



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Circulation analogues and uncertainty in the time-evolution of extreme event probabilities: evidence from the 1947 Central European heatwave

Citation for published version:

Harrington, LJ, Otto, FEL, Cowan, T & Hegerl, GC 2019, 'Circulation analogues and uncertainty in the time-evolution of extreme event probabilities: evidence from the 1947 Central European heatwave' *Climate Dynamics*. DOI: 10.1007/s00382-019-04820-2

Digital Object Identifier (DOI):

[10.1007/s00382-019-04820-2](https://doi.org/10.1007/s00382-019-04820-2)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Climate Dynamics

Publisher Rights Statement:

© The Author(s) 2019

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Circulation analogues and uncertainty in the time-evolution of extreme event probabilities: evidence from the 1947 Central European heatwave

Luke J. Harrington¹ · Friederike E. L. Otto¹ · Tim Cowan^{2,3} · Gabriele C. Hegerl²

Received: 31 July 2018 / Accepted: 15 May 2019
© The Author(s) 2019

Abstract

The science of extreme event attribution has rapidly expanded in recent years, with numerous studies dedicated to determining whether and to what extent anthropogenic climate change has increased the likelihood of specific extreme weather events occurring. However, the majority of such studies have focussed on extreme events which have occurred in the recent past (usually within the past 10 years) while minimal research efforts have considered the multitude of high-impact extreme climatic events which occurred throughout the instrumental record. This study proposes a framework to quantify how the likelihood of witnessing meteorological characteristics reminiscent of the 1947 Central European heatwave have evolved over time. We specifically examine circulation analogues as a tool to understand the relative role of dynamical and thermodynamic contributions to changes in the probability of experiencing similar heatwave events. Using a reanalysis-based dataset, our results show changes in the frequency of 1947-like extreme heat throughout the twentieth century to be highly sensitive to methodological choices, particularly in the context of disaggregating dynamic and thermodynamic changes in the risk of extreme heat. Evidence also suggests clear decadal variability in the occurrence of circulation patterns conducive to the 1947 heatwaves. Finally, we discuss how to appropriately consider the time-evolution of attribution statements, as well as the broader limitations of employing circulation analogues as a method to interrogate the dynamical contribution to the probability ratio of an extreme event.

Keywords Circulation analogues · Heatwaves · Extreme weather events · Signal-to-noise ratios · Climate change

1 Introduction

Understanding how the frequency and severity of extreme weather events change in response to global warming is of crucial importance for many elements of future decision making (Stott and Walton 2013; Sippel et al. 2015a; Otto

et al. 2018). In the past decade, significant progress has been made in the field of extreme event attribution, in which many different techniques have been developed to quantify whether and to what extent human influences have made the occurrence of a particular extreme event—often one which has occurred in the recent past and brought with it damaging impacts—more or less likely (NAS 2016; Stott et al. 2016). Substantive progress in understanding the physical processes which govern extreme weather events, as well as the modelling and statistical tools used to characterise these extremes (Otto 2017), have enabled a prolific expansion of the number of quantitative attribution analyses (Angélil et al. 2017).

The majority of probabilistic event attribution studies can be broadly characterised as determining how much more or less likely a recently observed extreme weather event was to occur in the present day (the factual climate), relative to a counterfactual climate which excludes the presence of human influences on the climate system. Of course, variations on this framework exist, depending on

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-019-04820-2>) contains supplementary material, which is available to authorized users.

✉ Luke J. Harrington
luke.harrington@ouce.ox.ac.uk

¹ Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX1 3QY, UK

² School of Geosciences, University of Edinburgh, James Hutton Road, Edinburgh EH9 3FE, UK

³ University of Southern Queensland and The Bureau of Meteorology, 700 Collins St, Melbourne 3008, Australia

whether observational records are being analysed or the type of climate modelling framework used to interrogate these changes. However, limited research (King et al. 2016) has considered how this factual-versus-counterfactual binary framework can be expanded upon to consider the evolution of risks from events witnessed over the last century, despite the fact that many of these historical events also caused significant damages at the time. Is it the case that only events which are occurring today carry a statistically detectable signal of anthropogenic climate change, or have there been events in the past which were already more likely to occur as a result of human influences at that time? Focusing on a European heatwave from 1947 as a case study, the first aim of this study is to investigate how such an analysis could be undertaken, and particularly how the framework of analysis for considering the time-evolution of attributable increases in risk of witnessing historically significant extreme events would need to be adjusted from the approaches used for attribution analyses of present-day events.

Since European heatwaves are known to preferentially occur in the presence of specific summertime atmospheric circulation states (Cassou et al. 2005; Pfahl and Wernli 2012; Sousa et al. 2018; Jézéquel et al. 2018b), it is important to understand how dynamic and non-dynamic processes leading to heatwaves are each affected by anthropogenic climate change. Consequently, a second topic of recent research for the attribution community has concerned the extent to which changes in the likelihood of a particular extreme event occurring can be further separated into dynamic and thermodynamic contributions (Trenberth et al. 2015; Shepherd 2016; Otto et al. 2016; Harrington 2017). Among the many different methodological frameworks which have been developed to address such questions (Meredith et al. 2015; Christidis and Stott 2015; Deser et al. 2016; Lehner et al. 2017, 2018), one particular technique to decompose the relative changes to dynamical and thermodynamic influences on extreme heat and rainfall events has been the use of circulation analogue methods (Yiou et al. 2007, 2017; Cattiaux et al. 2010, 2012; Vautard et al. 2016; Stott et al. 2016; Jézéquel et al. 2018a, b, c). This approach characterises the circulation features which occur during an extreme event, then identifies the most similar-looking sequences of weather systems in other parts of the observational record (or model ensemble) and compares the variable of interest for these examples with analogous circulation properties. By quantifying how frequently circulation analogues occur in the model- or observation-based record of interest, as well as what changes occur to the likelihood of witnessing extremes within these circulation analogues, attribution statements can be feasibly made about dynamical and thermodynamic components to changes in the likelihood of occurrence for the extreme event of interest (Yiou et al. 2017).

The second aim of this study will be to examine how circulation analogue methods characterise dynamic- and thermodynamic-driven changes to the properties of extreme heatwaves over the duration of the twentieth century. Specifically, we will explore how statements regarding the frequency of circulation analogues, as well as their associated heat anomalies, can vary according to different choices made in the formulation of these analogues. Particular attention will be paid to methodological choices such as the window of time from which analogues can be sampled, the stringency of analogue selection criteria, and whether the analogues represent daily-scale circulation properties or those aggregated over the length of the event.

Given the previous demonstrations of how circulation analogues are suited to the investigation of extreme heat over Europe (Jézéquel et al. 2018b), the anomalously warm summer of 1947 over Central Europe will be used as a case study.

1.1 The Central European summer of 1947

This study focuses on the anomalously hot summer of 1947 over Central Europe as a case study to examine the time-evolution of attribution statements, as well as the efficacy of circulation analogues as a tool for probabilistic event attribution. As discussed extensively in Grütter et al. (2013, hereafter G13), the summer of 1947 was a period of pronounced heat over large regions of Central Europe, and occurred during the middle of a five-year period of drought which affected the region between 1945 and 1949 (Hirschi et al. 2013).

The period of April–September 1947 witnessed record-breaking rainfall deficits over Switzerland, which by some estimates were up to twice as severe as for the summer of 2003 (Calanca 2007). This widespread dryness helped to exacerbate the favourable meteorological conditions for extreme heat in Central Europe: in this case, such conditions were characterised by a persistent high-pressure system over Central Europe and several low pressure systems forming to the west of the British Isles contributing to significant heat advection from the south-west. The severity and length of this anomalous heat during the summer of 1947 lead to widespread forest fires in Germany, as well as a near-total loss of crops across Switzerland (Grütter et al. 2013).

While the persistent state of drought over Central Europe in the late 1940s constituted a significant hydro-meteorological extreme event in its own right, we hereafter limit our focus on the extreme heat of July and August of 1947 only—it was during this two-month period that the majority of heatwave days occurred, as defined by G13. To confirm this, Fig. 1 considers monthly mean daily maximum temperatures (TX_{mon}) for July to August 1947, relative to a 1901–2000 baseline, using gridded temperature

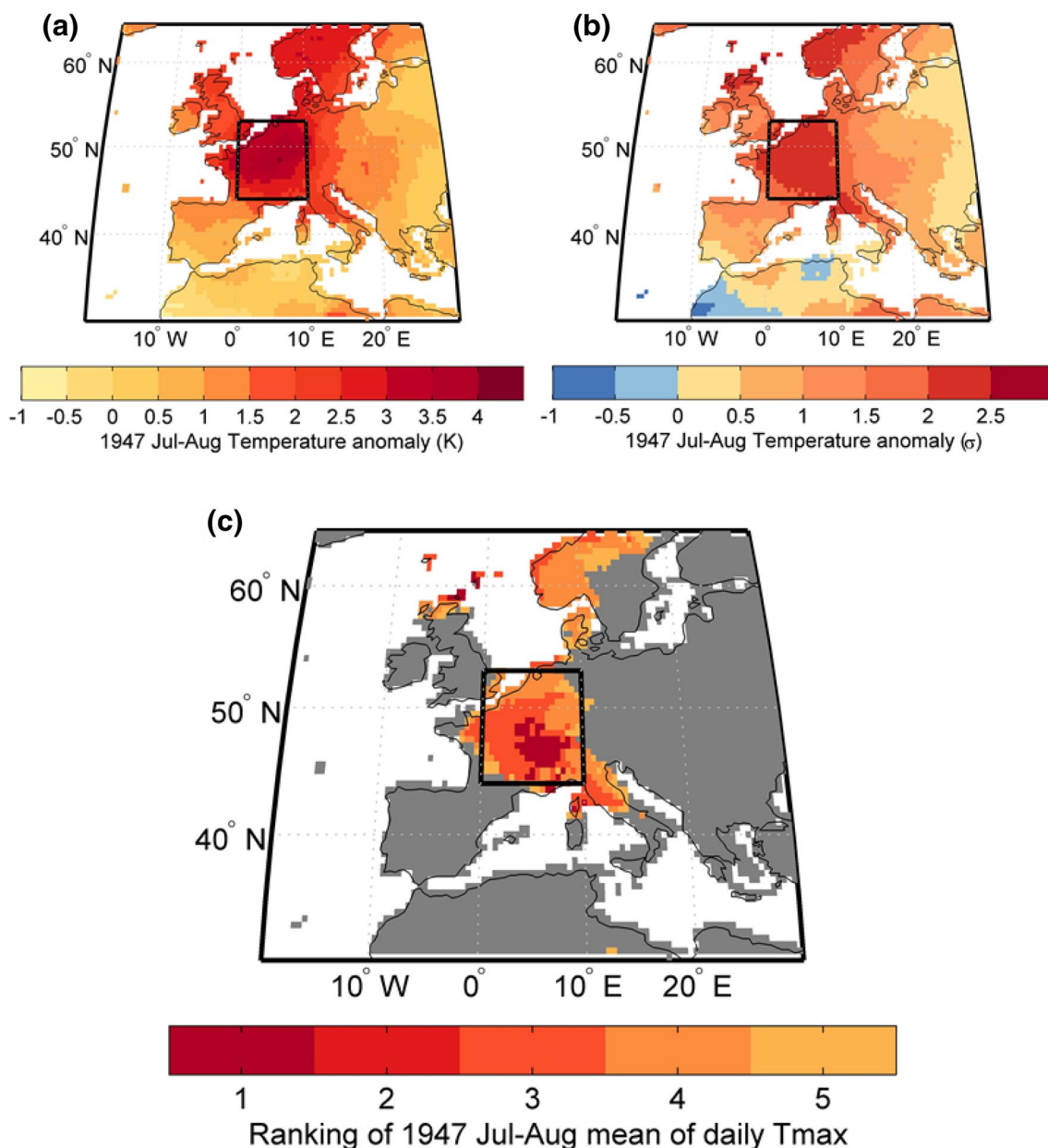


Fig. 1 Temperature anomalies for the 2 months of July–August 1947, using monthly mean daily maximum temperatures from the CRU-TS4.0 data set. **a** Absolute temperature anomaly (K) over Europe in 1947, relative to all Jul–Aug temperatures between 1901 and 2000; **b** normalised temperature anomaly by dividing the absolute anomaly at each grid cell by the standard deviation of all Jul–Aug temperatures

between 1901 and 2000. **c** Where the anomalies of 1947 rank relative to all other summers in the twentieth century, for each grid cell. The grey regions correspond to locations for which the 1947 event was outside the top 5. The black rectangle (0°E–10°E, 44°N–53°N) denotes the spatial domain of focus for all subsequent temperature-based analysis

data from the CRU-TS4.00 data set at a horizontal resolution of $0.5^\circ \times 0.5^\circ$ (Harris et al. 2014). Panels a and b respectively show absolute and relative temperature anomalies for July–August 1947, with the latter normalised by the standard deviation of TXmon over the baseline period at each grid point. Panel c shows the rank of 1947 summer heat relative to all other years in the twentieth century (Figure S1 presents an equivalent ranking map including

summers up to 2014, with similar results). Consistent with the narratives provided by previous analyses, these observations show the most severe anomalies to have occurred over a large region of Central Europe spanning parts of France, Belgium, the Netherlands, Germany and Switzerland. All subsequent temperature analysis will hereafter focus on the black rectangle outlined in Fig. 1 (0°E–10°E,

44°N–53°N) so to reflect the area which exhibited the most severe impacts from the 1947 event.

2 Data and methods

To explore the circulation and temperature characteristics which occurred during the summer of 1947 at a finer time-scale, we extract daily mean sea level pressure (MSLP) and daily maximum near-surface air temperature data from the state-of-the-art Twentieth Century Reanalysis data set Version 2c (hereafter 20CR; Compo et al. 2011). Many previous studies which consider circulation analogues in the context of heat extremes over the twentieth century have emphasised that geopotential height data at 500 hPa (Z500) is a more useful metric on which to constrain relevant circulation features for heatwaves, and reanalysis data from the National Center for Environmental Prediction (NCEP; Kalnay et al. 1996) is often preferred (Jézéquel et al. 2018b). We have chosen to instead use daily MSLP from 20CR for three reasons: first, NCEP reanalysis only extends back to 1948, while 20CR has data which extends well into the 19th Century. Second, we note that 20CR assimilates only surface pressure measurements and monthly sea surface temperatures into an atmosphere and land general circulation model (Compo et al. 2011): consequently, we have much greater confidence in the quality of the reanalysis-based MSLP data in the earlier parts of the twentieth century when compared against corresponding Z500 data (further discussion of the trade-offs between using Z500 and MSLP for analysis can be found in Sect. 4.4). And finally, G13 performed a specific analysis against station data for the summer of 1947 in Central Europe, and found 20CR to realistically represent the observed characteristics of the event (Grütter et al. 2013).

It is also worth emphasising that fewer observational records in the early part of the twentieth century fundamentally limit the reliability of any reanalysis products over this period. Therefore, while 20CR was shown to faithfully represent the key meteorological characteristics of the 1947 heatwave event (Grütter et al. 2013) as well as showing agreement with other reanalysis products in the evolution of summertime European temperatures (Figure S5 of Alvarez-Castro et al. 2018), there remains some uncertainty in the observed evolution of summertime circulation patterns, particularly prior to 1930 (Alvarez-Castro et al. 2018), which the reader should remain aware of when interpreting the results of this study.

2.1 Defining the 1947 heatwave using MSLP analogues in 20CR

Identifying when a particular extreme event actually occurred is far from straightforward (Otto 2017), and several

possible approaches exist (Christiansen 2015; Cattiaux and Ribes 2018). Discussions by G13 of both the local-scale meteorological conditions and subsequent societal impacts reveal there to be multiple periods within July and August 1947 where a heatwave could have been considered most severe, though the more significant periods of heat appear to have emerged in the final several days of July. To define the 1947 ‘heatwave’ hereafter analysed in this study, we choose to follow the methods of previous studies (Yiou et al. 2007; Cattiaux et al. 2010), and focus on when temperatures were hottest both in absolute terms, and with respect to other times in the historical record with similar circulation patterns.

For each day in the two-month period of July and August 1947, we extract the corresponding daily MSLP pattern. For each of these days, we find ten days across all other July and Augusts between 1901 and 2000 (excluding 1947) which have MSLP pattern matching most closely with the day of interest. This is found by calculating root mean square (RMS) errors over the larger European domain in Fig. 1 (spanning 20°W–30°E and 30°N–65°N, this domain size is suggested as most appropriate by Jézéquel et al. (2018b) for a comparable type of analysis). We then look at the daily maximum temperature averaged over the black rectangular box in Fig. 1 for each of these ten analogue days.

The red dashed line in Fig. 2 shows the daily Tmax data from 20CR for the summer of 1947, with the ten blue dots for each day denoting the corresponding values from each of

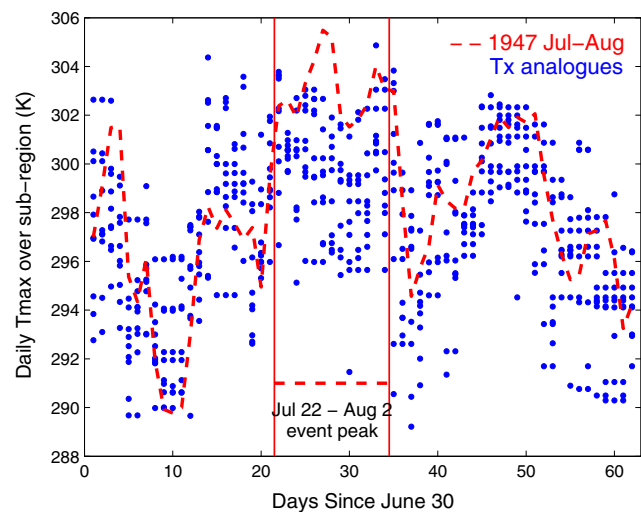


Fig. 2 Evolution of analogue temperatures for each day in July and August 1947. The red dashed line shows the daily maximum temperature evolution for the summer of 1947; the blue circles show the range of temperatures found for the top 10 circulation analogues for each individual calendar day: that is, the MSLP patterns for these analogue days most closely resemble the circulation state for the calendar day in 1947. The two vertical red lines indicate the temporal bounds hereafter chosen to define the 1947 heatwave

the ten MSLP analogues selected from anytime throughout the twentieth century. With these results, we find a clear candidate for a continuous 12-day period in the middle of the summer which we hereafter refer to as the ‘1947 heatwave’. While subjective, we have chosen the 12 days from July 22 to August 2 (inclusive), as not only are temperatures consistently near or above 302 K ($\approx 29^\circ\text{C}$), but they are also unusually high when compared with temperatures from other summer days where a similar circulation state persists.

Figure 3 presents these twelve daily MSLP patterns across the European domain. There exists clear evidence of

the quasi-persistent ridge of high pressure stretching from Portugal to Scandinavia which was emphasised by G13, as well as a continued presence of anticyclonic anomalies over Central Europe for the full twelve-day period. These patterns will also be the focus of all subsequent circulation analogue analysis for this study.

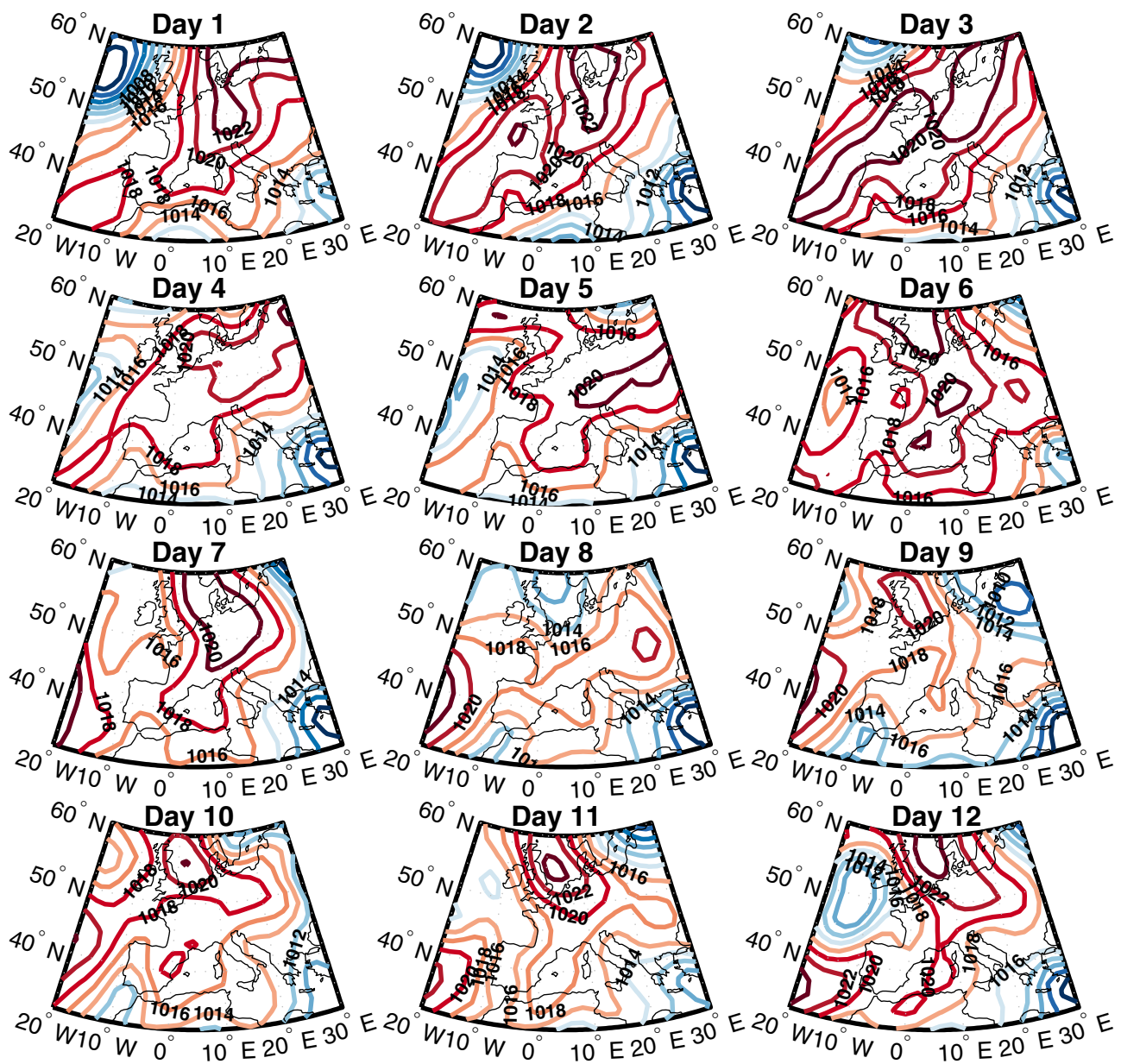


Fig. 3 Daily MSLP patterns for the 12 days of the 1947 heatwave. Day 1 refers to July 22 1947 and Day 12 refers to August 2 1947. The spatial domain presented is that which is used for all circulation analogue analyses in Sect. 3

2.2 Exploring sources of uncertainty in the time-evolution of circulation analogues and related temperature anomalies

In Sect. 3, we begin to explore the time-evolution of witnessing both circulation characteristics and temperature anomalies like that of the 1947 heatwave. Specifically, we are interested in: (1) how the probability of witnessing instances ‘like the 1947 event’ evolve through the twentieth century, using multiple different criteria; (2) whether any robust temporal trends exist in the frequency of circulation analogues being witnessed; (3) to what extent results differ when characterising analogues at a daily-scale relative to analogues of 12-day MSLP sequences; and (4) what uncertainty exists in our results from methodological choices, such as the time window from which circulation analogues are considered or the severity of the constraint as to what constitutes an ‘analogue’ circulation pattern to an observed event of interest.

3 Results

3.1 Daily-scale results

3.1.1 Evolving likelihoods of witnessing daily MSLP patterns like those of the 1947 heatwave

As a conservative estimate of witnessing circulation features similar to those which coincided with the 1947 heatwave through the twentieth century, we first assess changes to individual circulation patterns at the daily timescale—specifically, how closely they resemble each of the twelve days of the 1947 heatwave—as well as corresponding temperatures, through the twentieth century.

First, we establish baseline estimates of circulation distance (RMS error) distributions from the full 1901–2000 time series. To consider the effect of widening or narrowing the time-window of the calendar year from which we can sample days, we look at 11-day, 21-day and 31-day windows centred on each specific day of the 12 days of the 1947 heatwave. We note that a 31-day window is the maximum length of time considered, since only a 15-day window was recommended for other heatwave studies (Perkins and Alexander 2013; Perkins 2015; Perkins-Kirkpatrick and Gibson 2017). We do note that these time windows are narrower than those used by other studies of circulation analogues: as such, we also considered a 61-day window (see Supplementary Information), but found behaviour largely comparable to the 31-day-based results which are presented hereafter.

So to look at the similarity in the time-series with the MSLP pattern of July 22 1947, using an example time-window of 21 days, we consider all days from July 12 to August 1 and for all 99 years in the base period (that is, every year

from 1901 to 2000 excluding the summer of 1947 itself) to yield a distribution of $21 \times 99 = 2089$ possible days. For each of these days, we calculate the RMS error of the MSLP fields over the larger European domain presented in Fig. 1, as well as extracting the daily maximum temperatures averaged over the smaller black rectangular region of Fig. 1, corresponding to where the heat was most severe in 1947. For this distribution, we then select the value of the 10th, 20th and 30th lowest-ranked RMS errors, as arbitrary thresholds below which all days might be considered ‘analogues’ for the MSLP patterns of individual days from the 1947 event.

This process is then repeated for each of the twelve days of the 1947 event, each time extracting the RMS error corresponding to the 10th, 20th and 30th lowest-ranked values. We then aggregate the results from each individual day to yield a distribution of size $J \times Y \times 12$, where J is equal to the N -day time window considered ($J = \{11, 21, 31\}$) and $Y = 99$, which is the number of years over the base period being considered. For each of these three distributions (which differ in size according to the width of the N -day time window chosen), we extract three different RMS error values as possible criteria for ‘circulation analogues’ (as well as the probability of being below these thresholds in the base period): these correspond to the mean of all “10th-lowest ranked distances” extracted from each of the 12 daily distributions, as well as repeating this for the twelve estimates of the “20th-lowest” and “30th-lowest” distance rankings (see Fig. 6a for a schematic representation).

Now that we have a sequence of baseline thresholds to consider as criteria for a specific day being analogous to a day from the 1947 heatwave, we next consider running 20-year time periods from 1901 to 2014 (the full length of the 20CR dataset available), but excluding 1947 from consideration throughout (so the 20-year period ending 1960 will include data from 1940 to 1946 and 1948 to 1960). For each of these 20-year periods, we again obtain three $J \times Y \times 12$ distributions of daily circulation RMS errors, where J remains either 11, 21 or 31 and Y is now 20. For each of these twenty-year time periods, we look at the fraction of days, P_{BELOW} , where the RMS error sits below each of the three ‘analogue thresholds’ defined across the full base period. We then apply a 1000-member bootstrap process: this involves randomly resampling an equal number of days from each twenty-year period (each time sampling with replacement, which ensures all data points always have an equal chance of selection), re-calculating P_{BELOW} , and then repeating these steps 1000 times to yield uncertainty bounds for the P_{BELOW} term. In addition, we extract all temperatures from those days which lie below each of the specified RMS error thresholds—these will be considered hereafter as ‘analogue temperatures’.

To understand the time-evolution of this P_{BELOW} term, we present results for each of the three arbitrary analogue

thresholds chosen, for each of the distributions associated with a different N-day time window—shown separately in Fig. 4a–c for clarity. For each of the three panels (corresponding to a different N-day time window) the blue, red and purple solid lines denote the evolution of P_{BELOW} across the 20-year running periods from 1901 to 2014 (thin lines indicating the 90% bootstrapped confidence interval). The dashed line corresponds to the default probability obtained across the 1901–2000 baseline period.

To more effectively understand the changes in these probabilities of analogue circulation days occurring relative to the base period probability, we simply divide the former by the latter in each 20-year period to yield time-evolving probability ratios (PRs), and present these corresponding results in Fig. 4d–f. Similar to the frameworks that are considered for probabilistic event attribution more generally, we can interpret a PR greater (less) than one as indicating that a specific twenty-year period was more (less) likely to experience

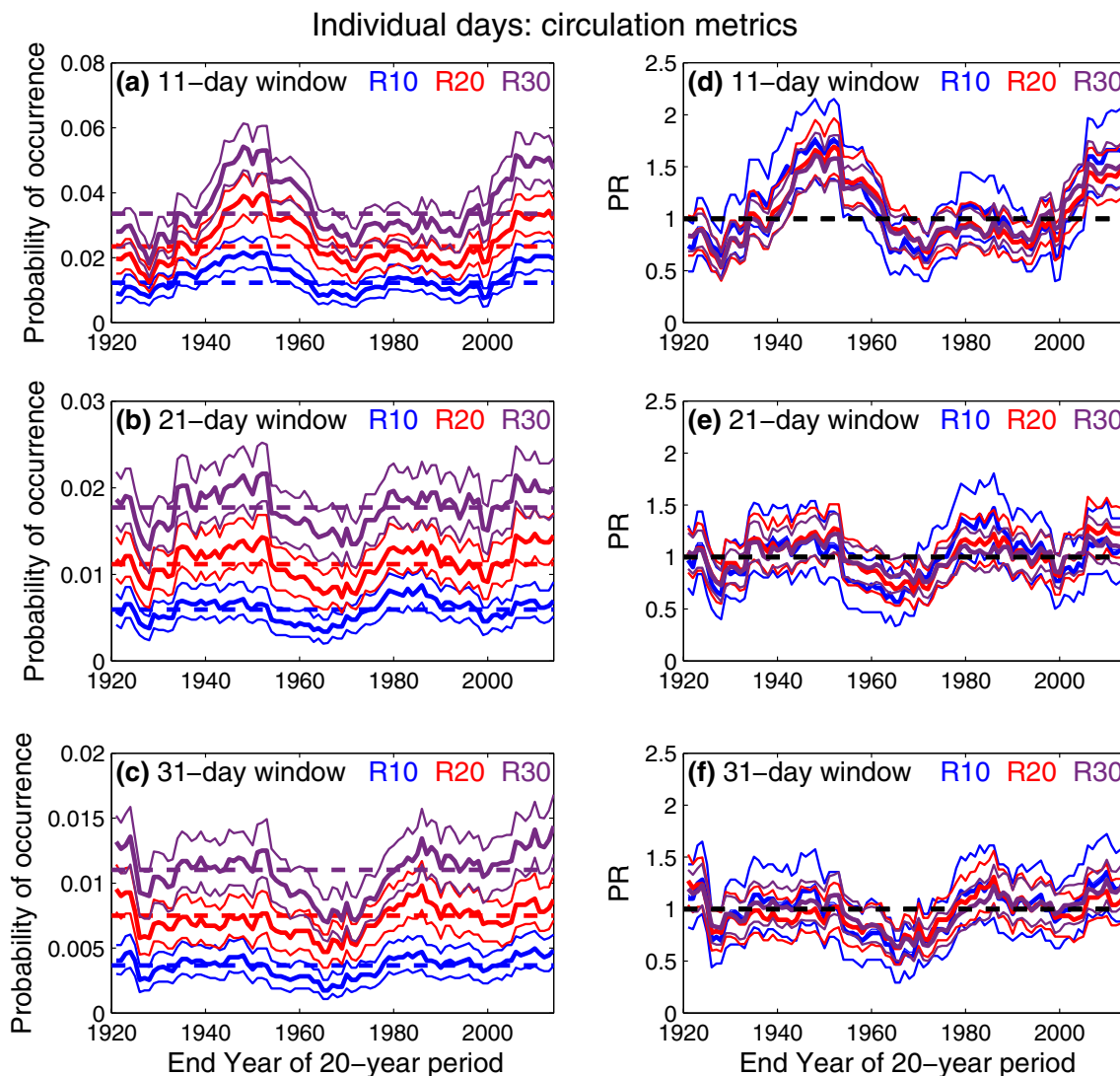


Fig. 4 Twentieth Century evolution of the probability of witnessing summer days with MSLP patterns comparable to any of the twelve individual days from the 1947 Central Europe heatwave. **a–c** Show the absolute probability of a random day in the period of interest exhibiting an MSLP pattern ‘analogous’ to one of the 12 days from the 1947 Central Europe heatwave. Dashed horizontal lines denote the fraction of days which are considered MSLP analogues over the full 1901–2000 base period (1947 excluded); different colours denote the use of three arbitrary thresholds for defining a circulation analogue (R10 refers to 10th-lowest ranked RMS error score, and so

on). The thick solid lines show the evolution of circulation analogue frequency over running 20-year periods through the time-series; thin solid lines represent the corresponding 5th–95th confidence interval using a 1000-member bootstrap resampling approach. **d–f** Show the same information as for panels **a–c**, but normalising the time-evolving probability over the running 20-year periods by the baseline probability estimates, thereby producing ‘Probability Ratios’ which are comparable across all event thresholds and across all the three distributions using different N-day time windows

daily MSLP patterns which could be characterised as ‘circulation analogues’, relative to the full baseline period.

Three important results emerge from Fig. 4d–f. Within each of the individual N -day window distributions, the probability ratio trajectories are nearly-identical between the three different circulation thresholds chosen—though the uncertainty bounds in PR are notably larger for the most extreme analogue constraint (the 10th-lowest ranked distance threshold). Second, there is no clear positive trend through time in the likelihood of witnessing daily circulation patterns like those of the 12 days of the 1947 extreme heatwave. Third, there are some individual decades—like the 1940s in Fig. 4d—which can show statistically significant probabilities of witnessing 1947-like MSLP patterns. But there are some notable variations depending on the length of the time window, and no substantive deviations from $PR = 1$ exist which are robust across all N -day time windows can be found in these results.

3.1.2 Time-evolution of daily analogue temperatures

Next we consider temperature changes during those days which satisfy the MSLP-based criteria for being a circulation analogue, and see how these changes compare against the corresponding evolution of summertime temperatures more generally over the same region.

To first consider the broader evolution of summertime temperatures, we aggregate all daily maximum temperatures over the period July 22–August 2 (consistent with the period of the 1947 heatwave) for running 20-year periods through the twentieth century, and extract the mean temperature from this aggregated data set. This mean temperature evolution is presented as a black line in Fig. 5a–c. We then select all days which satisfy the criteria for being ‘circulation analogues’ from the previous section, and consider the mean temperature from all of these analogue days for each of the different distributions which align with different N -day temperature windows considered. The time-evolution of these average analogue temperatures are also presented in Fig. 5a–c, with different coloured lines representing the three RMSE analogue thresholds of varying stringency that were considered, and the different N -day time windows again presented in separate panels.

For the case of general summertime temperatures, a clear warming signal is found through the twentieth century, with increases of more than $2\text{ }^{\circ}\text{C}$ for the twenty-year period ending in 2014 relative to the start of the 1900s. We also find for any given time period, the average of the analogue temperatures to be approximately $2\text{ }^{\circ}\text{C}$ warmer than the corresponding mean of all daily temperatures (black line in Fig. 5a–c), an inflation consistent with previous studies (Jézéquel et al. 2018b). This is clear evidence to support the notion that extreme summertime temperatures in 1947 were more likely

to occur in the presence of the corresponding MSLP patterns which were also present. Such a result which is entirely consistent with the types of circulation features associated with Central European summertime heatwaves (Fischer and Schar 2010). It is also further evidence to support the concept of using circulation analogues to interrogate the dynamical contributions to the occurrence of extreme weather events.

Interestingly however, we find the trends in average daily analogue temperatures through the twentieth century to mostly mirror the corresponding changes found for the distribution which includes all summer days. To interrogate this result more thoroughly, we present in Fig. 5d–f the mean analogue temperature as a percentile of the full distribution of daily summer temperatures. Results show a small decreasing trend through time, with variations between the 60th and 90th percentile of the corresponding distribution of all summer days, depending on the decade considered. However, the smoothed characteristics of the data violate independence assumptions, and thus prevent a formal assessment of whether such trends are statistically significant.

While the evolution of these analogue percentiles remain largely consistent across the different N -day time windows, the results for those days below the most stringent of the RMSE analogue thresholds (R10) do show more variable answers through the twentieth century. This is largely a sample size issue: the number of total analogue days using the R10 threshold is smaller, and thus more susceptible to variations in mean analogue temperatures—particularly during the twenty-year periods ending in the 1960s–70s when fewer analogue days are found overall (see Fig. 4).

Overall, these latter results suggest that, while positive ($\approx +2\text{ }^{\circ}\text{C}$) temperature anomalies are found during those days when MSLP patterns are approximately analogous to those witnessed during the 1947 heatwave (relative to all summer days), these analogue temperatures are not found to be increasing any faster than the general trend for daily summer temperatures over the last 114 years.

3.2 Transitioning from daily-scale circulation analogues to analogues spanning the 1947 event length

Understanding how changes to the frequency of daily circulation features like those witnessed during the 1947 extreme heatwave have evolved through the twentieth century, as well as the associated temperature anomalies are highly informative. However, to gauge a better estimate of changes to the likelihood of experiencing a 12-day event specifically like what was witnessed during the summer of 1947 requires a modified methodological framework, as has been suggested in previous research (Yiou et al. 2017; Jézéquel et al. 2018b).

As explained in Sect. 3.1.1, to obtain a distribution of daily RMS error values, we considered each of the 12 days

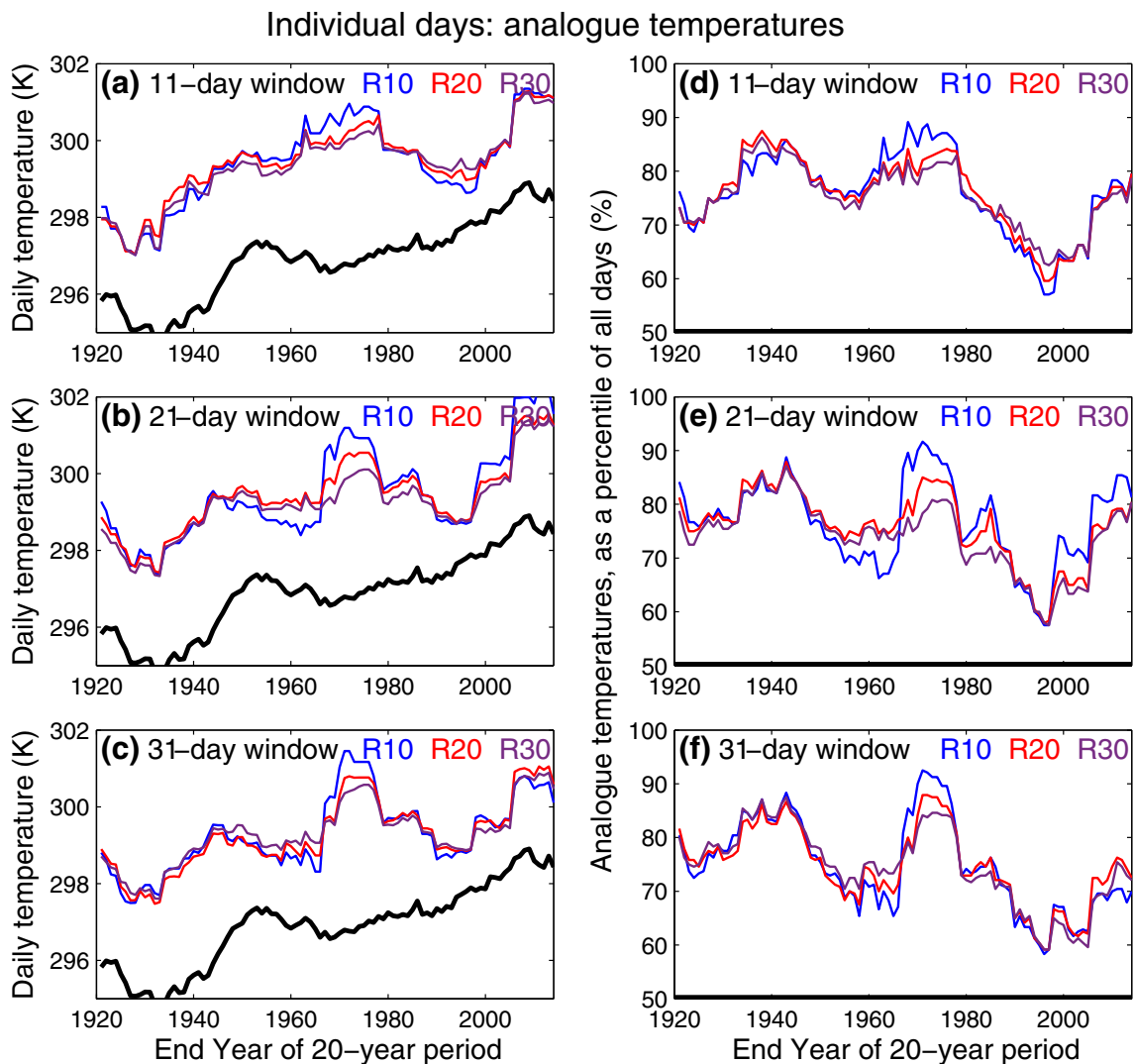


Fig. 5 Evolution of all daily summertime temperatures and daily analogue temperatures over the Twentieth Century. Black solid lines in panels **a–c** show the mean temperature of all days spanning July 22–August 2 for running twenty-year periods from 1901 to 2014 (excluding 1947); the corresponding blue, red and purple lines show

of the 1947 heatwave individually, and compared that day with all other days in an 11-, 21- or 31-day window centred on the day of interest and across the full 99-year base period, before extracting the RMS error corresponding to the 10th-, 20th- and 30th- lowest ranked days.

Repeating this for all 12 days of the heatwave then yielded a large distribution of daily RMS errors, and taking the mean of the lowest Nth-ranked values from each of the twelve individual distributions provided three objective distance thresholds, with any day falling below these thresholds consequently being considered a ‘circulation analogue’.

To instead understand the probability of witnessing the specific 12-day sequence of MSLP patterns which occurred over the course of the 1947 heatwave, we employ a Monte

Carlo sampling approach adapted from Yiou et al. (2017). First, we randomly select one day from all possible days across the full 1901–2000 base period and within the 11-day (or 21-day or 31-day) time window centred on the start of the heatwave, and then calculate the RMS error by comparing this MSLP pattern with the first day of the 1947 heatwave. A second random day is then selected, compared with the corresponding second day of the 1947 heatwave and so on, until twelve RMS errors are found. These are then averaged to produce a 12-day mean RMS error estimate, which represents how closely twelve randomly selected summer days throughout the time series resemble the specific 12-day sequence of MSLP patterns that occurred during the 1947 event.

This process is then repeated ten thousand times to produce a large distribution of resampled RMS errors, such that each value represents the similarity of a random 12-day sequence of MSLP patterns from anytime in the baseline period, with the corresponding sequence of MSLP patterns which specifically occurred during the 1947 heatwave.

As Fig. 6b, c reveal, the 12-daily average RMS error distribution has significantly different statistical characteristics compared to the aggregation of the twelve daily-scale RMS error distributions. Thus, to obtain a RMS error threshold which defines circulation analogues in a “12-day mean” setting, we use quantile mapping (Jeon et al. 2016) with the

thresholds found from the daily analogue analysis, such that the P_{BELOW} across the baseline period is the same for both the daily and the bootstrapped 12-day distributions. We note that this quantile mapping represents an RMSE adjustment by a factor of about 1.6, a value comparable to the ‘safety factor’ of 1.5 suggested by Yiou et al. (2017) for a similar approach.

We further note that many of the 12-day sequences within the upper tail of the PDF in Fig. 6c would likely lack dynamical coherence in their day-to-day MSLP evolution, by virtue of the random selection process. However, these potentially unphysical sequences escape any further scrutiny in our

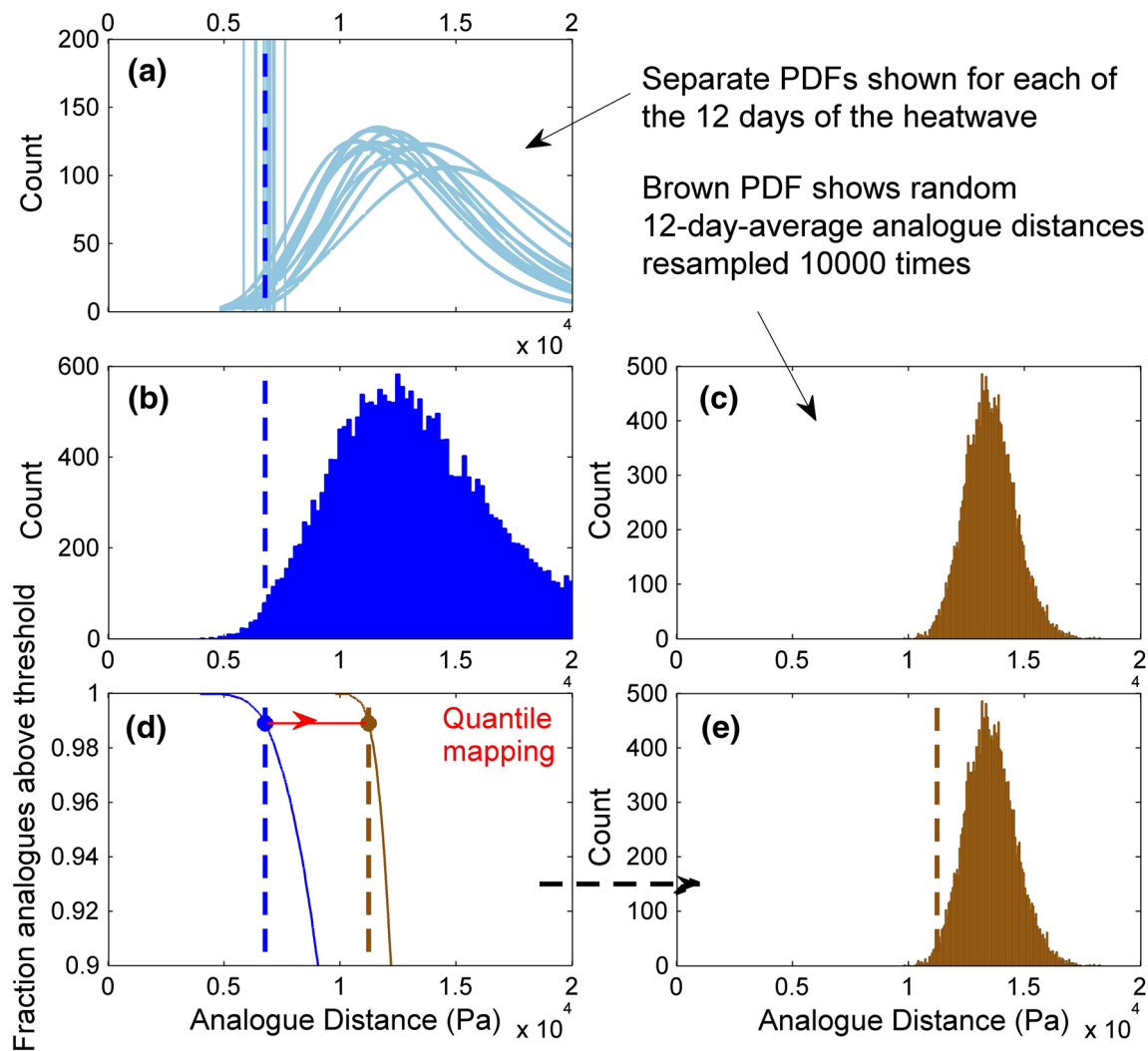


Fig. 6 Schematic illustration reconciling the difference between circulation analogues being assessed at a daily scale and for bootstrapped 12-day averages. **a** PDFs of RMS errors calculated for each of the 12 days of the 1947 heatwave across the full baseline period. The light blue lines show the threshold corresponding to the 20th lowest ranked distance in each distribution; the dark blue line shows the mean of these 12 individual thresholds and hereafter represents an arbitrary benchmark for defining MSLP circulation analogues related

to the 1947 heatwave, from a daily-scale perspective. **b** The PDFs for each of the twelve days in panel **a** aggregated into a single distribution. **c** shows a bootstrapped distribution of RMS errors for running 12-day periods; while panels **d**, **e** Show how quantile mapping is used to identify the corresponding threshold for defining a ‘circulation analogue’ within this resampled “12-day average” distribution of RMS errors

analysis, since specific focus is placed only on the lower tail of the RMSE distribution, where sequences exhibit a much closer resemblance to the actual 1947 event, by definition.

3.3 Resampled 12-day average results

3.3.1 Evolving likelihoods of witnessing 12-daily MSLP patterns like the 1947 heatwave

To explore the differences in the evolution of circulation analogue frequency between the daily-scale results presented in Sect. 3.3, and changes which more specifically reflect the full length of the 12-day heatwave in the summer of 1947, we repeat the process outlined in Sect. 3.2, using the same sequence of N -day time windows P_{BELOW} terms to consider, but also exploring multiple different bootstrap distribution sizes: specifically, using samples with 1000, 10,000 and 100,000 members. This investigation was particularly motivated by the fact that 1000-member resampled distributions have been used extensively in previous studies of circulation analogues related to extreme heatwaves (Yiou et al. 2017; Jézéquel et al. 2018b), yet the sensitivity of any potential trends to the size of these distributions remained poorly understood.

Figure 7 presents the results for the resampled 12-day average RMS error distributions, using the same methodological framework as for Sect. 3.3 aside from the caveats explained above. Because of the use of Monte Carlo resampling to produce these 12-day averaged PDFs, the figures presented in Fig. 7 show no confidence intervals as was the case for Fig. 4: instead, each of the three individual lines with the same colour in a given figure panel represent results using different bootstrap ensemble sizes.

Interestingly, these results collectively reveal changes in the time-evolution of ‘circulation analogue’ probabilities which are much different to those found in Sect. 3.1.1. While the results for daily-scale MSLP analogues showed few occurrences of statistically significant deviations from a probability ratio of 1, the opposite appears true when considering the frequency of occurrence of circulation analogues over resampled 12-day periods. Up until the 1940s—and to a lesser extent for those 20-year periods ending in the 1970s—the probability of witnessing 12-day MSLP sequences analogous to the 1947 heatwave event was substantially lower than average. Meanwhile, for every 20-year period since 1976–1995, there are substantive increases in the likelihood of witnessing such circulation analogues, with probability ratios reaching between three and six for in the most recent decades. While these results appear to qualitatively resemble corresponding changes in the Atlantic Meridional Oscillation over the same period (Sanderson et al. 2017), decomposing the relative effects of low-frequency modes of variability versus changes to European aerosol or greenhouse gas

emissions (King et al. 2016; Undorf et al. 2018) is beyond the scope of this analysis.

There are also important differences in these changes which emerge when closely examining the range of different N -day time windows used from which to sample random days, the size of the bootstrapped distribution as well as the choice of threshold to define a circulation analogue. While the latter two of these factors lead to surprisingly minimal variations in the evolution of probability ratios (Fig. 7d–f), the length of the time-window which is chosen to resample days from is found to modify the PR trajectories through the twentieth century substantially. When an 11-day window is applied (Fig. 7d), substantive increases in the likelihood of witnessing circulation analogues in the second quarter of the twentieth century is apparent (despite 1947 being excluded from consideration throughout), and probability ratios reach as high as six for the most recent 20-year periods. But if a 31-day time window is used to sample 12-day MSLP sequences, then the analogue occurrence frequencies for the 20-year running periods ending in the 1950s return to the baseline rate ($PR \approx 1$), while the PRs found in the most recent decades of the time series reduce by a factor of two (Fig. 7f). We do however note that results from the 31-day window appear largely consistent to those using a 61-day window (Supplementary Figure S3), suggesting that PR sensitivity is most pronounced only when the length of the time window selected is narrower than a month.

Collectively, these results highlight the sensitivity of discussions about low-probability event occurrences which can emerge when fundamental limitations exist in the size of the observational record from which resampled bootstraps can be taken.

3.3.2 Evolving changes to uchronic analogue temperatures

In keeping with the comparison between circulation analogue results at a daily scale versus at the resampled 12-daily scale, we next consider the changes to analogue temperatures for those 12-day MSLP sequences identified as circulation analogues for the 1947 heatwave. We will hereafter refer to these as uchronic analogue temperatures, as they are “temperature anomalies that might have occurred for a given circulation pattern sequence”, following the definition from Jézéquel and colleagues (2018b). Figure 8a shows results comparable to those presented in Fig. 5c: the black line shows the mean uchronic temperature found from the distribution of all random 12-day sequences calculated with the bootstrapping approach, while the coloured lines show the corresponding mean of all uchronic analogue temperatures in each running 20-year period. We note that the results using 1000 only bootstraps have been omitted from this section for showing volatile patterns of change, and are instead discussed

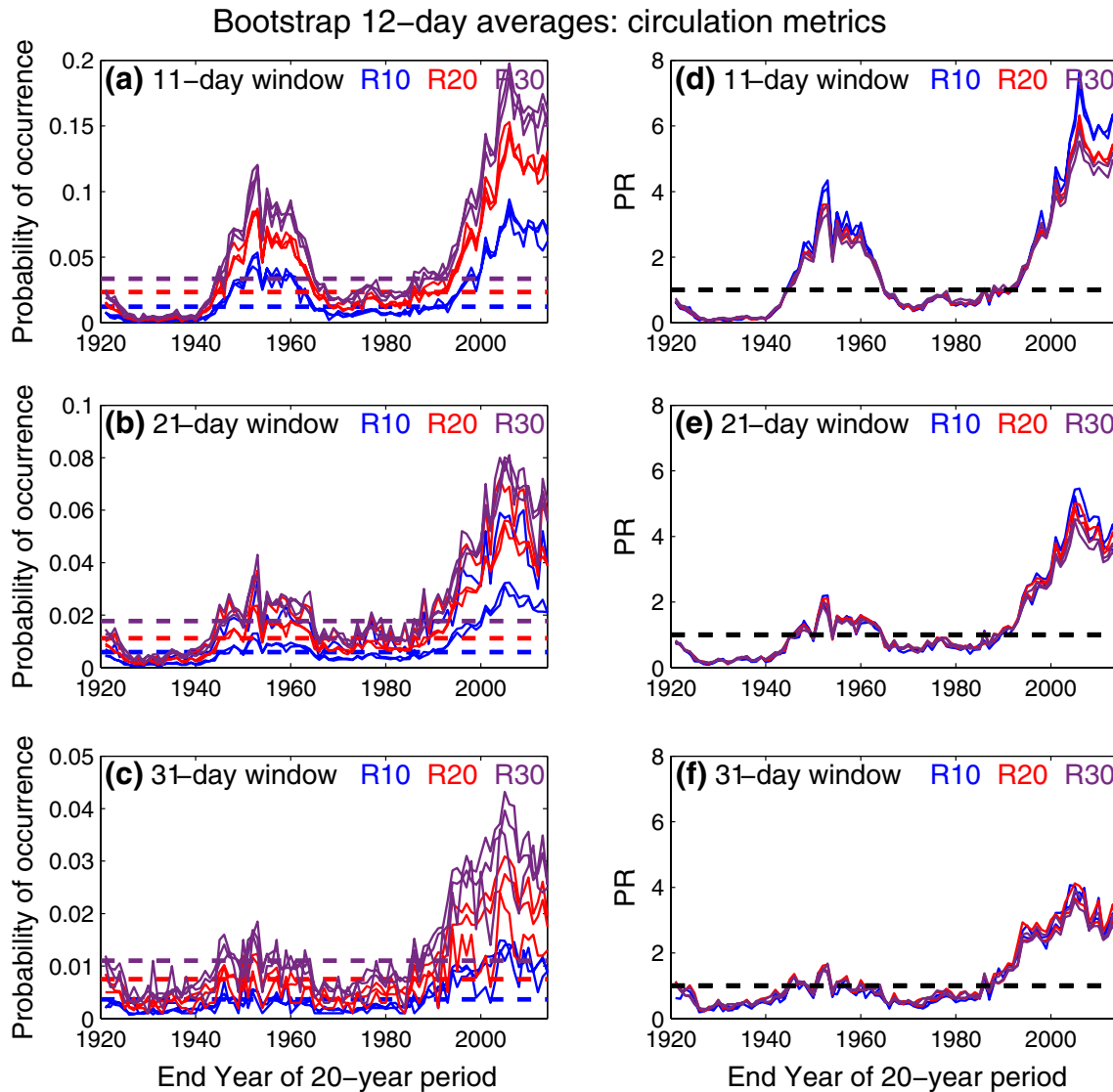
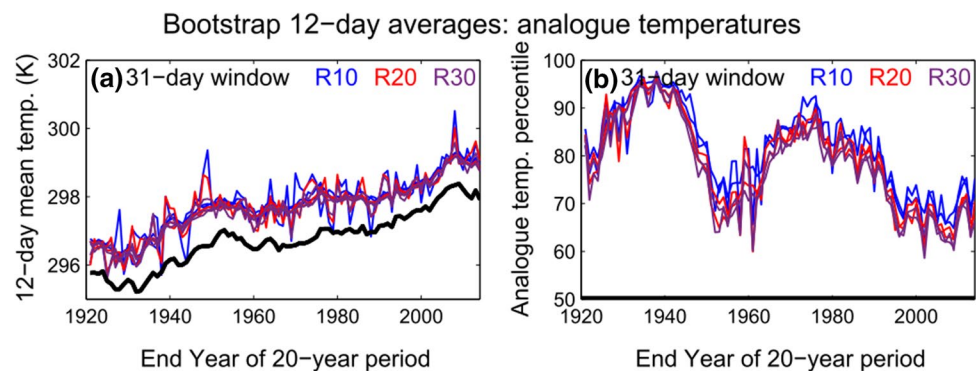


Fig. 7 Time-evolution of circulation analogue frequencies within the bootstrapped “12-day average” framework. Figure panels are the equivalent to those of Fig. 4, but with two key differences. First, three individual lines are presented for each colour in panels a–c: each of these represent distributions where the bootstrap sample size was dif-

ferent, containing either 1000, 10,000, or 100,000 members. Second, there are no confidence intervals associated with any of the trajectories, since the distributions are already the product of bootstrapped RMS error estimates

Fig. 8 Same as Fig. 5c, f, but with coloured lines representing the mean 12-daily temperatures for those resampled periods which are considered ‘circulation analogues’. Results have only been shown for the 10,000- and 100,000-member bootstrap ensembles: the 1000-member ensemble has been omitted here and is instead presented in more depth in Fig. 10



further in Fig. 10. We have also only shown results for the 31-day window, as minimal changes were observed by varying the N -day window for these analogue temperatures (see Figure S4).

Qualitatively, the average of all uchronic temperatures over the 1901–2014 period (black line in Fig. 8a) shows a similar evolution to those increases found at a daily-scale in Fig. 5, with warming of more than 2 °C observed between the start and the end of the time series. Similar increases over the length of the record are also found for the mean of the uchronic analogue temperatures.

Differences between the daily-scale analogue temperatures and those of the 12-daily uchronic analogue temperatures do exist however, specifically in terms of the added warming in the analogue-only temperatures relative to the values across all days: while an anomaly of about + 2 °C was found for a given 20-year period between the coloured and black lines in Fig. 5a–c, this anomaly is reduced to between + 0.5 °C and + 1 °C in the corresponding results for Fig. 8. However, variability in daily-scale temperatures found in the summertime is also generally much larger than for the distribution of resampled 12-day mean uchronic temperatures of Fig. 8 (not shown)—thus understanding these anomalies in a normalised context will instead provide a more useful interpretation.

Indeed, when understanding where the mean uchronic analogue temperature is placed as a percentile of all uchronic 12-day temperatures (Fig. 8b), a clearer narrative emerges as to what extent constraining on 12-day sequences with MSLP patterns comparable to the 1947 leads to temperature anomalies higher or lower than without such circulation constraints. Overall, the constrained uchronic analogue temperatures sit more towards the tail of the distribution than for the daily-scale results in Fig. 5, with most 20-year periods showing the uchronic analogue temperature mean spread approximately around the 80th percentile of the full uchronic temperature distribution. Better agreement is also found between different RMSE thresholds too, though we note that even these small changes in analogue percentiles between R10 and R30 (as with Fig. 5) could remain relevant, if they were used as event thresholds for a probabilistic attribution analysis (Harrington and Otto 2018).

Most importantly, there is no obvious increasing trend in this percentile threshold through the twentieth century, and if anything a small (but statistically insignificant) decreasing trend is apparent. This is consistent with the corresponding figures shown in Fig. 5, and shows even for the case of using mean temperatures from randomly resampled 12-day sequences, that no added increase in temperature anomalies is found associated when constraining on the circulation features of the 1947 heatwave, beyond the warming trends which are apparent for all summertime temperatures over the twentieth century.

4 Discussion and Implications

4.1 Sensitivity of resampled 12-day results to the number of bootstraps used

Some of the suggested techniques to interrogate probabilistic changes in circulation analogue frequency, particularly from an event attribution perspective, rely heavily on bootstrapping techniques (Yiou et al. 2017). This is particularly true for characterising event sequences of a specific length, such as the twelve days of the 1947 heatwave for example (Sect. 3.2).

However, the extent to which previous attribution-related results using circulation analogues might depend on the size of the bootstrap ensemble employed remains unclear. To specifically answer these questions in the context of the 1947 drought, we first repeat the process of calculating RMS error estimates for randomly resampled 12-day sequences over the full 1901–2000 baseline period (as explained further in Sect. 3.2). We do this for one million bootstrap samples, and then identify the percentile corresponding to the 1st percentile (provisionally referred to as a 1-in-100 year event hereafter). We then repeat this process, but using only 1000 bootstrap samples instead—a number commonly employed in several previous studies (Yiou et al. 2017; Jézéquel et al. 2018b). We record the 1st percentile estimate from this 1000-bootstrap ensemble, and then repeat this step twenty more times. As seen in Fig. 9, the estimates of a ‘1-in-100 year’ event based on the smaller 1000-member ensembles can vary substantially when looking at the much larger million-member ensemble, thereby highlighting the potential uncertainty which might emerge when wanting to understand changes to a specific return period of interest.

Another potential source of variability, when considering ensembles of 12-daily analogues with only 1000 members, is found when considering the ‘analogue-constrained’ temperature distributions relative to all 12-daily uchronic temperature sequences. As already shown in Fig. 8b, there are consistent results showing how hot the mean temperature of the ‘analogue-only’ sequences are, relative to all available days for the bootstrap ensembles with more than 10,000 members: they exhibit a stable evolution between the 70th and 90th percentile of the ‘all-days’ distribution, with some coherent variations from decade to decade. However, equivalent results from the ensemble with only 1000 members (Fig. 10d–f) reveals much more dramatic variability, even when shifting the 20-year window along by only several years. This instability suggests a bootstrap ensemble size of only 1000 is simply too small (and the related stochastic uncertainty too large) to appropriately infer where the analogue temperatures of a specific 20-year

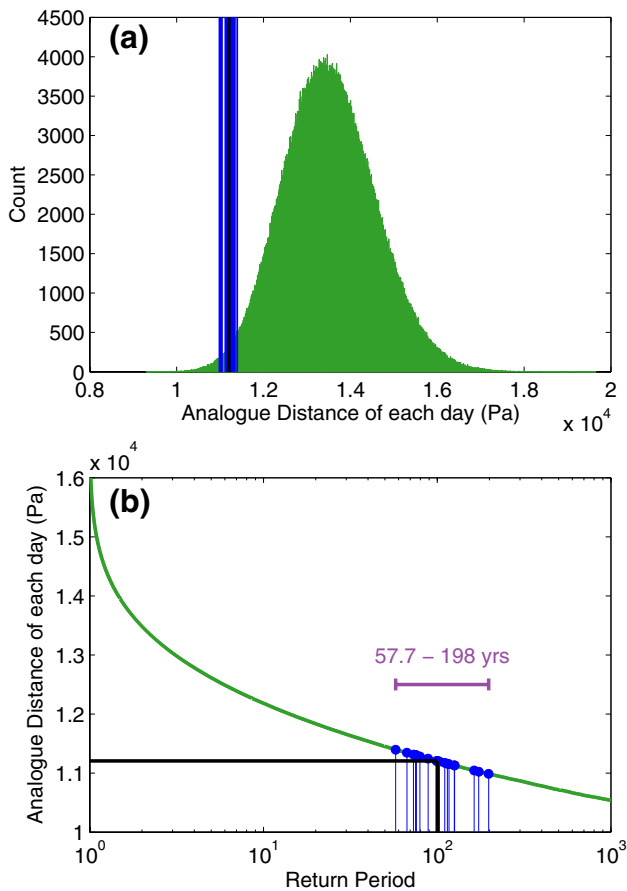
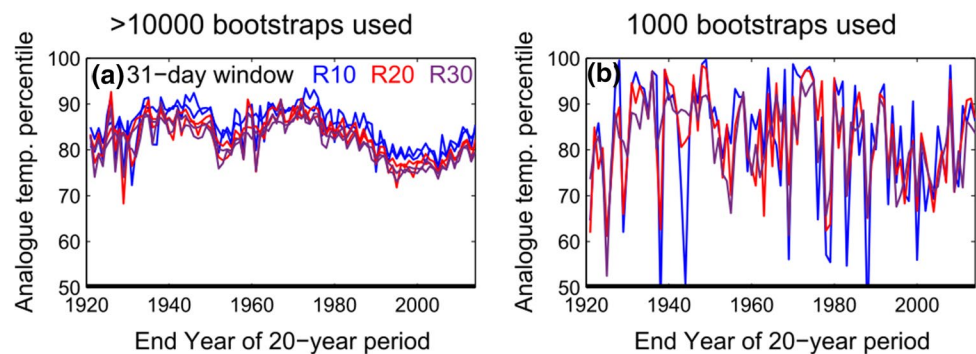


Fig. 9 **a** Green distribution shows 12-day average RMS error calculations with respect to the 1947 heatwave, based on one million bootstrap samples of random 12-day sequences from all summer days in the 1901–2000 baseline period. The black vertical line shows the RMS error value corresponding to a 1-in-100 year return period (1st percentile) based on the full 1,000,000-member ensemble. The twenty thin blue lines represent corresponding estimates of a 1-in-100 year return period, but each being based on a different bootstrap ensemble containing only 1000 members. **b** Presentation of the same results as panel **a**, but in a return period plot to emphasise the distribution tail. The black lines denote the 1-in-100 year event based on the 1,000,000-member bootstrap ensemble of 12-day RMS error values; the blue lines show the spread in ‘real’ return periods (i.e. with respect to the larger ensemble) for 1-in-100 year return levels derived from the smaller 1000-member bootstrap ensembles

Fig. 10 **a** Is identical to Fig. 8b, showing the average temperatures of all analogue 12-day periods as a percentile of all summertime temperatures for those 12-day ensembles with 10,000 members or more. **b** Shows equivalent results but for the 1000-member bootstrap ensemble only



period are positioned relative to the full distribution of summertime temperatures. Since this characterisation is crucial to understanding any decomposition of thermodynamic and dynamic changes in extreme temperatures over time, as well as to what extent the circulation metric (MSLP over wider Europe) is a useful constraint on witnessing extreme heat, these results emphasise that the size of the bootstrap ensemble employed for any analysis must be carefully considered.

4.2 Reconciling differences between changes to the frequency of daily-scale circulation analogues and bootstrapped 12-day analogues through the twentieth century

Arguably the most interesting result to emerge from Sect. 3 lies in the significant differences in analogue probability ratio evolution between the aggregated daily-scale distributions (Fig. 4) and the resampled 12-day sequence distributions (Fig. 7). These results are also broadly consistent with the fact that (Jézéquel et al. 2018a) found negligible trends in single-day circulation analogue frequency over the observational record for two European heat events, while (Yiou et al. 2017) found robust changes to the frequency of resampled 31-day analogue sequences related to the unusually wet January of 2014 over the Southern UK.

While the differences are stark between Figs. 4 and 7, they can be most likely explained by the differences in the daily and 12-daily distributions of RMS errors presented in Fig. 6b, c, and specifically the differences in variance between these distributions. Because a singular RMS error value for a resampled N -day sequence is found by taking the average of N individual RMS error values, the noise of any subsequent distribution will be implicitly reduced when looking at sequences of larger and larger values of N —that is, when characterising extreme events which span a longer period of time. However, the ‘signal’ of changes to the number of analogue days which occur through the time period will not change, whether daily-scale values are aggregated together or resampled 12-day sequences are used. Therefore, the signal-to-noise ratio, and thus increases in probability

ratios through the twentieth century, are proportionally more likely to emerge for those extreme events which persisted over a longer number of days. Such results are qualitatively equivalent to previous research finding a robust anthropogenic signal in the occurrence of exceptionally hot years over a specific location, yet finding a negligible attributable signal when looking at any of the constituent seasons (King et al. 2016): the former distribution is inherently narrower, so less of a signal is required for statistically significant differences to emerge.

Nevertheless, the fact that changes between Figs. 4 and 7 are so distinct is somewhat surprising, and further research is needed to establish what fraction of the probability ratios are simply due to the dilution of noise from resampling MSLP sequences over multiple days, versus any physical reasons which might explain the interesting decade-to-decade PR variations seen in Fig. 7d–f, and particularly the surges in analogue frequency from the 1980s onwards.

4.3 Challenges in characterising changes to the likelihood of witnessing conditions ‘like the 1947’ heatwave through time

Another important aspect of this analysis concerns the feasibility of quantifying the time-evolution of probabilistic changes in witnessing conditions (either circulation or temperature anomalies) like that witnessed during the summer of 1947. Some potential research questions in this context might include, ‘was the type of extreme heat witnessed in 1947 more likely to occur in the early 2000s than in the 1940s?’ or ‘was the extreme heat witnessed in 1947 more likely to occur in the 1940s than the beginning of the twentieth century?’ To answer these questions, an obvious first choice for a baseline distribution might be the 20-year period centred about the 1940s. However, there are multiple reasons why this choice is troublesome: first, if a time-series of probability ratios pass through the baseline period, (Sippel et al. 2015b) have demonstrated this would lead to artificial changes in the risk ratio when comparing years within the baseline with those out of the baseline period. Further, in the context of propagating uncertainties, Hawkins and Sutton (2015) have highlighted how the magnitude of uncertainty in probability ratios through the twentieth century will, by definition, be negligible as you pass through the baseline period for the counterfactual, and become proportionally larger depending on how far away from the baseline period is being considered, whether that is forwards or backwards in time. This also has clear implications for identifying the time of emergence of a robust anthropogenic signal on the increased likelihood of extreme heatwaves.

These issues collectively suggest there is no suitable counterfactual period to benchmark the time-evolution of the risk ratios against, so a subjective decision was made in

this analysis for the ‘baseline’ to span 1901–2000. Of course, a consequence of this decision means we have not made any specific statements regarding the probabilistic increase in witnessing extreme 1947-like heatwaves in the early 2000s relative to the 1940s. While such answers could be inferred by the results in Sect. 3, we emphasise that comparisons between any two time periods should be made with caution, given the decadal variability in PR which exists in Figs. 4 and 7, as well as decadal variability in the ‘severity’ of analogue-constrained temperatures found in Fig. 8b.

4.4 Ambiguity in quantifying the ‘dynamical’ contribution to an attribution statement

There remain open questions about the extent to which constraining an ensemble according to matching MSLP patterns actually meaningfully selects those days which experience extreme heat more frequently, and whether or not this actually matters. In their sensitivity analysis, Jézéquel and colleagues (2018b) found that analogue temperatures constrained using 500 hPa geopotential height (Z500) anomalies captured higher temperature anomalies in the observational record over Europe, relative to when MSLP anomalies were used. However, one could interpret differences in the type of dynamical constraint used as simply producing differences in the percentile of all temperatures that the analogue-constrained temperatures correspond to (as with Figs. 5 and 8). There is no clear answer as to what constitutes an optimal result: one could add additional criteria as to what an ‘analogue’ for the characteristics of the extreme event looks like (such as vertical velocity profiles, or heat advection pathways) and doing so would mean that fewer observational analogues exist but they would show hotter and hotter temperatures. Yet, from an attribution perspective, this is just another way of changing how one frames the event definition, and the added value of providing more and more specific constraints on ‘what happened’ during a specific extreme event, especially when questioning whether or not such events will become more common in a warmer world, remains unclear (Hannart et al. 2016; Otto et al. 2016; Harrington 2017).

A further consideration is whether or not warming-driven changes to background circulation patterns should be considered as part of the analogue analysis a priori. Jézéquel and colleagues (2018b) argue that mean changes to both Z500 and MSLP fields generally constitute a thermodynamic signal, and thus any overall trend in Z500 or MSLP through time should be removed before an analogue analysis to infer dynamical changes to the likelihood of extreme heatwaves is performed. While such an approach would be reasonable to account for the overall thermodynamic increase in tropospheric depth when considering Z500 trends, mean changes to MSLP could equally reflect dynamically driven

changes in the prevalence of blocking high pressure systems in general, and thus could be argued as a dynamical signal (Smoliak et al. 2015; Parsons et al. 2016; Harrington et al. 2016; Lehner et al. 2017; Gibson et al. 2017; Woollings et al. 2018). Hence, we have chosen not to consider any removal of long-term mean trends in MSLP before performing the circulation analogue analysis in this study.

Any potential assumption of independence between changes in the likelihood of witnessing 1947-like circulation patterns, and changes to the likelihood of witnessing 1947-like temperatures *given* the presence of analogous circulation patterns, also requires further scrutiny. If a conventional attribution analysis using climate models were to be performed (Vautard et al. 2016; Stott et al. 2016; Uhe et al. 2016; van der Wiel et al. 2017; Hauser et al. 2017; Otto et al. 2018) and a robust increase in the likelihood of witnessing 1947-like circulation patterns found (like in Fig. 7), then the top N analogues from both the counterfactual and factual distributions could no longer be used to evaluate the risk ratios of analogue-only temperatures. Otherwise, if there is a risk ratio in the analogue-only temperatures, which might be interpreted as the ‘thermodynamic signal’, this may actually just reflect the fact that those extra days which bear a closer dynamical resemblance to the event in the factual world also experience higher temperatures, rather than there being any robust signal if the two distributions of RMS error metrics were otherwise identical. In reality, an intermediate step would be required to sub-sample circulation analogues with comparable RMS errors, and only then could these analogue temperatures be compared to infer a thermodynamic risk ratio.

The collection of issues discussed above highlights a broader question: to what extent can any singular metric—like the sum of Euclidean distance anomalies across a regional Z500 or MSLP field—fully characterise the contribution of ‘dynamical climate changes’ to the likelihood of an extreme weather event occurring? In reality, no metric will serve as a perfect proxy for the weather systems conducive to an extreme heatwave event. Therefore, any attempt to look at changes in the frequency of witnessing such weather regimes—which one might consider to reflect the ‘dynamical’ component of a probability ratio—will in fact just be a partial representation of the ‘true’ role of human influence on circulation changes conducive to heatwave occurrence, with the remainder instead being (perhaps mistakenly) ascribed as a thermodynamic component of the probability ratio.

There are also further complicating aspects to the binary narrative of either thermodynamic or dynamic changes to the atmosphere combining to influence the likelihood of an extreme heatwave occurring. One example is the role of land-use changes and land–atmosphere feedbacks. Antecedent soil moisture anomalies can modulate the potential

severity of a heatwave (Seneviratne et al. 2010), and these processes can be further influenced by anthropogenic changes to land use over the twentieth century unrelated to climate (Cook et al. 2009; Lejeune et al. 2018). Such issues have not been addressed in this analysis, and warrant further consideration.

5 Conclusions

Circulation analogues can be highly useful tools to interrogate both thermodynamic and dynamically-driven changes to the frequency and intensity of extreme weather events as global temperatures continue to rise. However, particular care must be taken to understand how results may be influenced by methodological choices when defining and quantifying these analogues.

Previous research (Vautard et al. 2016; Yiou et al. 2017; Jézéquel et al. 2018b) has provided exceptionally useful insights as to some of the uncertainties which influence the application of circulation analogue techniques for understanding recent changes to extreme weather events. The results presented in this study extend these previous analyses in two ways: first, by interrogating the sensitivity of results to some previously overlooked methodological choices; and second, by examining how the use of circulation analogue constraints on extreme temperature analysis can be best understood for a case study which happened not in the present day, but in the first half of the twentieth century. The latter issue extends to wider questions of attribution framing and how to consider the time-evolution of probabilistic changes to the frequency and intensity of extreme weather events.

Our sensitivity analysis of the 1947 Central European heatwave, using only 20CR data spanning 1901–2014, yielded several new insights:

1. When considering changes in the frequency of witnessing individual daily MSLP patterns like those witnessed during the 1947 heatwave, we find negligible increases in observing such analogues through time. However, we do find substantive increases in the likelihood of witnessing 12-day MSLP sequences analogous to the 1947 event in more recent decades. This suggests that the identification of robust temporal changes (either positive or negative) in the likelihood of witnessing a specific sequence of circulation features ‘analogous’ to those of a specific extreme event will be more likely for events which persist over a longer period of time (due to an enhanced signal-to-noise ratio).
2. Whether a robust difference exists in the frequency of witnessing circulation analogues between any two time periods in the twentieth century is also highly sensitive

to the length of the N -day time window chosen from which analogue days can be selected, particularly when $N < 31$. However, negligible differences are found to result from different choices in the RMS error threshold chosen to define a circulation analogue.

3. Sensitivity exists as to the number of bootstrap samples needed to yield robust results for an event of specific length. When considering only 1000 bootstrapped estimates of resampled 12-day MSLP sequences from the observational record (compared with $> 10,000$ runs), we found, (1) much greater variability to exist in both identifying what RMS error constitutes an event of a specific return period (Fig. 9); and more importantly, (2) substantive variations in where the analogue-constrained temperature distribution was placed relative to the temperature distribution of all days (Fig. 10). This suggests care is needed when inferring the importance of a particular dynamical state increasing the likelihood of witnessing extreme heat more frequently, and shows the importance of multiple sensitivity tests.
4. Comparing 20-year periods ending in 2014 versus ending in 1920 shows increases of up to 3 °C in summertime temperatures over the Central European domain of interest. However, there are no clear differences in the temperature trends between those days which can be classified as circulation analogues of the 1947 heatwave, and those which are not.
5. Both increases to the frequency of witnessing 12-day sequences of circulation characteristics like the 1947 heatwave, coupled with robust and consistent warming trends in the distribution of temperatures general over the twentieth century, suggest witnessing the combined characteristics reminiscent of the heatwave which occurred in late July and early August 1947 are now more likely than at any other time over the twentieth century (Figs. 7 and 8, respectively). However, quantifying the exact extent to which these likelihoods have increased, and the relative contributions of dynamical and thermodynamic changes therein, remain highly sensitive to decadal variations which are in turn sensitive to choices in the methodological framework used.

Important questions remain in the scientific community about, (1) how best to partition and characterise different components of the state of the earth system at the time when an extreme weather event occurs; (2) understanding the extent to which each of these components combined to make the event as extreme as it was; and then finally (3) interrogating how anthropogenic changes to the climate system might have modified each of these components, with the potential for that event to consequently produce more severe impacts than if it otherwise occurred in the absence of these human influences. This study sought to understand the strengths and

limitations of employing circulation analogue techniques as a tool to answer aspects of each of these three questions. The answer to the first remains a fundamentally subjective choice (for example, whether to constrain circulation analogues using MSLP or Z500), and these choices are known to consequently modify the interpretation of the answers to both questions 2 and 3 (Harrington 2017). However, the results presented in this study also show how substantive variations in the answers to questions 2 and 3 can also emerge, depending on several methodological choices which are necessary to proceed with any extreme event analysis, yet might have previously been considered inconsequential.

A warming world might influence how often weather patterns conducive to exceptionally hot weather will occur over Central Europe (Jézéquel et al. 2018a; Woollings et al. 2018), as well as how warm summertime temperature maxima will reach. However, our results show further work is needed to properly quantify the uncertainty of these changes, both to understand the evolution of changes in the past and as warming continues into the future.

Acknowledgements The 20th Century Reanalysis data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, and available at <http://www.esrl.noaa.gov/psd/>. Support for the Twentieth Century Reanalysis Project version 2c data set is provided by the U.S. Department of Energy, Office of Science Biological and Environmental Research, and by the National Oceanic and Atmospheric Administration Climate Program Office. The authors thank Robert Vautard and Aglae Jézéquel for helpful discussions on topics related to the manuscript. LJH acknowledges support from the Transition into the Anthropocene (TITAN) project, funded by a European Research Council (ERC) Advanced Grant (EC-320691).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Angéilil O, Stone D, Wehner M et al (2017) An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. *J Clim* 30:5–16. <https://doi.org/10.1175/JCLI-D-16-0077.1>
- Calanca P (2007) Climate change and drought occurrence in the Alpine region: how severe are becoming the extremes? *Global Planet Change* 57:151–160. <https://doi.org/10.1016/j.gloplacha.2006.11.001>
- Cassou C, Terray L, Phillips AS (2005) Tropical Atlantic influence on European heat waves. *J Clim* 18:2805–2811. <https://doi.org/10.1175/JCLI3506.1>
- Cattiaux J, Ribes A (2018) Defining single extreme weather events in a climate perspective. *Bull Am Meteorol Soc* 99:1557–1568. <https://doi.org/10.1175/BAMS-D-17-0281.1>

- Cattiaux J, Vautard R, Cassou C et al (2010) Winter 2010 in Europe: a cold extreme in a warming climate. *Geophys Res Lett*. <https://doi.org/10.1029/2010gl044613>
- Cattiaux J, Yiou P, Vautard R (2012) Dynamics of future seasonal temperature trends and extremes in Europe: a multi-model analysis from CMIP. *Clim Dyn* 38:1949–1964
- Christiansen B (2015) The role of the selection problem and non-gaussianity in attribution of single events to climate change. *J Clim* 28:9873–9891. <https://doi.org/10.1175/JCLI-D-15-0318.1>
- Christidis N, Stott PA (2015) Extreme rainfall in the United Kingdom during winter 2013/14: the role of atmospheric circulation and climate change. *Bull Am Meteorol Soc* 96:S46–S50. <https://doi.org/10.1175/BAMS-D-15-00094.1>
- Compo GP, Whitaker JS, Sardeshmukh PD et al (2011) The twentieth century reanalysis project. *QJR Meteorol Soc* 137:1–28. <https://doi.org/10.1002/qj.776>
- Cook BI, Miller RL, Seager R (2009) Amplification of the North American “Dust Bowl” drought through human-induced land degradation. *Proc Natl Acad Sci* 106:4997–5001. <https://doi.org/10.1073/pnas.0810200106>
- Deser C, Terray L, Phillips AS (2016) Forced and internal components of winter air temperature trends over North America during the past 50 years: mechanisms and implications. *J Clim* 29:2227–2258. <https://doi.org/10.1175/JCLI-D-15-0304.1>
- Fischer EM, Schar C (2010) Consistent geographical patterns of changes in high-impact European heatwaves. *Nat Geosci* 3:398–403. <https://doi.org/10.1038/ngeo866>
- Gibson PB, Pitman AJ, Lorenz R, Perkins-Kirkpatrick SE (2017) The role of circulation and land surface conditions in current and future Australian heat waves. *J Clim* 30:9933–9948. <https://doi.org/10.1175/JCLI-D-17-0265.1>
- Grütter J, Lehmann S, Auchmann R et al (2013) The heatwaves in Switzerland in summer 1947. *Geographica Bernensia*. <https://doi.org/10.4480/gb2013.g89.08>
- Hannart A, Carrassi A, Bocquet M et al (2016) DADA: data assimilation for the detection and attribution of weather and climate-related events. *Clim Change* 136:155–174. <https://doi.org/10.1007/s10584-016-1595-3>
- Harrington LJ (2017) Investigating differences between event-classes and probability density-based attribution statements with emerging climate change. *Clim Change* 141:641–654. <https://doi.org/10.1007/s10584-017-1906-3>
- Harrington LJ, Otto FEL (2018) Adapting attribution science to the climate extremes of tomorrow. *Environ Res Lett*. <https://doi.org/10.1088/1748-9326/aaf4cc>
- Harrington LJ, Gibson PB, Dean SM et al (2016) Investigating event-specific drought attribution using self-organizing maps. *J Geophys Res Atmos* 121:12–766. <https://doi.org/10.1002/2016jd025602>
- Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int J Climatol* 34:623–642. <https://doi.org/10.1002/joc.3711>
- Hauser M, Gudmundsson L, Orth R et al (2017) Methods and model dependency of extreme event attribution: the 2015 European drought. *Earth's Future* 5:1034–1043. <https://doi.org/10.1002/2017EF000612>
- Hawkins E, Sutton R (2015) Connecting climate model projections of global temperature change with the real world. *Bull Am Meteorol Soc* 97:963–980. <https://doi.org/10.1175/BAMS-D-14-00154.1>
- Hirschi E, Auchmann R, Martius O, Brönnimann S (2013) The 1945–1949 droughts in Switzerland. *Geographica Bernensia*. <https://doi.org/10.4480/gb2013.g89.09>
- Jeon S, Paciorek CJ, Wehner MF (2016) Quantile-based bias correction and uncertainty quantification of extreme event attribution statements. *Weather Clim Extremes* 12:24–32. <https://doi.org/10.1016/j.wace.2016.02.001>
- Jézéquel A, Cattiaux J, Naveau P et al (2018a) Trends of atmospheric circulation during singular hot days in Europe. *Environ Res Lett* 13:054007. <https://doi.org/10.1088/1748-9326/aab5da>
- Jézéquel A, Yiou P, Radanovics S (2018b) Role of circulation in European heatwaves using flow analogues. *Clim Dyn* 50:1145–1159. <https://doi.org/10.1007/s00382-017-3667-0>
- Jézéquel A, Yiou P, Radanovics S, Vautard R (2018c) Analysis of the exceptionally warm december 2015 in France using flow analogues. *Bull Am Meteorol Soc* 99:S76–S79. <https://doi.org/10.1175/BAMS-D-17-0103.1>
- Kalnay E, Kanamitsu M, Kistler R et al (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–472. [https://doi.org/10.1175/1520-0477\(1996\)077%3c0437:TNYRP%3e2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077%3c0437:TNYRP%3e2.0.CO;2)
- King AD, Black MT, Min S-K et al (2016) Emergence of heat extremes attributable to anthropogenic influences. *Geophys Res Lett*. <https://doi.org/10.1002/2015gl067448>
- Lehner F, Deser C, Terray L (2017) Toward a new estimate of “time of emergence” of anthropogenic warming: insights from dynamical adjustment and a large initial-condition model ensemble. *J Clim* 30:7739–7756. <https://doi.org/10.1175/JCLI-D-16-0792.1>
- Lehner F, Deser C, Simpson IR, Terray L (2018) Attributing the U.S. Southwest’s recent shift into drier conditions. *Geophys Res Lett* 45:6251–6261. <https://doi.org/10.1029/2018GL078312>
- Lejeune Q, Davin EL, Gudmundsson L et al (2018) Historical deforestation locally increased the intensity of hot days in northern mid-latitudes. *Nat Clim Change* 8:386. <https://doi.org/10.1038/s41558-018-0131-z>
- Meredith EP, Semenov VA, Maraun D et al (2015) Crucial role of Black Sea warming in amplifying the 2012 Krymsk precipitation extreme. *Nat Geosci* 8:615–619. <https://doi.org/10.1038/ngeo2483>
- National Academies of Sciences, Engineering, and Medicine, Committee on Extreme Weather Events and Climate Change Attribution, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies (2016) Attribution of Extreme Weather Events in the Context of Climate Change. National Academies Press, Washington, D.C
- Otto FEL (2017) Attribution of weather and climate events. *Annu Rev Environ Resour* 42:627–646. <https://doi.org/10.1146/annurev-environ-102016-060847>
- Otto FEL, van Oldenborgh GJ, Eden J et al (2016) The attribution question. *Nat Clim Change* 6:813–816. <https://doi.org/10.1038/nclimate3089>
- Otto FEL, Philip S, Kew S et al (2018) Attributing high-impact extreme events across timescales—a case study of four different types of events. *Clim Change*. <https://doi.org/10.1007/s10584-018-2258-3>
- Parsons S, Renwick JA, McDonald AJ (2016) An assessment of future southern hemisphere blocking using CMIP5 projections from four GCMs. *J Clim* 29:7599–7611. <https://doi.org/10.1175/JCLI-D-15-0754.1>
- Perkins SE (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmos Res* 164–165:242–267. <https://doi.org/10.1016/j.atmosres.2015.05.014>
- Perkins SE, Alexander LV (2013) On the measurement of heat waves. *J Clim* 26:4500–4517. <https://doi.org/10.1175/JCLI-D-12-00383.1>
- Perkins-Kirkpatrick SE, Gibson PB (2017) Changes in regional heatwave characteristics as a function of increasing global temperature. *Sci Rep* 7:12256. <https://doi.org/10.1038/s41598-017-12520-2>
- Pfahl S, Wernli H (2012) Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophys Res Lett*. <https://doi.org/10.1029/2012gl052261>

- Sanderson MG, Economou T, Salmon KH, Jones SEO (2017) Historical trends and variability in heat waves in the United Kingdom. *Atmosphere* 8:191. <https://doi.org/10.3390/atmos8100191>
- Seneviratne SI, Corti T, Davin EL et al (2010) Investigating soil moisture—climate interactions in a changing climate: a review. *Earth Sci Rev* 99:125–161. <https://doi.org/10.1016/j.earscirev.2010.02.004>
- Shepherd TG (2016) A common framework for approaches to extreme event attribution. *Curr Clim Change Rep* 2:28–38. <https://doi.org/10.1007/s40641-016-0033-y>
- Sippel S, Walton P, Otto FEL (2015a) Stakeholder perspectives on the attribution of extreme weather events: an explorative enquiry. *Wea Clim Soc* 7:224–237. <https://doi.org/10.1175/WCAS-D-14-00045.1>
- Sippel S, Zscheischler J, Heimann M et al (2015b) Quantifying changes in climate variability and extremes: pitfalls and their overcoming. *Geophys Res Lett* 42:9990–9998. <https://doi.org/10.1002/2015gl066307>
- Smoliak BV, Wallace JM, Lin P, Fu Q (2015) Dynamical adjustment of the northern hemisphere surface air temperature field: methodology and application to observations. *J Clim* 28:1613–1629. <https://doi.org/10.1175/JCLI-D-14-00111.1>
- Sousa PM, Trigo RM, Barriopedro D et al (2018) European temperature responses to blocking and ridge regional patterns. *Clim Dyn* 50:457–477. <https://doi.org/10.1007/s00382-017-3620-2>
- Stott PA, Walton P (2013) Attribution of climate-related events: understanding stakeholder needs. *Weather* 68:274–279. <https://doi.org/10.1002/wea.2141>
- Stott PA, Christidis N, Otto FEL et al (2016) Attribution of extreme weather and climate-related events. *WIREs Clim Change* 7:23–41. <https://doi.org/10.1002/wcc.380>
- Trenberth KE, Fasullo JT, Shepherd TG (2015) Attribution of climate extreme events. *Nat Clim Change* 5:725–730. <https://doi.org/10.1038/nclimate2657>
- Uhe P, Otto FEL, Haustein K et al (2016) Comparison of methods: attributing the 2014 record European temperatures to human influences. *Geophys Res Lett* 43:8685–8693. <https://doi.org/10.1002/2016gl069568>
- Undorf S, Bollasina MA, Hegerl GC (2018) Impacts of the 1900–74 increase in anthropogenic aerosol emissions from North America and Europe on Eurasian Summer Climate. *J Clim* 31:8381–8399. <https://doi.org/10.1175/JCLI-D-17-0850.1>
- van der Wiel K, Kapnick SB, van Oldenborgh GJ et al (2017) Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrol Earth Syst Sci* 21:897–921. <https://doi.org/10.5194/hess-21-897-2017>
- Vautard R, Yiou P, Otto F et al (2016) Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events. *Environ Res Lett* 11:114009. <https://doi.org/10.1088/1748-9326/11/11/114009>
- Woollings T, Barriopedro D, Methven J et al (2018) Blocking and its response to climate change. *Curr Clim Change Rep* 4:287–300. <https://doi.org/10.1007/s40641-018-0108-z>
- Yiou P, Vautard R, Naveau P, Cassou C (2007) Inconsistency between atmospheric dynamics and temperatures during the exceptional 2006/2007 fall/winter and recent warming in Europe. *Geophys Res Lett* 34:L21808. <https://doi.org/10.1029/2007GL031981>
- Yiou P, Jézéquel A, Naveau P et al (2017) A statistical framework for conditional extreme event attribution. *Adv Stat Climatol, Meteorol Oceanogr* 3:17–31. <https://doi.org/10.5194/ascmo-3-17-2017>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.