Climate Dynamics manuscript No.
(will be inserted by the editor)

On the nonlinearity of spatial scales in extreme weather attribution statements

Oliver Angélil · Daíthí Stone · Sarah Perkins-Kirkpatrick · Lisa V. Alexander · Michael Wehner · Hideo Shiogama · Piotr Wolski · Andrew Ciavarella · Nikolaos Christidis

Received: date / Accepted: date

O. Angelil
Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, UNSW Australia, Sydney NSW 2052, Australia
E-mail: oliver.angelil@student.unsw.edu.au

D. Stone
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

S. Perkins-Kirkpatrick
Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, UNSW Australia, Sydney NSW 2052, Australia

L.V. Alexander
Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, UNSW Australia, Sydney NSW 2052, Australia

M. Wehner
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

H. Shiogama
National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan

P. Wolski
Climate Systems Analysis Group, Environmental and Geographical Science, University of Cape Town, Rondebosch, South Africa

A. Ciavarella
Met Office Hadley Centre, Exeter EX1 3PB, UK

Abstract In the context of ongoing climate change, extreme weather events are drawing increasing attention from the public and news media. A question often asked is how the likelihood of extremes might have changed by anthropogenic greenhouse-gas emissions. Answers to the question are strongly influenced by the model used, duration, spatial extent, and geographic location of the event – some of these factors often overlooked. Using output from four global climate models, we provide attribution statements characterised by a change in probability of occurrence due to anthropogenic greenhouse-gas emissions, for rainfall and temperature extremes occurring at seven discretised spatial scales and three temporal scales. An understanding of the sensitivity of attribution statements to a range of spatial and temporal scales of extremes allows for the scaling of attribution statements, rendering them relevant to other extremes having similar but non-identical characteristics. This is a procedure simple enough to approximate timely estimates of the anthropogenic contribution to the event probability. Furthermore, since real extremes do not have well-defined physical borders, scaling can help quantify uncertainty around attribution results due to uncertainty around the event definition. Results suggest that the sensitivity of attribution statements to spatial scale is similar across models and that the sensitivity of attribution statements to the model used is often greater than the sensitivity to a doubling or halving of the spatial scale of the event. The use of a range of spatial scales allows us to identify a nonlinear relationship between the spatial scale of the event studied and the attribution statement.

Keywords Attribution · Extremes · C20C+ · AGCMs

1 Introduction

Event attribution literature has been populated by targeted studies investigating the influence of human activity on the properties and probability of recent major weather events (e.g. Stott et al, 2004; Dole et al, 2011; Peterson et al, 2012, 2013; Herring et al, 2014, 2015). Each of these studies focused on one or a few extreme

N. Christidis
Met Office Hadley Centre, Exeter EX1 3PB, UK

weather events which adversely impacted human health, infrastructure, or agriculture (or a combination of these), usually attracting substantial attention from the public and media. However, attribution statements for these events as well as the nature of associated impacts, vary according to the spatial and temporal scales chosen to define them. These definitions are often somewhat arbitrarily chosen. For example, in Stott et al (2004), who examined the 2003 European heatwave, there is a mismatch between the spatial scales for which the severest impacts were felt, and those defined in their analyses – only roughly two-thirds of the area examined in their study was European land (the remaining area was over North Africa), and mortality was mostly a consequence of a two-week heatwave in August mostly confined to western Europe (Robine et al, 2008), not a hot summer. As demonstrated by Angélil et al (2014b), the sensitivity of attribution statements to the spatial and temporal scales of the extreme event can be substantial (increasing as the spatial scale increases), however their study only tests sensitivity to very large changes in the spatial scale: from $2 \cdot 10^6$km^2 to the resolution of the two models they used, being $\sim$22500km^2 and $\sim$40000km^2 ($\sim 1.5^\circ$ and $\sim 2^\circ$).

The endogenous variability of the atmosphere depends on the spatial scale (Hawkins and Sutton, 2012) and this would be expected to translate into a dependence of event attribution calculations because of their sensitivity to the magnitude of endogenous variability (Bellprat and Doblas-Reyes, 2016). Angélil et al (2014b) revealed the existence of this scale dependence in climate model simulations, but they did not determine the functional form of the relationship. Here we expand on Angélil et al (2014b) by determining this functional relationship in a number of climate models examining the robustness of the relationship across those models, by calculating attribution statements for extremes occurring over a set of discretized spatial scales – all at subcontinental domains. The range of spatial scales allows us to more precisely characterise the relationship between the spatial scale and attribution statement – a relationship that is potentially nonlinear. Results can enable us to, for example, scale previously published attribution statements such that they are relevant to extremes occurring at slightly different spatial scales. A by-product of this sensitivity analyses are results showing the magnitudes of attribution statements across models. We therefore additionally explore reasons for differences in attribution statements between models.

We use the probabilistic event attribution framework designed by Pall et al (2011), using large initial condition ensembles from four Atmosphere-only Global Climate Models (AGCMs). The large ensembles better resolve the statistics of the rare weather events we are interested in, and such ensemble sizes are feasible since AGCMs are less computationally expensive to run compared to their coupled counterparts. Similar to Pall et al (2011) who use the Fraction of Attributable Risk (FAR), we characterise the anthropogenic contribution to the chance of the extreme with the Probability Ratio (PR) which is given by ratio of the probability of exceeding an extreme threshold in model runs forced by natural and anthropogenic influences (ALL) to the probability of exceeding the same threshold in model runs forced by only natural influences (NAT). If the PR > 1, anthropogenic greenhouse-gas emissions have increased the chance of the event. If the PR < 1, they have decreased the chance of the event. Using this framework, we take a brute force approach by calculating attribution statements on a global scale for daily, 5-day, and monthly temperature and rainfall extremes occurring at seven different spatial scales over thousands of different locations (see Fig. 1).

The goal of this paper is to understand the dependence of event attribution conclusions on the spatial scale for subcontinental domains. Some event attribution studies, for example Stott et al (2004), have considered events occurring over regions approaching continental scales (i.e. 4 million km^2 and larger). We do not consider these larger scales in this paper because our method of using fixed regions becomes more of an issue at larger scales, with the sample of regions being smaller and thus it being more difficult to distinguish between region-specific properties (for example if a region happens to contain a section of Arctic coast) and generic properties for that type of region (e.g. mid-latitude continental). Refinement of event attribution techniques to smaller scale events is identified as a major direction and challenge in the field (and others National Academies of Sciences, Engineering and Medicine, 2016) so continental-scale analyses might be expected to become less frequent in the future.

While there remain pressing questions on issues that challenge event attribution assessments, such as how multi-model ensembles can be best used to optimise key properties seen in observations (e.g. variability; dynamical response to boundary conditions), this paper directly addresses a particular question, that being on the sensitivity of event attribution analyses to the definition of the event examined.

2 Data

We use four AGCMs, each run under two forcing scenarios. The first being a factual scenario forced with natural and anthropogenic influences (ALL) simulating weather that might have occurred under observed historical boundary conditions. The second set of ensembles are run under a counterfactual "natural" scenario (NAT), in which emissions from human activities had not interfered with the climate system.

The ALL scenario is forced with observed boundary conditions for greenhouse gases, tropospheric aerosols, volcanic aerosols, ozone concentrations, solar irradiance, sea surface temperatures (SST), sea ice coverage (SIC), and land cover. In the NAT scenario, greenhouse gases, tropospheric aerosols and ozone were altered to estimate pre-industrial levels, while ocean temperatures were cooled and sea ice coverage expanded according to an estimate based on output from the international CMIP5 climate modelling effort (`http://portal.nersc.gov/c20c/input_data/C20C-DandA_dSSTs_All-Hist-est1_Nat-Hist-CMIP5-est1.pdf`). The NAT SST variability is based on observed ocean surface conditions, which preserves month-to-month and year-to-year variability, such as the El Niño-Southern Oscillation phenomenon (ENSO).

The AGCMs used are part of the C20C+ detection and attribution project (`http://portal.nersc.gov/c20c/`): the CAM5.1, MIROC5, HadGEM3-A-N216, and HadAM3P-N96 AGCMs, run at resolutions of $\sim$1.4°, $\sim$1°, $\sim$0.5° and $\sim$1.8° respectively. The area covered by grid cells varies with latitude, decreasing with increasing distance from the equator. In CAM5.1, prescribed SSTs up to 1982 are an adjusted version of the HadISST1 dataset (Rayner et al, 2003), after which the NOAA-OI.v2 dataset is used (Hurrell et al, 2008). In HadAM3P-N96, SSTs were prescribed using NOAA-OI.v2. The HadGEM3-A-N216 (Christidis et al, 2013) and MIROC5 (Shiogama et al, 2013, 2014) prescribed monthly SST and SIC were taken from the HadISST1 dataset. Any differences between the AGCMs may be partially due to CAM5.1 and HadAM3P-N96 using prescribed aerosol burdens (black carbon, organic carbon, sulfate and sea salt), while MIROC5 and HadGEM3-A-N216 simulate aerosol distributions from prescribed aerosol emissions. The MIROC5 and HadGEM3-A-N216 experimental setups therefore allow for interactions between the simulated weather and atmospheric chemistry, while in CAM5.1 and HadAM3P-N96 the absence of this interaction may prevent the occurrence of feedbacks relevant in the simulation of extremes, particularly hot events.

In HadGEM3-A-N216, all ensembles members have the same initial conditions but differences are generated using parameter perturbations and a stochastic kinetic energy backscatter scheme (Christidis et al, 2013). In CAM5.1, MIROC5 and HadAM3P-N96, each ensemble member from each AGCM differs from the next only in its initial conditions. Each model run has been trimmed to cover the January 2008 - December 2012 period. Daily means of two meter air temperature and precipitation are used in this analyses. Extremes are by definition rare, therefore in order to resolve the statistics of these events, we use the maximum number of available simulations from each AGCM. This consists of 100 ALL and NAT members from CAM5.1, 60 ALL and 50 NAT counter-factual members from MIROC5, 15 ALL and NAT members from HadGEM3-A-N216, and 50 ALL and NAT members from HadAM3P-N96.

3 Method

In order to better resolve the statistics of extreme events, all members and all years from each AGCM are pooled before any further calculations are made. All data are remapped to the coarsest model being HadAM3P-N96 using a first order conservative remapping procedure (Jones, 1999). We calculate PRs for the probability of exceeding daily, 5-day and monthly one-in-ten-year (0.027% chance of occurrence; also expressed as a 1 in 365×10 chance of occurrence) hot and cold temperature extremes, and one-in-one-year (0.27% chance of occurrence) wet extremes. However, we exclude one-in-ten-year rainfall extremes, because over many regions, they are too extreme to be accurately sampled in the NAT scenario, particularly for monthly extremes as the averaging across time increases the signal (anthropogenic) to noise (natural variability) ratio. 5-day and monthly weather are calculated by averaging daily output with 5-day and 30-day running windows with a 1-day step. This procedure simply smooths the distributions and will not systematically increase or decrease exceedance probabilities.

We select wet and cold event thresholds from ensembles driven by the ALL scenario. However since one-in-ten-year hot extremes simulated under the ALL scenario rarely occur in weather simulated under the NAT scenario, we select these thresholds from the NAT ensembles. Therefore, for hot extremes, P_{NAT} is fixed at 0.027%, and P_{ALL} varies according to the chance of exceeding the thresholds obtained from the NAT ensembles. For cold and wet extremes, P_{ALL} is fixed at 0.027% and 0.27% respectively, and P_{NAT} varies according to the probability of events being colder or wetter than the thresholds derived from the ALL ensembles. When the desired percentile lies between two data points, the value is estimated via linear interpolation.

We use this method to calculate PRs for extremes occurring over almost all land regions of the globe, at 7 different spatial scales (Fig. 1) – the largest being demarcated by the 58 regions in the Weather Risk Attribution Forecast (WRAF, Fig. 1a), and the smallest being the resolution of HadAM3P-N96 (not shown). These regions are on average $2.18 \cdot 10^6$ km^2 with a standard deviation of $4.64 \cdot 10^5$ km^2 . We define the second largest spatial scale ("$\frac{1}{2}$ WRAF"; Fig. 1b) by halving the area of each of the WRAF regions. The axis (latitudinal or longitudinal) along which regions are split is always perpendicular to the axis with the greater maximum latitudinal or longitudinal distance. After a region has been split, the areas of the two halves are equal, with accuracy being to the nearest grid cell. For the 3$^{\text{rd}}$ spatial scale ("$\frac{1}{4}$ WRAF"; Fig. 1c), we halve the areas of the "$\frac{1}{2}$ WRAF" regions. We continue to halve regions until the 6^{th} spatial scale ("$\frac{1}{32}$ WRAF") has been defined. The average area of the regions from the WRAF scale to the 6th spatial scale are: $2.18 \cdot 10^6$ km^2; $1.09 \cdot 10^6$ km^2; $5.44 \cdot 10^5$ km^2; $2.72 \cdot 10^5$ km^2; $1.36 \cdot 10^5$ km^2; and $6.81 \cdot 10^4$ km^2. The 7^{th} and smallest spatial scale is defined as the resolution of the coarsest model being HadAM3P-N96. For the 1^{st} to 6^{th} spatial scales, area-weighted averages are taken from the temperature and rainfall grid cell values at every time-step.

To prevent positive and negative infinity log(PR) values interfering with the calculations, we have artificially adjusted all cases where either P_{ALL} or P_{NAT} = 0% to a probability assuming one-tenth of an event (day, 5-day or month) exceeded (or fell below for cold events) the threshold. Depending precisely on the temporal scale examined, this for example equates to an exceedance probability of $\sim$0.000055% (0.1 in 5×100 years) in CAM5.1 (the probability will be slightly greater for other models given the number of runs is less). These cases occur over a negligible percentage of the regions, and are therefore hardly expected to effect the results.

4 Results

PRs have been computed for hot, cold, and wet extremes; occurring at 3 temporal scales; 7 spatial scales; over 58 regions of the world; using output from 4 AGCMs. Before we discuss summarised results for all models, variables, spatial and temporal scales, we begin with Fig. 2, which highlights a key contribution this analyses makes beyond Angélil et al (2014b). The figure summarises the PR for hot day extremes in CAM5.1 for tropical regions (y-axis; being the average PR for all regions occurring within the tropics at a given spatial scale) as a function of the spatial scale of the extreme (x-axis). Linear interpolation for scaling PRs had the Angélil et al (2014b) method been applied (dashed line), fails to characterise the non-linear relationship seen when 7 spatial scales are used (pink markers). An interpolated attribution statement can differ by approximately 20% for a given spatial scale where the vertical distance between the two lines is greatest.

Next, we present the main results: PRs for temperature extremes averaged separately over the tropics and extra-tropics, and PRs for rainfall extremes averaged across each of the 58 WRAF regions. The reason for summarising the results for temperature and rainfall in such a way is that PRs for temperature extremes exhibit similar values across bands of latitude, while PRs for extreme rainfall tend to vary more across smaller spatial scales (Angélil et al, 2014a,b). Results calculated for every variable over each spatial and temporal scale (without averaging across the tropics, extra-tropics, or the WRAF regions), can be found in the Supplementary Material as scatter plots (Figs. S2-S10) and as maps (Figs. S11-S13).

The pink curve in Fig. 2 is again shown in Fig. 3(a). Here however, the axes are logarithmic (log10 on the y-axis and log2 on the x-axis) so the six curves (tropics and extra-tropics for each of the temporal scales) can be visualised more comfortably within each panel. Given the log-log axes, the relationships *are* linear, offering a straightforward way to interpolate PRs for events occurring at different spatial scales. Regions are defined as falling within the tropics if more than half of the grid cells of which they are comprised fall between 23.5°N and 23.5°S. As expected, PRs in Fig. 3 are above Unity (the dashed horizontal line) for hot events and below Unity for cold events. The PRs for hot extremes over the tropics (denoted as 'T') are greater in CAM5.1 and HadGEM3-A-N216 than MIROC5 by a factor of roughly 5. PRs for cold extremes are more similar between AGCMs, decreasing slightly from HadGEM3-A-N216 to MIROC5 to CAM5.1 and HadAM3P-N96. Because estimates at each scale are based on the same data, confidence intervals on the actual PR value would not provide an accurate indication of confidence intervals on the difference in values between different spatial scales; the difference in values will depend very strongly on the correlation in variability between scales. For this reason, we do not plot confidence intervals because they would be misleading.

We see a clear division between PRs over tropical and extra-tropical regions when attributing hot extremes (Fig. 3a-d). This characteristic can be explained by the fact that temperature variability generally decreases with decreasing latitude: as distance from the equator decreases, the anthropogenic signal tends to emerge more clearly from the noise of natural variability, resulting in a PR tending away from Unity (Angélil et al, 2014b; Harrington et al, 2016). This concept is additionally relevant when values are averaged across space or time, as we are essentially smoothing the noise of natural variability. Thus, in all six panels, PRs tend away from Unity as the spatial or temporal scales of the events increase. The PR is found to have a log-log relationship with spatial scale here.

Although PRs for temperature typically exhibit a smooth transition from weak (a near-unity PR) to strong as distance from the equator decreases, PRs can vary significantly within WRAF regions – in Figs. S2-7 and S11-12 we see larger spread between PRs which occur at small spatial scales. This suggests that the WRAF regions may have a North-South spatial extent large enough (not excluding other possible factors contributing to PRs) to result in a range of PRs – a consequence of noise/seasonality being highly sensitive to distance from the equator. PRs vary less across WRAF regions close to the equator (e.g. red markers for Africa) than those in the extra-tropics, as the change of temperature variability as a function of latitude is low near the equator (see Fig. 4c).

Angélil et al (2016) evaluate the shapes of extreme rainfall and temperature tails in three of the models used in this analyses (CAM5.1, MIROC5, and HadGEM3-N-216). Their results suggest there is substantial tail bias mostly in favour of overly strong attribution statements for one-in-ten-year hot and cold daily extremes, because the simulated tails tend to be shorter than those in reanalyses products, thereby increasing the anthropogenic signal to the noise of natural variability. The exception being attribution statements for hot extremes over North America and parts of Asia, which were found to be biased in favour of being overly weak (Angélil et al, 2016). Extremes in all of the current generation reanalyses used in (Angélil et al, 2016) except for ECMWF Interim Reanalysis (ERA-Interim; Dee et al (2011)) have not yet been thoroughly evaluated against observations. Extremes in ERA-Interim were briefly evaluated against gridded observations over Australia in Angélil et al (2016) and thoroughly in Donat et al

(2014). Of all reanalyses evaluated in Donat et al (2014), ERA-Interim performed best and was therefore a reason to use it in Angélil et al (2016).

The difference between PRs for hot extremes over the tropics and extra-tropics vary depending on the AGCM. The difference is smallest in MIROC5 where PRs over the extra-tropics and tropics are roughly a factor of 4 apart. The gap is larger in HadAM3P-N96 and even larger in CAM5.1 and HadGEM3-A-N216, being roughly an order of magnitude. The inter-model differences are mostly a result of inter-model variations of PRs over the tropics. We further explore reasons for these difference in Fig. 4 by separating internal variability in the AGCMs from mean temperature response to forcings. In both panels statistics are calculated using the pooled runs from each AGCM. Panels (a) and (b) show the difference of zonal mean (land only) temperature between both scenarios (ALL minus NAT) in each AGCM. Panel (a) shows the raw differences, while in panel (b) we divide by the difference in the global mean temperature between both scenarios (ALL minus NAT), which allows us to visualise the sensitivity of mean temperature to anthropogenic forcing at every latitude per degree Kelvin of global warming. In panel (c), curves of the zonal mean standard deviations calculated at the grid-point level with daily data, are plotted for both scenarios in each AGCM.

Panels (a) and (b) suggest that temperature differences are largest at the poles (particularly the north pole in agreement with Stott and Jones (2009)), a phenomenon known as polar amplification. The raw differences (panel (a)) over the tropics are lowest in MIROC5 and similar in CAM5.1, HadGEM3-A-N216, and HadAM3P-N96, which corresponds to the PRs for hot extremes in Fig. 3. Although the sensitivities of tropical temperature to a degree of global warming in CAM5.1 and HadAM3P-N96 are similar to that of MIROC5 (panel (b)), CAM5.1 and HadAM3P-N96 result in PRs more similar to HadGEM3-A-N216 since their global mean temperature differences are $\sim$0.35K greater than that of MIROC5 and HadGEM3-A-N216. Panel (c) suggests that anthropogenic influences on our climate have reduced temperature variability at the poles, but have hardly caused change in variability over the tropics between AGCMs or scenarios. Internal variability is therefore not responsible for the differences in PRs between AGCMs over the tropics in Fig. 3.

There is little reason to average PRs for one-in-one-year rainfall extremes over spatial domains larger than the WRAF scale, because although PRs for rainfall do vary geographically, we do not see a systematic difference between PRs over the tropics and over the extra-tropics (see Figs S8-10 and S13) like we do for temperature extremes. In Fig. 5, for PRs calculated over each of the 7 spatial scales, we average results across each of the 58 WRAF regions – each line representing a different WRAF region. In other words, for one curve, no averaging has been applied to the first marker as the event occurs at the WRAF scale. The value for the second marker is the arithmetic mean of two values as there are two regions within each WRAF region, each occurring at the $\frac{1}{2}$ WRAF scale. Only results from CAM5.1 are shown here (see Fig. S1 for results from all models). To avoid a saturated figure, we separate PRs for daily, 5-day and monthly extremes into individual panels. The colours represent distance from the equator, being the absolute value of arithmetic mean of the latitude of every gridcell within a WRAF region. Regions near the equator are magenta, and those furthest from it are cyan.

As in Fig. 3, PRs tend away from Unity as spatial and temporal scale increases. Regions furthest from the equator tend to be the regions with PRs closest to Unity, while regions closest to the equator have higher and lower PRs. Similar results are seen in the other 3 models (Fig. S1).

Averaging PRs within the spatial domain of a WRAF region, as done in Fig. 5, can result in a loss of useful information. PRs for rainfall can be very sensitive to small scale changes in location – for example neighbouring grid cells can have strikingly different attribution statements as shown in Angélil et al (2014a), as rainfall extremes can be very localised. Angélil et al (2014a) use a bootstrap sampling procedure to show that the difference was not a consequence of noise due to sampling, but rather a dynamical response native to the model. However since models resolve the dynamics at the grid cell scales poorly, PRs for rainfall over individual grid cells are unlikely to be reliable.

Fig. 6 highlights the evolution of the PR and uncertainty around it due to internal variability as the spatial scale changes. The spread of PRs within one WRAF region (northwestern United States; the red curve in the red box in Fig. 5(c)) is shown. Here best estimate PRs from CAM5.1 are shown for wet extremes lasting a month occurring over the whole region and within the region. This is a region of particular interest as PRs are split between being above and below Unity. PRs shown are those before averaging, as in Figs. S2-13. The spread of the raw PRs at the grid cell scale is ~0.4 to ~2.5, however when an average is taken across the WRAF scale for events occurring at each of the 7 spatial scales, PRs lie between ~0.9 and ~1.3 (red curve in the red box in Fig. 5(c)).

Uncertainty due to internal variability on the best estimate (BE) of the PRs are described by their colours, and calculated by generating 10000 bootstrap datasets of the ALL and NAT realisations. Simulations are shuffled, not days, in order to preserve sequencing information. For each dataset the corresponding PR is calculated (on the log scale) per the procedures discussed in the Methods section. This gives a sample of 10000 PR values that characterise the sampling distribution of the PR estimate. To quantify uncertainty in the estimated PR, we used the basic bootstrap confidence interval procedure (not to be confused with the percentile bootstrap confidence interval), by which lower and upper uncertainty bars are calculated by $BE - (E95 - BE)$ and $BE - (E05 - BE)$ respectively, where $E95$ and $E05$ represent the 95th and 5th percentiles of the 10000 bootstrapped PR values (Davison and Hinkley, 1997; Davison and Huser, 2015). With ensemble sizes of 50-100 simulations per scenario, this bootstrap estimate should provide a decent approximation of the uncertainties in the probabilities of exceedance; however, for the 15-member ensembles of HadGEM3-A-N216 this will be a rather poorer estimator. The legend depicts the range of uncertainty for each coloured marker. Uncertainty due to internal variability on average increases with decreasing spatial scale and the higher the PR is – the latter being a sign of the extreme threshold being further out into the tail (Fischer and Knutti, 2015).

Results shown in Figs. 5 and 6 suggest the PR may be sensitive enough to small changes in the exact location of the defined extreme and its spatial scale, to, for example, change the attribution statement from being 'positive' (roughly that anthropogenic influence increased the chance of the event) or 'negative' (roughly that anthropogenic influence decreased the chance of the event), or vice versa. Although there is currently no strict definition of a 'positive' or 'negative' attribution statement (and others National Academies of Sciences, Engineering and Medicine, 2016), studies should properly justify their choice of spatial scale and location for extreme rainfall events.

Relationships between attribution statements for sequential pairs of spatial scales are identified to gauge the reliability of the scaling. We regress 58 PRs (one for each WRAF region, each value being the average of PRs across that WRAF region) against 58 PRs for events occurring at one larger or smaller spatial scale. Figure 7 demonstrates this for 5-day wet extremes in MIROC5. We regress PRs for extremes occurring at the $\frac{1}{2}$ WRAF scale against those occurring at the $\frac{1}{4}$ WRAF scale. The correlation coefficient of 0.93 denotes a strong relationship, and the gradient of less than one (0.73) indicates that PRs on average tend away from Unity as spatial scale increases. The advantage of this method is that the sensitivity of PRs to spatial scale is based on sensitivity within all of the WRAF regions. This means that the resulting regression is also helpful to scale attribution statements for extremes occurring within regions where the average PR across the region is not very sensitive to the spatial scale (Fig. 7).

Correlation coefficients for all combinations of: the AGCM; pairs of spatial scales; temporal scale of the event; and event type, are plotted in Fig. 8, and coefficients to two decimal places can be found in Table S1. All but a few correlation coefficients for hot and cold extremes lie between 0.95 and 1. For wet extremes the coefficients lie between 0.75 and 1. The higher the correlation coefficient, the more reliable the scaling. The high coefficients between the smallest spatial scales may be artefacts of the experimental setup. Because the effective dynamical resolution of a climate model is greater than the resolution it is run at, the variability near and at the grid scale is expected to be under-represented. Reduced variability (noise) increases the strength of the attribution statement (the anthropogenic signal), resulting in higher correlation coefficients with statements for events occurring at slightly larger spatial scales, where this artefact is not as prominent and noise is rather reduced through averaging over space. Caution should therefore be taken when scaling events at near-grid cell spatial scales. Scaling can be performed with the gradients and y-intercepts (for all regressions) found in Table S2.

For example, a PR of 10 (whether it be a statement already published or not) for a 5-day heatwave occurring over the $\frac{1}{2}$ WRAF scale can be scaled to one occurring over the $\frac{1}{4}$ WRAF scale using the following relationship found in CAM5.1: $y = 0.95x + 0.01$. A PR of 9.51 results when $x = 10$. Given the relationships found in MIROC5, HadGEM3-A-N216, and HadAM3P-N96; PRs of 9.59, 9.18, and 9.07 result respectively.

5 Discussion

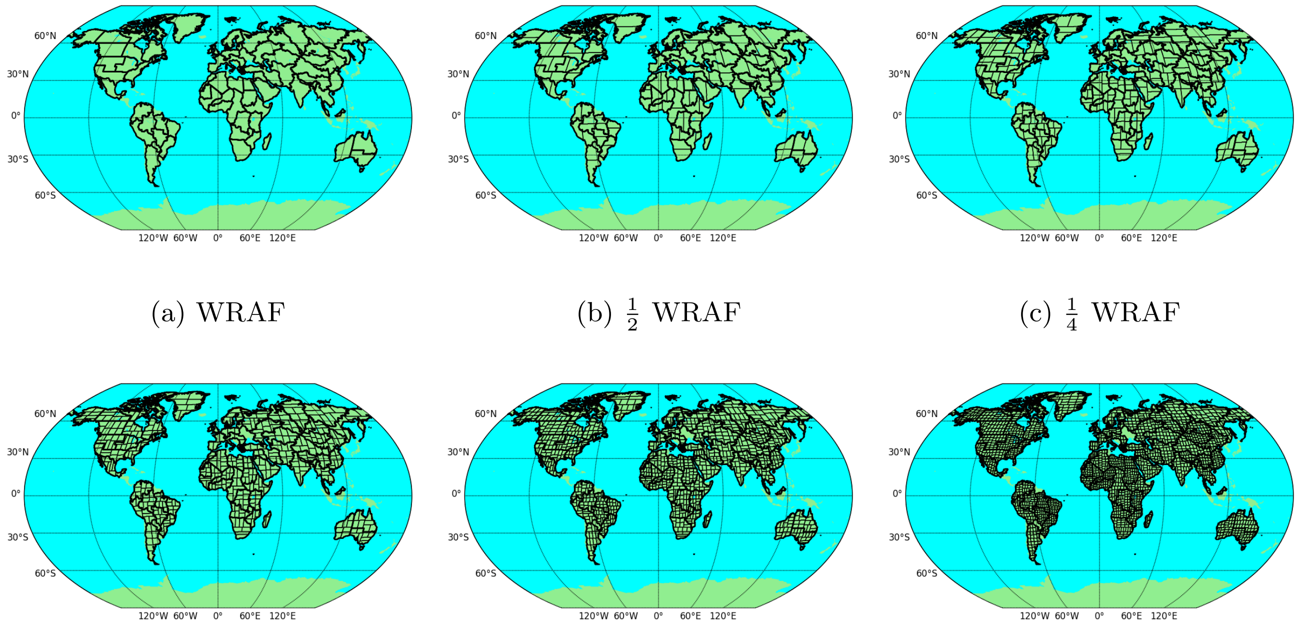
This study characterises functions representing the relation between the spatial scale of the extreme and its attribution statement. Although global mean temperature differences between the NAT and ALL scenarios are $\sim$0.35°K greater in CAM5.1 and HadAM3P-N96 than MIROC5 and HadGEM3-A-N216, zonal mean land temperature difference in the AGCMs hardly correspond to the global response. The response is also highly sensitive to latitude. For example, HadGEM3-A-N216 has a higher temperature sensitivity over the tropics per degree global warming than CAM5.1, MIROC5 and HadAM3P-N96, resulting in comparable attribution statements with CAM5.1 and HadAM3P-N96 for extremes occurring over this region. In essence, it appears that zonal mean absolute temperature differences correspond closely to attribution statements for temperature extremes, suggesting mean temperature is a low order proxy for extreme temperature in agreement with Seneviratne et al (2012). Given the sensitivity of results to the model used, we stress the importance of model evaluation in event attribution studies.

PRs for hot extremes over the extra-tropics were found to lie anywhere between $\sim$5 and $\sim$30 depending on the spatial and temporal scale of the event, and the

AGCM used – the highest from HadGEM3-A-N216 and the lowest from MIROC5. PRs for hot extremes over the tropics were anywhere between $\sim$12 and $\sim$250, the highest PRs coming from both CAM5.1 and HadGEM3-A-N216. For cold extremes, PRs ranged between $\sim$0.05 and $\sim$0.2 over the extra-tropics and between $\sim$0.008 and $\sim$0.15 over the tropics. For the PRs over individual regions within the extra-tropics or tropics, see Figs. S11 & S12. In general, PRs for temperature extremes are less sensitive to variations in the spatial scales of the events than to the AGCM used. PRs for wet events may be similarly sensitive to the AGCM used as to slight changes in the spatial scales (see Figs S8-10 and S13), but further statistical analyses would be required to test this robustly. Although, it is clear that model responses to anthropogenic forcings do not impact PRs for rainfall as directly as it impacts PRs for temperature, which may be due to limited moisture availability over land. Statements do however vary largely between AGCMs in terms of whether they are positive (PR $>$ 1) or negative (PR $<$ 1), as shown in Figs. S8-10 and S13. On average, PRs for wet events are greater than Unity but only marginally, in agreement with Pall et al (2011); Peterson et al (2013); Herring et al (2014, 2015); Fischer and Knutti (2015). In this study we only look at one-in-one year wet extremes. Studies have shown that PRs increase as the anomaly of the wet extreme increases (Angélil et al, 2014a; Fischer and Knutti, 2015), owing to the Clausius-Clapeyron relation – a relation most pertinent to short-lived extreme rainfall (Allen and Ingram, 2002; Christensen and Christensen, 2003; Pall et al, 2007; Jones et al, 2010; Westra et al, 2014), influencing the limit on the most extreme wet event possible as a function of temperature. Warming raises this limit.

Results shown in Figs. 3 and 5 clearly show a nonlinear relationship between the PR and the spatial scale. The correlation coefficients between PRs for temperature extremes occurring at different spatial scales are almost all greater than 0.95 (3 are between 0.9 and 0.95). For rainfall extremes the correlations are all greater than 0.75, although most are greater than 0.9. Such results should encourage the scaling of attribution statements to provide real-time statements for new extremes occurring at different spatial domains. Since PRs between models can vary substantially, there is future work to be done in order to reduce this uncertainty. However since the sensitivity of the PR as a function of the spatial scale is similar between models, scaling could still be performed in the future as model uncertainty is reduced. Furthermore, because real extremes do not have clear-cut physical borders, it is important to understand how attribution results scale as a consequence of uncertainty around the event definition.

In some cases such as the region examined in Fig. 6, PRs tend to be close and on both sides of Unity – in close proximity to thresholds which could categorise an attribution statement as either ‘positive’, ‘neutral’ or ‘negative’. Although no such definitions have yet been established, failure to thoroughly justify the spatial scale and location of an event can result in biased attribution statements, possibly leading to a change in the sign of the statement.



(a) WRAF (b) $\frac{1}{2}$ WRAF (c) $\frac{1}{4}$ WRAF

(d) $\frac{1}{8}$ WRAF (e) $\frac{1}{16}$ WRAF (f) $\frac{1}{32}$ WRAF

Fig. 1 The 1^{st}(a) to 6^{th}(f) spatial scales over which grid cell values are aggregated. Regions derived from the original WRAF regions (panel a and thick black lines in all panels), are demarcated by the thin black lines. The 7^{th} and smallest spatial domain is not shown as it is the grid cell scale. The regions shown here were derived from a high resolution grid of the WRAF regions (1440 x 720), such that the smaller regions could be most accurately defined before remapping them to the resolution of the coarsest model for the analyses.

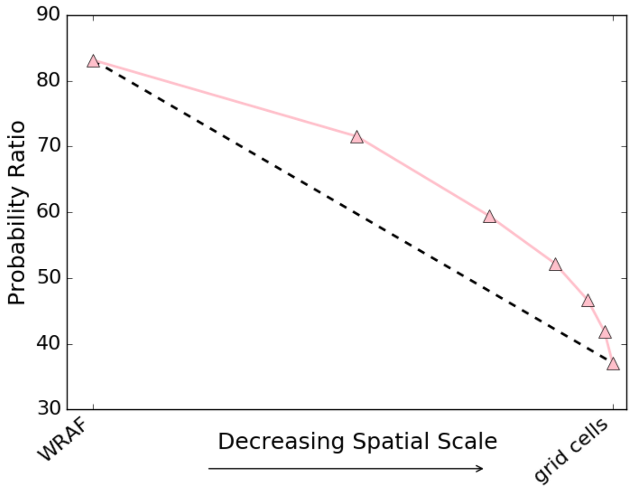
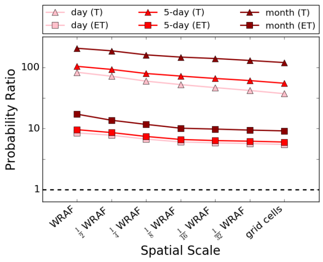
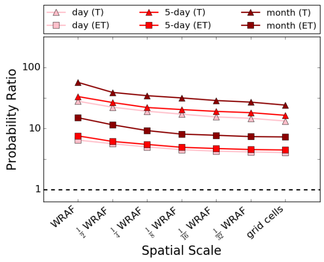


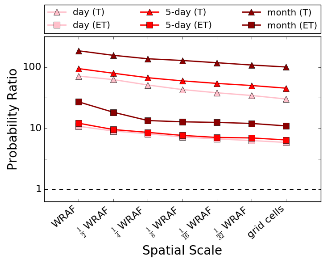
Fig. 2 One-in-ten-year hot day PRs from CAM5.1 as a function of spatial scale. Each of the seven pink markers is the arithmetic mean of PRs for all regions within the tropics, occurring at a given spatial scale. The dashed line represents the relationship had two spatial scales been used, as was performed in Angélil et al (2014b).



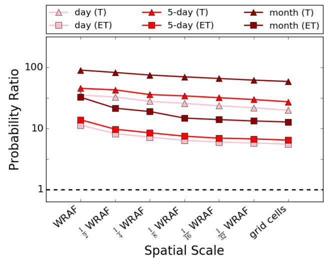
(a) CAM5.1, hot



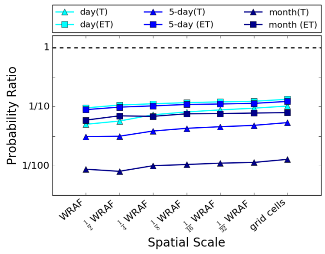
(b) MIROC5, hot



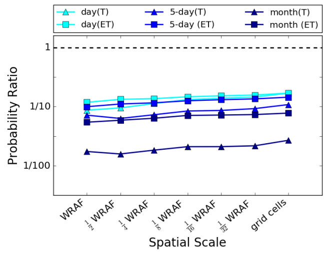
(c) HadGEM3-A-N216, hot



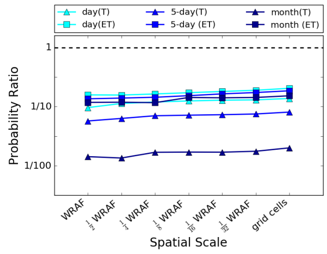
(d) HadAM3P-N96, hot



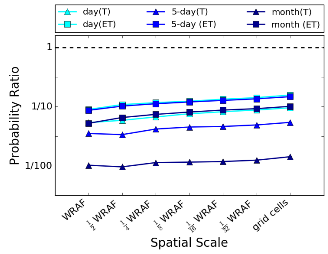
(e) CAM5.1, cold



(f) MIROC5, cold



(g) HadGEM3-A-N216, cold



(h) HadAM3P-N96, cold

Fig. 3 PRs for one-in-ten-year hot (top panels) and cold (bottom panels) extremes, calculated from CAM5.1, MIROC5, HadGEM3-A-N216, and HadAM3P-N96 output. Each marker represents the arithmetic mean of PRs calculated over either tropical ('T') or extra-tropical ('ET') regions, for one-in-ten-year hot and cold extremes occurring at a specified spatial and temporal domain. The dashed black line represents a PR of Unity.

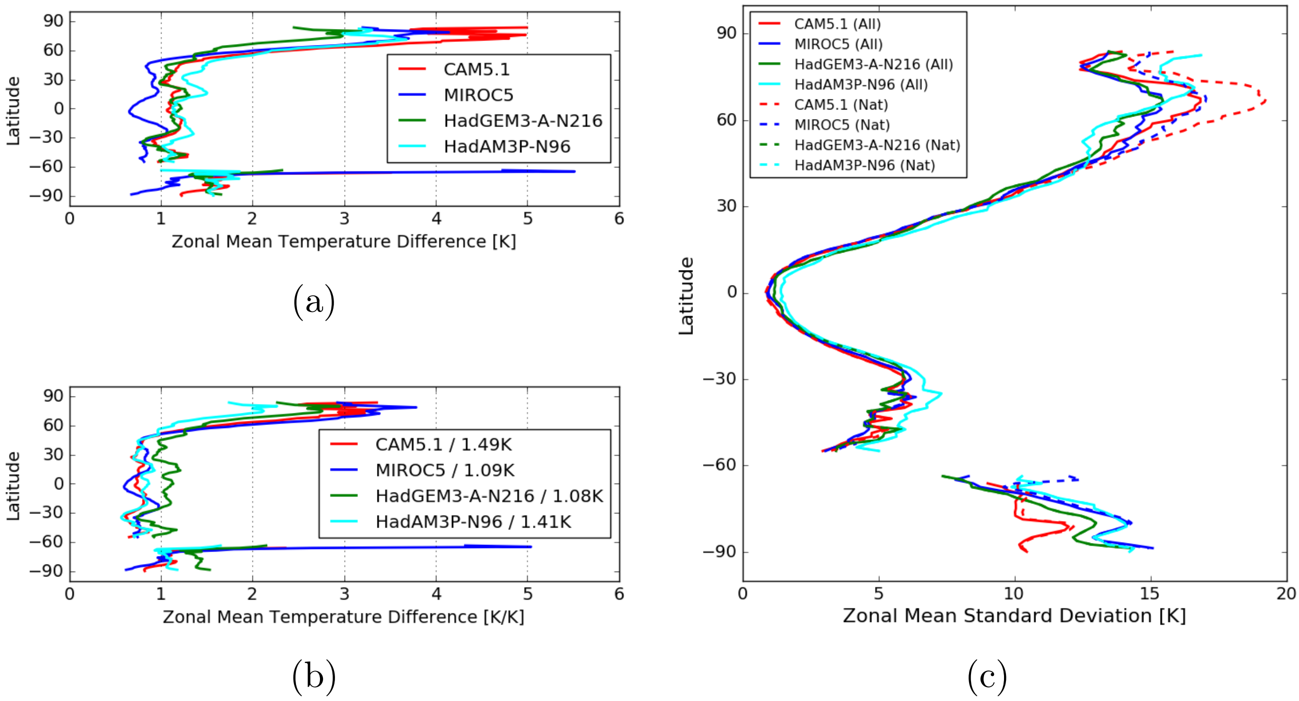
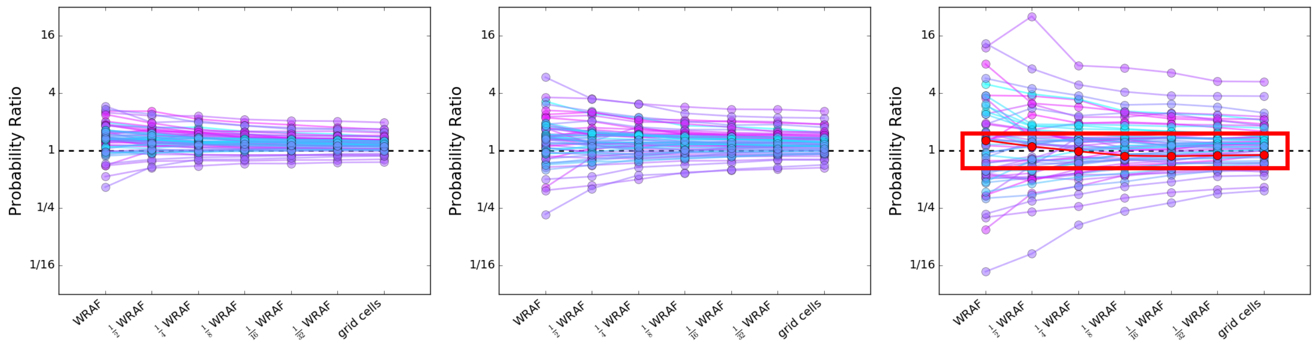


Fig. 4 Zonal mean of land-only temperature (a & b) and land-only standard deviation (c), across all time-steps in all ensemble members in each AGCM. Zonal means in the panel (b) have been divided by the global mean temperature difference (All minus Nat) in each model, to highlight changes per degree Kelvin warming.



(a) CAM5.1 (day) (b) CAM5.1 (5-day) (c) CAM5.1 (month)

Fig. 5 One-in-one-year wet day (first row), 5-day (second row), and month (third row) PRs, from CAM5.1, MIROC5, HadGEM3-A-N16, HadAM3P-N96. Each marker represents the PR for extremes occurring at one of seven different spatial scales, averaged at the WRAF scale. Each line represents a different WRAF region. The dashed black line represents a PR of Unity. The red curve in the red box in panel (c) is examined in more detail in Fig. 6. The colours represent distance from the equator, being the absolute value of arithmetic mean of the latitude of every gridcell within a WRAF region. Regions near the equator are magenta, and those furthest from it are cyan.

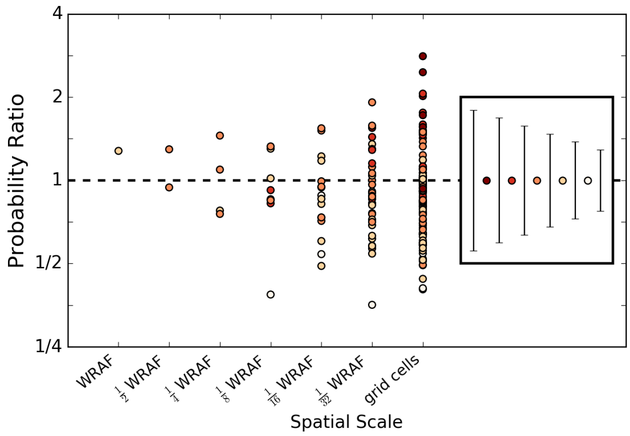


Fig. 6 PRs for wet months in CAM5.1 for the red curve in the red box in figure 5(c), before averaging over space. The WRAF region is Northwestern United States. The dashed black line represents a PR of Unity. The markers can be one of 5 colours, denoting a range of uncertainty due to internal variability around the best estimate. The uncertainty range for each colour appears in the legend, and has been calculated using a Monte Carlo sampling procedure.

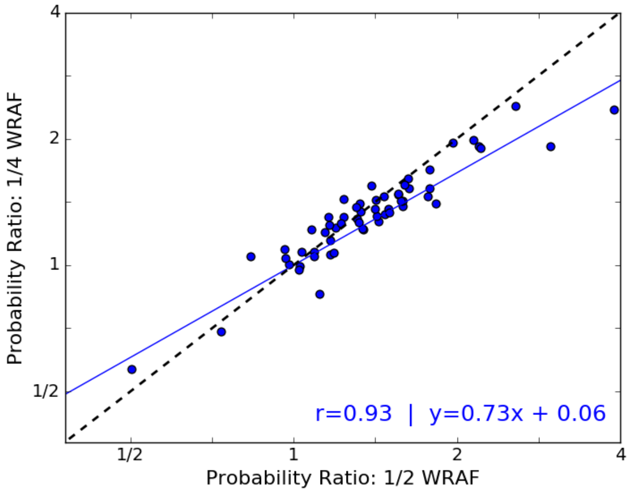
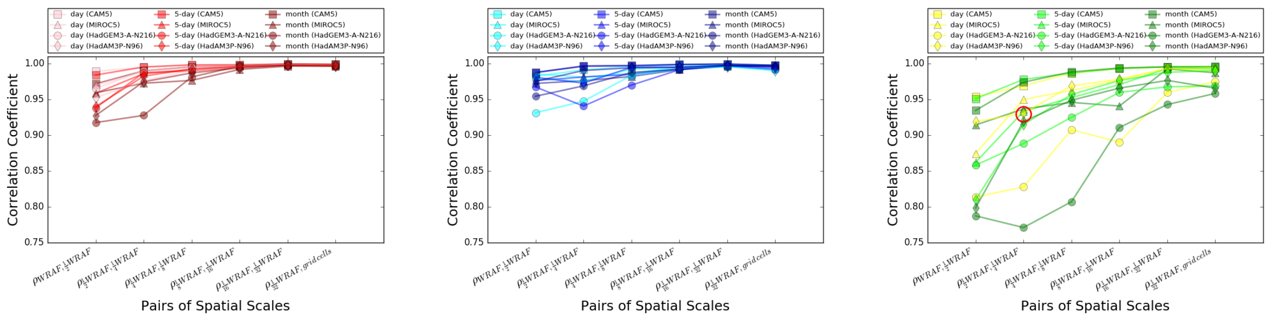


Fig. 7 Regressions between PRs derived at "$\frac{1}{2}$ WRAF" and "$\frac{1}{4}$ WRAF" scales over each of the 58 WRAF regions, for one-in-one-year 5-day wet extremes. The position of each marker is determined by the average of four PRs in a WRAF region (y-value) and the average of two PRs in the same WRAF region (x-value). Data used are from MIROC5.



(a) Hot (b) Cold (c) Wet

Fig. 8 Correlation coefficients between pairs of spatial domains, for day, 5-day, and month-long hot (a), cold (b), and wet (c) events, from CAM5.1, MIROC5, HadGEM3-A-N216, and HadAM3P-N96. For explanatory purposes, the marker encircled in red is the correlation coefficient from Fig. 7

Acknowledgements OA, SP-K, and LVA were supported by Grant CE110001028. In addition SP-K was supported by DECRA grant DE140100952. DS and MW were supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under contract number DE-AC02- 05CH11231. HS was supported by the Program for Risk Information on Climate Change. PW was funded by the RSA National Research Foundation grant number 90964. AC and NC were supported by the Joint UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101) and by the EUCLEIA project funded by the European Unions Seventh Framework Programme [FP7/20072013] under grant agreement number 607085.

References

Allen MR, Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. Nature 419(6903):224–232, DOI 10.1038/nature01092

Angélil O, Stone DA, Pall P (2014a) Attributing the probability of South African weather extremes to anthropogenic greenhouse gas emissions: Spatial characteristics. Geophysical Research Letters 41(9):3238–3243, DOI 10.1002/2014GL059760, URL `http://doi.wiley.com/10.1002/2014GL059760`

Angélil O, Stone DA, Tadross M, Tummon F, Wehner MF, Knutti R (2014b) Attribution of extreme weather to anthropogenic greenhouse gas emissions: Sensitivity to spatial and temporal scales. Geophysical Research Letters 41(6):2150–2155, DOI 10.1002/2014GL059234

Angélil O, Perkins S, Alexander L, Stone D, Donat M, Wehner M, Shiogama H (2016) Comparing regional precipitation and temperature extremes in climate model and reanalysis products. Weather and Climate Extremes 13:In prep., DOI 10.1016/j.wace.2016.07.001, URL `http://dx.doi.org/10.1016/j.wace.2016.07.001`

Bellprat O, Doblas-Reyes F (2016) Unreliable climate simulations overestimate attributable risk of extreme weather and climate events. Geophysical Research Letters pp n/a–n/a, DOI 10.1002/2015GL067189, URL `http://doi.wiley.com/10.1002/2015GL067189`

Christensen JH, Christensen OB (2003) Climate modelling: Severe summertime flooding in Europe. Nature 421(6925):805–806, DOI 10.1038/421805a

Christidis N, Stott Pa, Scaife Aa, Arribas A, Jones GS, Copsey D, Knight JR, Tennant WJ (2013) A New HadGEM3-A-Based System for Attribution of Weather- and Climate-Related Extreme Events. Journal of Climate 26:2756–2783, DOI 10.1175/JCLI-D-12-00169.1

Davison A, Hinkley D (1997) Bootstrap methods and their application. Cambridge university press, URL http://www.citeulike.org/group/17501/article/12121847

Davison A, Huser R (2015) Statistics of Extremes. Annual Review of Statistics and Its Application 2:203–235, DOI 10.1179/003962659792003612, URL http://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-010814-020133

Dee DP, Uppala SM, Simmons aJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda Ma, Balsamo G, Bauer P, Bechtold P, Beljaars aCM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer aJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, Mcnally aP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137(656):553–597, DOI 10.1002/qj.828

Dole R, Hoerling M, Perlwitz J, Eischeid J, Pegion P, Zhang T, Quan XW, Xu T, Murray D (2011) Was there a basis for anticipating the 2010 Russian heat wave? Geophysical Research Letters 38(6):1–6, DOI 10.1029/2010GL046582

Donat MG, Sillmann J, Wild S, Alexander L, Lippmann T, Zwiers FW (2014) Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. Journal of Climate 27(13):5019–5035, DOI 10.1175/JCLI-D-13-00405.1

Fischer EM, Knutti R (2015) Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. Nature Climate Change 5(April):1–6, DOI 10.1038/nclimate2617, URL http://www.nature.com/doifinder/10.1038/nclimate2617

Harrington LJ, Frame DJ, Fischer EM, Hawkins E, Joshi M, Jones CD (2016) Poorest countries experience earlier anthropogenic emergence of daily temperature extremes. Environmental Research Letters 11(5):055,007, DOI 10.1088/

1748-9326/11/5/055007, URL http://stacks.iop.org/1748-9326/11/i=5/a=055007?key=crossref.477f55eb444e412d7b6db31d0ce2788b

Hawkins E, Sutton R (2012) Time of emergence of climate signals. Geophysical Research Letters 39(December 2011):1–7, DOI 10.1029/2011GL050087

Herring SC, Hoerling MP, Peterson TC, Stott Pa (2014) Explaining Extreme Events of 2013 from a Climate Perspective. Bulletin of the American Meteorological Society 95(9):S1–S96

Herring SC, Hoerling MP, Kossin JP, Peterson TC, A SP (2015) Extreme Events of 2014. Bulletin of the American Meteorological Society 96(12)

Hurrell JW, Hack JJ, Shea D, Caron JM, Rosinski J (2008) A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. Journal of Climate 21(19):5145–5153, DOI 10.1175/2008JCLI2292.1

Jones PW (1999) First- and Second-Order Conservative Remapping Schemes for Grids in Spherical Coordinates. Monthly Weather Review 127(9):2204–2210, DOI 10.1175/1520-0493(1999)127⟨2204:FASOCR⟩2.0.CO;2

Jones RH, Westra S, Sharma A (2010) Observed relationships between extreme sub daily precipitation , surface temperature , and relative humidity. Geophysical Research Letters 37(September):1–5, DOI 10.1029/2010GL045081

and others National Academies of Sciences, Engineering, Medicine (2016) Attribution of Extreme Weather Events in the Context of Climate Change. The National Academies Press, DOI 10.17226/21852, URL http://www.nap.edu/catalog/21852

Pall P, Allen MR, Stone DA (2007) Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO2 warming. Climate Dynamics 28(4):351–363, DOI 10.1007/s00382-006-0180-2

Pall P, Aina T, Stone DA, Stott Pa, Nozawa T, Hilberts AGJ, Lohmann D, Allen MR (2011) Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. Nature 470(7334):382–385, DOI 10.1038/nature09762

Peterson TC, Stott Pa, Herring SC (2012) Explaining Extreme Events of 2011 from a Climate Perspective. Bulletin of the American Meteorological Society 93(7):1041–1067, DOI 10.1175/BAMS-D-12-00021.1

Peterson TC, Hoerling MP, Stott Pa, Herring SC (2013) Explaining Extreme Events of 2012 from a Climate Perspective. Bulletin of the American Meteorological Society 94(9):S1–S74

Rayner N, Parker D, Horton E, Folland C, Alexander L, Rowell D, Kent E, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late Nineteenth Century. Journal of Geophysical Research: Atmospheres 108(D14), DOI 10.1029/2002JD002670, URL `http://dx.doi.org/10.1029/2002JD002670`

Robine JM, Cheung SLK, Le Roy S, Van Oyen H, Griffiths C, Michel JP, Herrmann FR (2008) Death toll exceeded 70,000 in Europe during the summer of 2003. Comptes Rendus - Biologies 331(2):171–178, DOI 10.1016/j.crvi.2007.12.001

Seneviratne S, Nicholls N, Easterling DR, Goodess C, Kanae S, Kossin J, Luo Y, Marengo J, McInnes K, Rahimi M, Reichstein M, Sorteberg A, Vera C, Zhang X (2012) Changes in climate extremes and their impacts on the natural physical environment. Managing the Risk of Extreme Events and Disasters to Advance Climate Change Adaptation A Special Report of Working Groups I and II of the IPCC, Annex IIanaging the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation pp 109–230

Shiogama H, Watanabe M, Imada Y, Mori M, Ishii M, Kimoto M (2013) An event attribution of the 2010 drought in the South Amazon region using the MIROC5 model. Atmospheric Science Letters 14(May):170–175, DOI 10.1002/asl2.435

Shiogama H, Watanabe M, Imada Y, Mori M, Kamae Y, Ishii M, Kimoto M (2014) Attribution of the June-July 2013 Heat Wave in the Southwestern United States. Sola 10(0):122–126, DOI 10.2151/sola.2014-025, URL `http://jlc.jst.go.jp/DN/JST.JSTAGE/sola/2014-025?lang=en{\&}from=CrossRef{\&}type=abstract`

Stott PA, Jones GS (2009) Variability of high latitude amplification of anthropogenic warming. Geophysical Research Letters 36(10):1–7, DOI 10.1029/2009GL037698

Stott PA, Stone DA, Allen MR (2004) Human contribution to the European heatwave of 2003. Nature 432(7017):610–4, DOI 10.1029/2001JB001029, URL `http://dx.doi.org/10.1038/nature03089`

Westra S, Fowler H, Evans J, Alexander LV, Berg P, Johnson F, Kendon E, Lenderink G, Roberts N (2014) Future changes to the intensity and frequency of short-duration extreme rainfall. Rev Geophys 52:522–555, DOI 10.1002/2014RG000464