1630–1648 (2015).
48. D. M. 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).
49. E. Winsberg, N. Oreskes, E. Lloyd, Severe weather event attribution: Why values won't go away. *Stud. Hist. Philos. Sci.* **84**, 142–149 (2020).
50. D. A. Stone, M. R. Allen, The end-to-end attribution problem: From emissions to impacts. *Clim. Change* **71**, 303–318 (2005).
51. A. Jézéquel et al., Behind the veil of extreme event attribution. *Clim. Change* **149**, 367–383 (2018).
52. S. Philip et al., A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **6**, 177–203 (2020).
53. P. Pall et al., Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013. *Weather Clim. Extrem.* **17**, 1–6 (2017).
54. S. Sippel et al., A novel bias correction methodology for climate impact simulations. *Earth Syst. Dyn.* **7**, 71–88 (2016).
55. S. Jeon, C. J. Paciorek, M. F. Wehner, Quantile-based bias correction and uncertainty quantification of extreme event attribution statements. *Weather Clim. Extrem.* **12**, 24–32 (2016).
56. S. Li et al., Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation. *Geosci. Model Dev.* **12**, 3017–3043 (2019).
57. T. G. Shepherd et al., Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Clim. Change* **151**, 555–571 (2018).
58. ECMWF, Summary of cycle 45r1. https://www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/summary-cycle-45r1. Accessed 28 May 2020.
59. ECMWF, M-climate, the ENS model climate. https://confluence.ecmwf.int/display/FUG/M-climate%2C+the+ENS+Model+Climate. Accessed 28 May 2020.
60. J. R. M. Hosking, L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Stat. Soc. Ser. B Methodol.* **52**, 105–124 (1990).
61. H. Hersbach et al., The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
62. R. J. 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).
63. R. J. Dunn, K. M. Willett, C. P. Morice, D. E. Parker, Pairwise homogeneity assessment of HadISD. *Clim. Past* **10**, 1501–1522 (2014).
64. R. J. H. Dunn, K. M. Willett, D. E. Parker, L. Mitchell, Expanding HadISD: Quality-controlled, sub-daily station data from 1931. *Geosci. Instrum. Methods Data Syst.* **5**, 473–491 (2016).
65. A. Smith, N. Lott, R. Vose, The integrated surface database: Recent developments and partnerships. *Bull. Am. Meteorol. Soc.* **92**, 704–708 (2011).
66. N. J. Leach, Perturbed CO_2 forecasts of the February 2019 European heatwave. National Environment Research Council Environmental Data Service Centre for Environmental Data Analysis Archive. https://catalogue.ceda.ac.uk/uuid/dd6a312c701f47778390de50cd052071. Deposited 17 November 2021.

EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES

Leach et al.
Forecast-based attribution of a winter heatwave within the limit of predictability

PNAS | 9 of 9
https://doi.org/10.1073/pnas.2112087118