Early Benefits of Mitigation in Risk of Regional Climate Extremes

Andrew Ciavarella*[1], Peter Stott[1] and Jason Lowe[1]

[1]Met Office Hadley Centre, FitzRoy Road, Exeter, EX1 3PB, UK

Large differences in climate outcomes are projected by the end of this century depending on whether greenhouse gas (GHG) emissions continue to increase or are reduced sufficiently to limit total warming to below 2°C.[1] However it is generally thought that benefits of mitigation are hidden by internal climate variability until later in the century.[2] Here we show that if the likelihood of extremely hot seasons is considered, the benefits of mitigation emerge more quickly than previously thought. It takes less than twenty years of emissions reductions in many regions for the likelihood of extreme seasonal warmth to reduce by more than half following initiation of mitigation. Additionally we show that the latest possible date at which the probability of extreme seasonal temperatures will be halved through emissions reductions consistent with the 2°C target is in the 2040s. Exposure to climate risk is therefore reduced markedly and rapidly with substantial reductions of GHG emissions demonstrating that the early mitigation needed to limit eventual warming below potentially dangerous levels benefits societies in the nearer-term not just in the longer-term future.

Future changes in climate depend in part upon emissions to date (so-called climate change commitment) and emissions over the coming decades with an approximately linear relationship between total anthropogenic CO_2 emissions and overall warming.[3] The impacts on societies around the globe will depend upon regional exposure and vulnerability to the hazards posed by a changing climate, particularly those related to changing climate extremes.

*andrew.ciavarella@metoffice.gov.uk

Coupled climate model simulations together with observations and fundamental physical understanding form the basis for assessment of past and future changes in climate. The generation of models supporting the IPCC's 5th Assessment Report is phase 5 of the Coupled Model Intercomparison Project (CMIP5)[4] comprising simulations of 20thC historical climate and scenarios of 21stC climate driven by emissions prescribed in a set of Representative Concentration Pathways (RCPs).[5] These range from an aggressive mitigation scenario, consistent with the scale of emissions needed to likely limit warming to 2°C above pre-industrial levels (RCP 2.6), to a scenario in which emissions continue to increase unchecked by climate mitigation policy (RCP 8.5[i]).

Indicators of key interest are those such as time of emergence (ToE) at which climate changes in a given emissions scenario become detectable above some baseline variability.[6] Tropical warming leads to emergence of a regional signal earlier than higher latitudes where, despite greater magnitude of expected warming, greater variability and lower signal-to-noise retard ToE by several decades. ToE is sensitive to choice of variable and definition of signal, however. Broadly speaking temperature studies concerning larger spatial scales, earlier baselines and tests for change of the full temperature distribution find emergence has already occurred or is soon to do so[7,8,9] while later baselines and signal-to-noise type metrics show later emergence.[6,10]

The ToE approach has been adapted to examine the emergence of climate changes between two future emissions scenarios to address the benefits of emissions mitigation. Tebaldi and Friedlingstein[11] used CMIP5 models to find dates at which a regionally significant difference in trend in seasonal warming emerges between the aggressive mitigation scenario RCP 2.6 and the high emissions scenario RCP 8.5. This date is not until the middle decades of this century in most continental and sub-continental scale regions of the world. The conclusion was drawn that detection of the benefits of mitigation is thus delayed to this time.

The most obvious adverse impacts of climate change however are likely to be felt initially though society's vulnerabilities to locally rare "extreme events".[12] Furthermore, extreme events defined by some fixed threshold may see dramatic changes in likelihood upon only small changes in the mean state of the climate.[9,13,14,15] For example, if a climate variable with normally distributed annual values sees a positive shift in its mean of 0.3 standard deviations (a signal to noise of much less than one) the probability of fixed threshold events originally at a 1 in 20 year return level approximately doubles to 1 in 10 years. For the same shift in the mean, extremes at larger return

[i]The scenarios are named according to the resulting median total radiative forcing estimated at 2100.

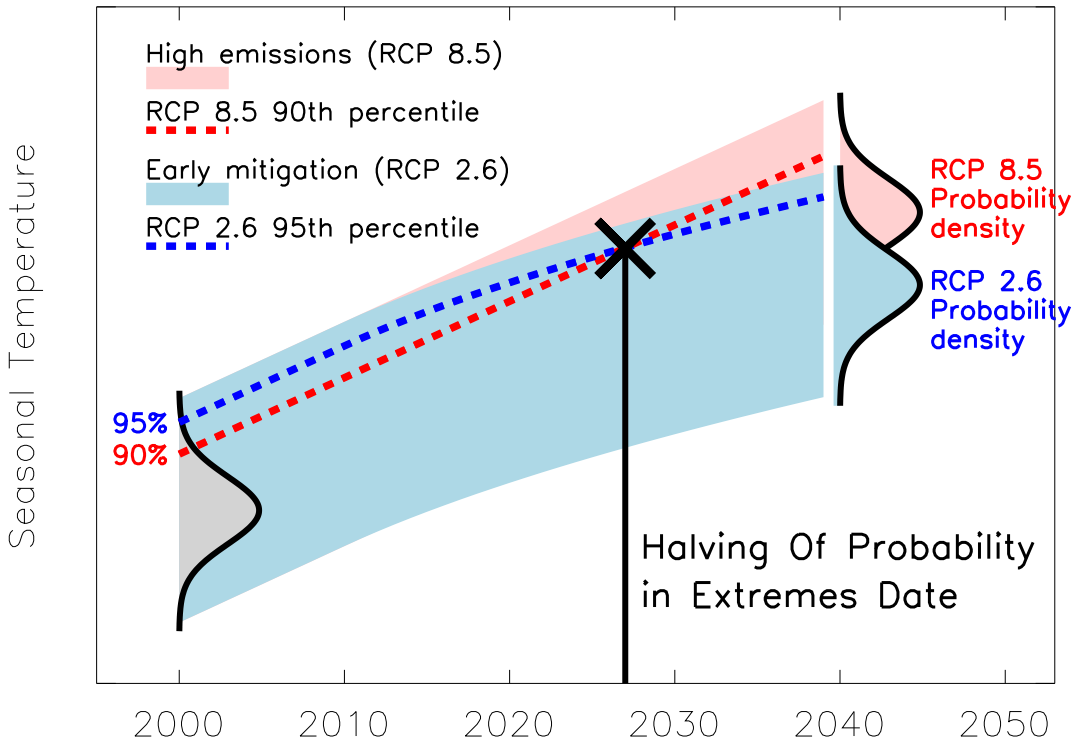

Figure 1: Definition of the Halving of Probability in Extremes Date (HOPE Date). Depicted are idealised non-stationary probability distributions of modelled temperatures and relevant percentiles with time. Distribution widths include contributions to the likelihood from both internal climate variability as well as inter-model uncertainty in response to climate forcings. The grey probability distribution represents a time when emissions scenarios are indistinguishable. Into the 21stC the distribution of seasonal temperatures in a region diverges under different emissions scenarios. The HOPE Date gives an indication of when the likelihood of extreme seasonal warmth in an emissions mitigated world is half that of the same warmth in the unmitigated world.

times become more likely by increasingly large factors. This demonstrates the importance of changes in the expected frequency of extremes before traditional detection is possible and adds further penalties to delayed emissions reductions.[16]

This study investigates how quickly benefits of mitigation are realised through reductions in the risk of extreme temperatures over land. We calculate dates when the modelled likelihood of regional extreme seasonal warmth in an aggressive mitigation scenario (RCP 2.6) has substantially decreased compared to its likelihood in an unmitigated scenario (RCP 8.5) as represented through a multi-model ensemble (MME). We focus on events at 1 in 10 year return levels in the unmitigated world and calculate the date when they become half as likely, i.e. representing 1 in 20 year return levels with early mitigation. We call this date the Halving Of Probability in Extremes Date (HOPE Date), illustrated in Figure 1.

Adopting the event attribution paradigm[17] the HOPE Date is the date in a high emissions

world beyond which more than 50% of the likelihood of extremes is attributable to the unmitigated component of emissions.

Using the 26 SREX regions[12] (Supplementary Information (SI) Figure 1) we focus on local warm seasons, being those in which warm extremes are likely of greatest impact. Despite the relatively large spatial and temporal scale and associated averaging out of more localised, shorter duration extremes, such regions and periods are routinely of interest in attribution studies concerning extreme climate events and their impacts.[18,19]

The HOPE Date is the time at which the 90^{th} percentile of the distribution of temperatures under RCP 8.5 is expected to be equal to the 95^{th} percentile under RCP 2.6. With 27 models simulating RCP 2.6 and 8.5 (see SI Table 1) we obtain 58 and 64 member ensembles respectively. The time evolution of percentiles is therefore subject to noise due solely to finite MME size, so Figure 1 is idealised. We largely remove this noise by calculating HOPE Dates from 15-year running means of the raw percentiles; different smoothing periods do not significantly affect the results (SI Figure 2).

We address model uncertainty by bootstrapping the MME: for every model included we recalculate the HOPE Date with all members of this single model removed to produce a range of HOPE Date estimates. We refer to the estimate which includes all available models as the best estimate although we stress the importance of the resulting range.

We conduct validation tests on the MME (see Methods), failed by regions where the observations simply do not look like a member of the MME. First we ask for a statistically significant mean correlation between 15-year smoothed observations and MME members over 1905 - 2004 as a check on the timing and sign of response to external forcings. Secondly total time series variance is tested as a check on the consistency of the amplitude of variations (on which our first test is silent) by asking that the observed variance lie within the MME range over the same period. 5 of 26 SREX regions (CNA, WAF, CAS, SAS & SAU) fail the first test and two fail the second (AMZ & SAS) leaving 20 regions whose representation of warm seasons by the CMIP5 MME is adequate for the purposes of our analysis.

Figure 2 (left hand panels) depicts warm season HOPE Dates calculated for a small selection of regions, repeated for all regions in SI Figure 3. Best estimates are used in panel (a) of Figure 3 and column 3 of Table 1. HOPE Dates are found to occur no later than the early 2040s with ranges spanning less than a decade in most regions. Half of SREX regions passing the

validation tests see best estimate HOPE Date by 2030. HOPE Dates thus occur considerably sooner (one to two decades) than those typical of RCP-derived mitigation efforts based on a detectable difference in scenario trends.[11] The spatial pattern is nevertheless consistent (given quoted uncertainty and for comparable regions and seasons) whereby regions with higher signal to noise in seasonal temperatures emerge first. We would therefore expect HOPE Dates for sub-seasonal temperatures such as dailies and variables with lower signal to noise such as at the grid box scale to be considerably later.

RCP inter-scenario emissions diverge significantly from 2010[5] while to date real world emissions have followed a trajectory much closer to the unmitigated scenario.[20,21] Therefore such reductions in risk are unlikely to be realisable in practice as early as the RCP 2.6-based HOPE Dates so we next consider replacing RCP 2.6 with delayed mitigation scenarios.

It is desirable for delayed mitigation scenarios to warm no more than the existing scenario, RCP 2.6. We therefore calculate latest possible HOPE Dates for hypothetical scenarios with delayed emissions reductions where regional temperatures may experience small overshoots but are constrained to not exceed their peak values under RCP 2.6. This constraint together with the fixed RCP 8.5 reference determines a latest date we may find the HOPE Date, which is hence defined as the date when the smoothed 90^{th} percentile of RCP 8.5 is equal to the peak value of the smoothed 95^{th} percentile of RCP 2.6, as demonstrated in Figure 2 (b), (d), (f) and (h) and SI Figure 4. The results are therefore independent of the detailed mitigation trajectory and avoid scenarios with larger overshoots which Lowe et al.[22] show require larger mitigation efforts to return to lower temperatures within century timescales.

Correspondingly in Figure 3 (a) we first display the duration between the onset of dramatic emissions reductions (i.e. 2010 in RCP 2.6) and the HOPE Date for all regions passing validation along side the latest possible HOPE Dates in Figure 3 (b), which fall around a decade later than the RCP 2.6-based HOPE Dates themselves (see also Table 1).

In panel (d) of Figure 3 we show values of the difference in global annual mean land temperatures between scenarios, $\Delta T_{\text{land}}^{\text{Global}}$, at the time of the RCP 2.6-derived HOPE Dates as another measure that should possess some scenario independence. Again, the spatial pattern of $\Delta T_{\text{land}}^{\text{Global}}$ compares well with the pattern of committed global (land + ocean) warming resulting in emergence of local signals above pre-industrial variability found by Mahlstein et al.[7] where 1°C of global mean annual warming was sufficient to obtain emergence in most countries.

Along similar lines a study by Tebaldi et al.[23] finds $\Delta T^{\mathrm{Global}}_{\mathrm{ocean+land}} = 0.4°\mathrm{C}$ sufficient for emergence of signals in median 20-year smoothed seasonal mean temperatures over half of the land surface. Here we find that by $\Delta T^{\mathrm{Global}}_{\mathrm{land}} = 0.3°\mathrm{C}$ between scenarios most regions (15 of 20 passing validation) have already experienced a halving in the risk of warm season extremes. As a land-sea contrast $\Delta T^{\mathrm{Global}}_{\mathrm{land}} / \Delta T^{\mathrm{Global}}_{\mathrm{land+ocean}} > 1$ is a robust feature of simulated and observed warming we conclude that $\Delta T^{\mathrm{Global}}_{\mathrm{land+ocean}} < 0.3°\mathrm{C}$ is required for most regions to experience a halving of the probability in warm season extremes.

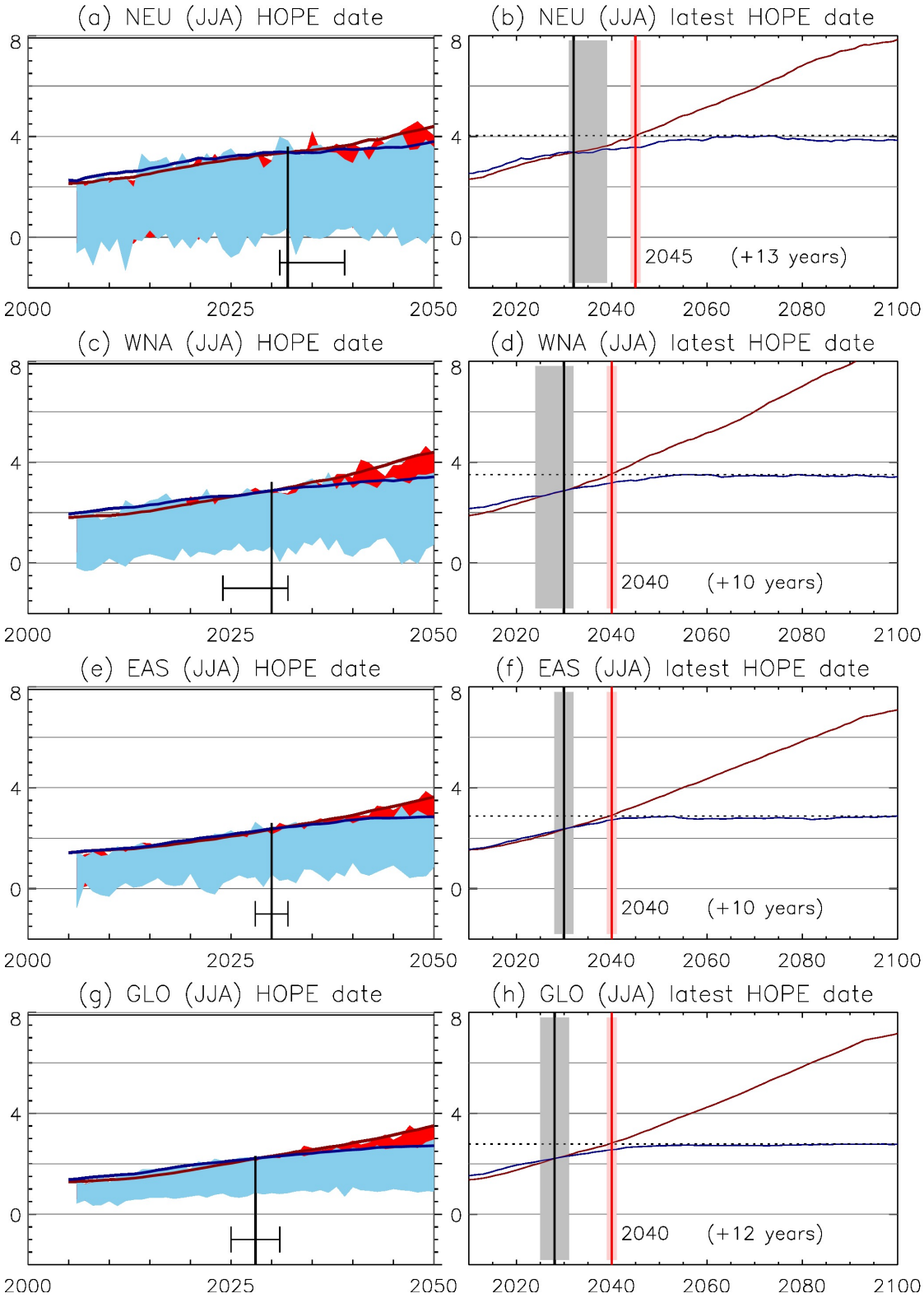

Figure 2: HOPE dates for four selected regions. Left panels show HOPE dates calculated from distributions of regional warm season temperature anomalies (degrees Kelvin with respect to '61-'90 baseline) displayed over 2005 - 2050 from 5 - 95% under RCP 2.6 (early mitigation, light blue) and from 10 - 90% under RCP 8.5 (non-mitigation, red). HOPE Dates are depicted by vertical (best estimate) and horizontal (range) black lines. Right panels show latest possible HOPE dates as determined by the peak value of the 15-year smoothed RCP 2.6 95th percentile, with year stated and delay in years from best estimate HOPED Date in parenthesis. Best estimate HOPE dates are vertical black lines with range in grey, latest possible HOPE dates are vertical red lines with range in pale red. Also plotted in both left and right panels are the relevant 15-year smoothed percentiles (dark blue and dark red correspondingly).

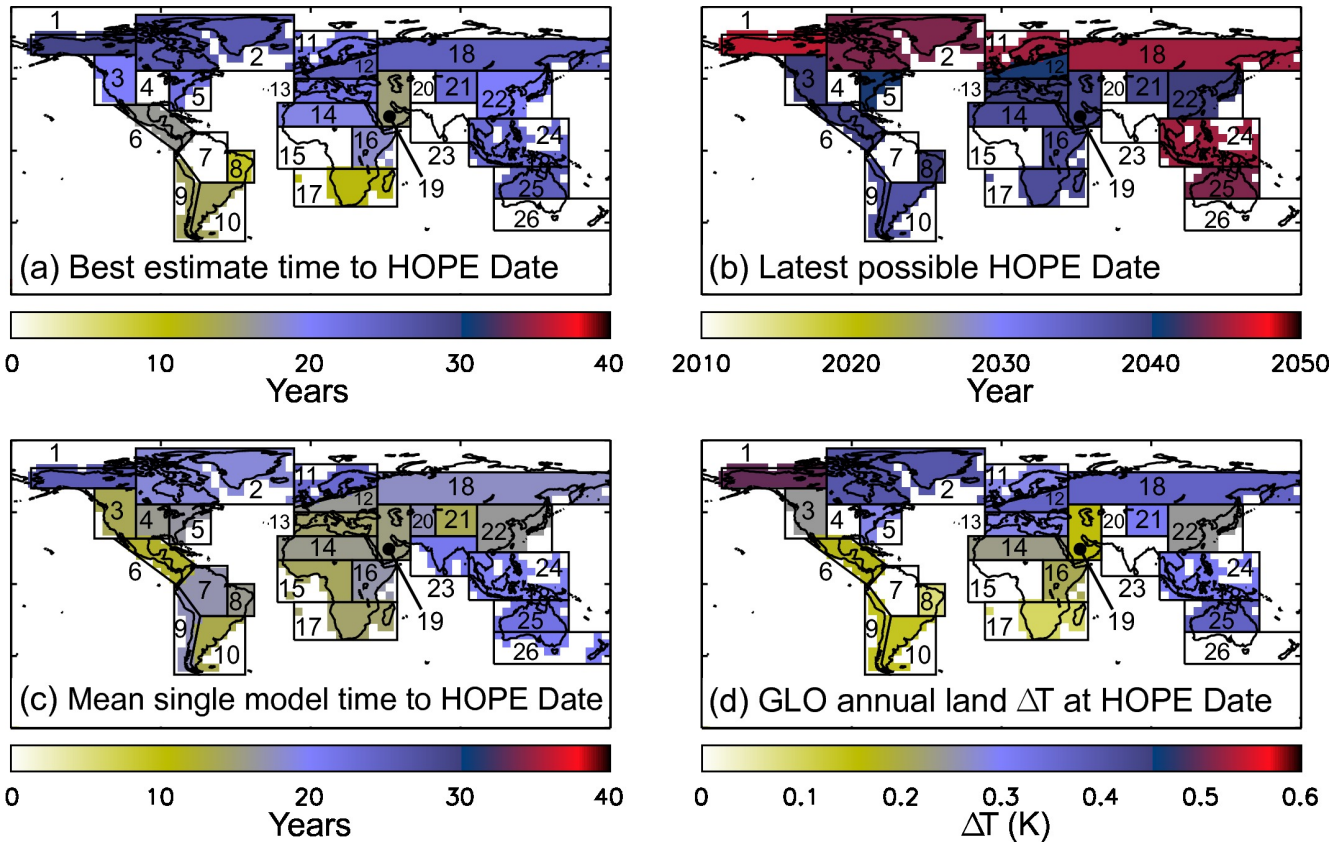

Figure 3: The early benefits of mitigation as represented by regional HOPE Dates. (a) multi-model ensemble (MME) best estimate time between the onset of dramatic emissions reductions (which in RCP 2.6 is 2010) and the HOPE Date when the likelihood of extremes is halved, (b) latest possible HOPE Date itself with onset of emissions reductions delayed from the present and temperatures not exceeding those in the RCP 2.6 scenario, and (c) the mean across single model estimates of the time between the onset of dramatic emissions reductions and the HOPE Date. (d) displays the difference in global mean annual land temperature $\Delta T_{\text{land}}^{\text{Global}}$ between emissions scenario median values at the time of the RCP 2.6 based HOPE Dates implied in panel (a). Regions 4, 7, 15, 20, 23 & 26 fail the MME validation while the remaining 20 regions are adequately represented by the MME for the purposes of the study. HOPE Date ranges can span a couple of years to a decade while ranges of single model estimates can be larger but with the spatial pattern showing broad similarities: an axis north-west to south-east of later dates, including Europe, while Central & South America through Africa and the middle east tend to have earlier dates. Refer to Table 1 and SI Figures 3 - 5 for precise dates and HOPE Date ranges.

We additionally investigate model sensitivity by calculating estimates of HOPE Dates from a selection of individual models (offering RCP ensemble sizes ≥ 3, see SI), hence removing model response differences from the probability distributions. Figure 3 (c) shows mean time to HOPE Dates across models (SI Figure 5 shows the full range). The HOPE Dates thus produced are consistent with the multi-model values in all regions although the full range of individual model dates is often as large as two decades. Individual model dates tend to be sooner than MME best estimates implying that future constraints on warming rates in both scenarios, which effectively removes ensemble spread due to model response differences, could mean regional HOPE Dates come sooner than the MME best estimates.

Based on this analysis we find that benefits of mitigation through a substantial reduction in the likelihood of extreme seasonal warmth could be felt at regional scales across the world within three decades, even with rapid emissions reductions delayed until 2020 (representing a 10 year delay from the RCP 2.6 scenario). In addition to well known long term benefits provided by early emissions reductions, for example to keep global mean warming below 2°C, our work demonstrates that early mitigation provides short term benefits by reducing exposure to the risks of regional climate extremes.

Region number: name (code)	Warm season	Years until HOPE Date	Latest possible HOPE Date	Validation failure?
8: N. E. Brazil (NEB)	SON	9 [7, 15]	2040 [2039, 2041]	
17: S. Africa (SAF)	DJF	11 [11, 15]	2038 [2037, 2039]	
9: W. Coast South America (WSA)	DJF	14 [8, 17]	2037 [2035, 2038]	
10: S. E. South America (SSA)	DJF	14 [14, 16]	2037 [2036, 2037]	
19: W. Asia (WAS)	JJA	15 [15, 19]	2039 [2038, 2039]	
6: Central America / Mexico (CAM)	JJA	16 [13, 17]	2039 [2037, 2040]	
16: E. Africa (EAF)	MAM	18 [18, 19]	2038 [2035, 2039]	
14: Sahara (SAH)	JJA	19 [16, 20]	2038 [2036, 2040]	
3: W. North America (WNA)	JJA	20 [14, 22]	2040 [2039, 2041]	
22: E. Asia (EAS)	JJA	20 [18, 22]	2040 [2039, 2041]	
5: E. North America (ENA)	JJA	22 [22, 24]	2041 [2040, 2043]	
11: N. Europe (NEU)	JJA	22 [21, 29]	2045 [2044, 2046]	
13: S. Europe / Mediterranean (MED)	JJA	22 [18, 24]	2040 [2040, 2041]	
21: Tibetan Plateau (TIB)	JJA	23 [11, 25]	2039 [2038, 2040]	
24: S. E. Asia (SEA)	MAM	23 [17, 23]	2045 [2044, 2046]	
12: Central Europe (CEU)	JJA	25 [24, 26]	2041 [2041, 2042]	
18: N. Asia (NAS)	JJA	25 [21, 27]	2045 [2044, 2046]	
25: N. Australia (NAU)	DJF	25 [24, 25]	2044 [2043, 2046]	
2: E. Canada / Greenland / Iceland (CGI)	JJA	26 [26, 28]	2044 [2041, 2045]	
1: Alaska / N. W. Canada (ALA)	JJA	29 [26, 29]	2046 [2045, 2047]	
Validation failures				
20: Central Asia (CAS)	JJA	13 [10, 14]	2040 [2037, 2040]	Test 1
26: S. Australia / New Zealand (SAU)	DJF	17 [17, 21]	2040 [2039, 2040]	Test 1
15: W. Africa (WAF)	MAM	18 [12, 19]	2040 [2038, 2041]	Test 1
23: S. Asia (SAS)	JJA	20 [18, 22]	2042 [2039, 2043]	Test 1 & 2
4: Central N. America (CNA)	JJA	23 [22, 23]	2041 [2039, 2041]	Test 1
7: Amazon (AMZ)	SON	30 [19, 31]	2048 [2043, 2049]	Test 2
Continental regions				
31: Africa (regs 14 to 17) (AFR)	MAM	12 [12, 14]	2039 [2037, 2040]	
29: S. America (regs 7 to 10) (SAM)	DJF	17 [14, 18]	2039 [2038, 2040]	
27: Global (all regions) (GLO)	JJA	18 [15, 21]	2040 [2039, 2041]	
32: Asia (regs 18 to 23) (ASI)	JJA	19 [18, 23]	2039 [2037, 2041]	
33: Australasia (regs 24, 25, 26) (AUS)	DJF	19 [17, 19]	2040 [2039, 2042]	
28: N. America (regs 1 to 6) (NAM)	JJA	21 [18, 23]	2040 [2039, 2041]	
30: Europe (regs 11, 12, 13) (EUR)	JJA	22 [21, 25]	2041 [2041, 2042]	

Table 1: Detailed dates, ranges and warm seasons for the regions used in the study. Column three lists the duration in years until the HOPE Date (with model bootstrap ranges described in the text) from the onset of aggressive emissions reductions, which in the mitigation simulations used was in 2010. The fourth column lists the latest possible warm season HOPE Date itself under hypothetical scenarios with emissions reductions delayed beyond the present and temperatures constrained to not exceed their peak under the RCP 2.6 scenario. Validation tests are detailed in the methods and an empty fifth column entry implies both tests were passed.

References

[1] Collins, M. *et al.* Long-term climate change: projections, commitments and irreversibility (2013).

[2] Kirtman, B. *et al.* Near-term climate change: projections and predictability. *Climate change* 953–1028 (2013).

[3] Allen, M. R. *et al.* Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).

[4] Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of cmip5 and the experiment design. *Bulletin of the American Meteorological Society* **93**, 485–498 (2012).

[5] Meinshausen, M. *et al.* The rcp greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**, 213–241 (2011).

[6] Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophysical Research Letters* **39** (2012).

[7] Mahlstein, I., Knutti, R., Solomon, S. & Portmann, R. Early onset of significant local warming in low latitude countries. *Environmental Research Letters* **6**, 034009 (2011).

[8] Irina Mahlstein, Gabriele Hegerl, and Susan Solomon. Emerging local warming signals in observational data. *Geophysical Research Letters*, 39(21), 2012.

[9] Andrew D King, Markus G Donat, Erich M Fischer, Ed Hawkins, Lisa V Alexander, David J Karoly, Andrea J Dittus, Sophie C Lewis, and Sarah E Perkins. The timing of anthropogenic emergence in simulated climate extremes. *Environmental Research Letters*, 10(9):094015, 2015.

[10] Diffenbaugh, N. S. & Scherer, M. Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Climatic Change* **107**, 615–624 (2011).

[11] Tebaldi, C. & Friedlingstein, P. Delayed detection of climate mitigation benefits due to climate inertia and variability. *Proceedings of the National Academy of Sciences* **110**, 17229–17234 (2013).

[12] Field, C. B. *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change* (Cambridge University Press, 2012).

[13] Christidis, N., Jones, G. S. & Stott, P. A. Dramatically increasing chance of extremely hot summers since the 2003 european heatwave. *Nature Climate Change* (2014).

[14] Russo, S., Sillmann, J. & Fischer, E. M. Top ten european heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters* **10**, 124003 (2015).

[15] Sanderson, B. M., Oleson, K. W., Strand, W. G., Lehner, F. & ONeill, B. C. A new ensemble of gcm simulations to assess avoided impacts in a climate mitigation scenario. *Climatic Change* 1–16 (2015).

[16] Allen, M. R. & Stocker, T. F. Impact of delay in reducing carbon dioxide emissions. *Nature Climate Change* **4**, 23–26 (2014).

[17] Allen, M. Liability for climate change. *Nature* **421**, 891–892 (2003).

[18] Herring, S. C., Hoerling, M. P., Peterson, T. C. & Stott, P. A. Explaining extreme events of 2013 from a climate perspective. *Bulletin of the American Meteorological Society* **95**, S1–S104 (2014).

[19] Herring, S. C., Hoerling, M. P., Kossin, J. P., Peterson, T. C. & Stott, P. A. Explaining extreme events of 2014 from a climate perspective. *Bulletin of the American Meteorological Society* **96**, S1–S172 (2015).

[20] Peters, G. P. *et al.* The challenge to keep global warming below 2 c. *Nature Climate Change* **3**, 4–6 (2013).

[21] Friedlingstein, P. *et al.* Persistent growth of co2 emissions and implications for reaching climate targets. *Nature geoscience* **7**, 709–715 (2014).

[22] JA Lowe, C Huntingford, SCB Raper, CD Jones, SK Liddicoat, and LK Gohar. How difficult is it to recover from dangerous levels of global warming? *Environmental Research Letters*, 4(1):014012, 2009.

[23] Claudia Tebaldi, Brian O'Neill, and Jean-Franois Lamarque. Sensitivity of regional climate to global temperature and forcing. *Environmental Research Letters*, 10(7):074001, 2015.

Corresponding author

Correspondence and requests for materials associated with this article should be directed to Andrew Ciavarella.

Acknowledgements

This work was 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 no. 607085. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in SI Table 1) for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author contributions

AC conducted the research and wrote the first draft of the paper. PS initially suggested a study concerning changing extremes under mitigation. Both PS and JL contributed to discussion of the results and revisions of the paper.

Competing financial interests

The authors declare no competing financial interests.

Methods

The full CMIP5 archive provides 60 climate models from over 20 institutions contributing between 1 and 10 simulations per experiment type, including the 20^{th}C historical experiment and future scenarios RCP 2.6 and RCP 8.5. We use 27 of these models taking simulations differing only through their initial conditions. See SI Table 1 for the full model list and number of members used per experiment.

To arrive at the two percentile time series from which the HOPE Date is found we require MME cumulative probability distribution functions (CDFs) at each year, found as follows.

Monthly mean fields are first converted to anomalies with respect to 1961 - 1990 at each cell by subtracting the climatological month mean calculated from that ensemble member. Fields are masked using the native resolution model land fraction mask (in three cases where the mask was unavailable that of another model of identical horizontal resolution), removing data in cells with land fraction < 0.25. We then regrid to $5^\circ \times 5^\circ$ resolution before masking with the spatiotemporal observational coverage of CRUTEM4,[24] which after the present uses mean coverage of the final decade of observations (applying a temporal missing data tolerance of 50% of months at each cell). Anomalies are taken again with respect to the same period to avoid distortions introduced upon regridding (again applying a missing data tolerance of 50%) and the resulting data are then closely comparable with the observations used in the validation. We mask CRUTEM4 with land fractions (≥ 0.25) used in the construction of the HadCRUT4 data set.[25]

Seasonal mean fields are then produced (using seasons DJF, MAM, JJA, SON), retaining data in cells where all three months are present. Regional area weighted averages are taken to produce each time series using the masks depicted in SI Figure 1. Continental regions use the combined masks of regions indicated in Table 1 while the GLO region uses the combination of all masks, which hence excludes Antarctica.

Regional warm seasons are found by obtaining seasonal means of absolute near surface air temperatures from CRU CL version 2.0[27] then taking regional means.

The MME empirical CDF for each season is obtained by ranking values at that year in ascending order and assigning each a weight such that the sum over weights is normalised to one and each model is weighted equally (regardless of the number of simulations contributed by a model). The form of the weights assigned to member j of model α used in calculating seasonal CDFs is therefore of the form

$$w_{\alpha j} = \frac{1}{N_\alpha M} \tag{1}$$

where N_α is the number of members contributed by model α and M is the number of models. The weights satisfy

$$\sum_{\alpha}^{M} \sum_{j}^{N_\alpha} w_{\alpha j} = 1. \tag{2}$$

After ordering weights according to their corresponding ranked values the temperature value

of percentile $100p$ ($0 < p \leq 1$) may be found as the value at which the sum over the ordered weights up to that corresponding to this value is equal to p. Our method is very similar to that used in a recent study.[26] A small difference is that instead of adopting the temperature value of the ensemble member lying closest to the percentile of interest we linearly interpolate between the two nearest values. The value of the 100^{th} percentile is defined to be equal to that of the highest member and the 0^{th} to that of the lowest (which is therefore coincident with the value corresponding to the lowest weight).

We are interested in the time when the 95^{th} percentile of RCP 2.6 passes below the 90^{th} percentile of RCP 8.5. Noise due to the finite ensemble size means that percentile series may cross one another several times before separating clearly (a similar issue is discussed in a different context by Hawkins et al., 2014[28]), biasing HOPE Dates artificially late. To address this we smooth the percentile series using a 15-year box-car window and define the HOPE Date as the final year that the smoothed 95^{th} percentile of RCP 2.6 is greater than smoothed 90^{th} percentile of RCP 8.5.

Note that smoothing of percentile series does not reduce the width of the MME distribution. This width represents both inter-annual variability within each model and the component of inter-model spread which incorporates uncertainty in response to external forcings.

Bootstrap re-sampling of the MME is performed by removing all members associated with a single model at a time and recalculating weights and percentiles of the MME accordingly. In calculating ranges of latest possible HOPE Dates we simply replace the peak smoothed percentile value with each of its bootstrapped alternatives.

Single model HOPE Dates are found by the same method but with only a handful of members per model we instead find percentiles by assuming that single model ensemble members $x(t)$ are generated at time t by a normal stochastic process with variance σ^2 superimposed on the 15-year smoothed empirical model mean $\bar{x}(t)$. σ^2 is calculated from the residuals $r(t) = x(t) - \bar{x}(t)$ aggregated across model members and years in the period 2006 - 2099. We required a model have at least 3 ensemble members in each scenario in order that the ensemble mean be sufficiently free of internal variations. This leaves 9 models (see SI Table 1) across which the means displayed in Figure 3 (c) are taken.

Annual global mean land temperature difference values, $\Delta T_{\text{land}}^{\text{Global}}$, which appear in Figure 3 (d) are calculated from differences between scenario multi-model mean annual anomalies for the

GLO region produced by the mean over the four seasonal multi-model median values produced by the same method described above.

In the absence of observations from the future with which to validate model responses to future scenario forcings (let alone two possible futures) we assess the performance of the MME against 20$^{\text{th}}$C observations from CRUTEM4.4.0.0.[24] We acknowledge that we cannot expect this to be an entirely adequate means of judging model performance into the 21$^{\text{st}}$C as, for example, forcings that cancel one another in an earlier period may fail to do so into the future leading to disguised model deficiencies.[26,29] We shall adopt the following simple and objective validation paradigm: the MME will be considered adequate in a given region and season if we cannot tell the observational series from a member of the MME itself. As such we require that the MME of historical experiments pass two simple tests.

First (Test 1) we assess the timing and relative magnitude of 20$^{\text{th}}$ C variability (1905 - 2004) through a comparison of the mean correlation coefficient $\langle r \rangle$ between the 15-yr rolling means of the observed time series and members of the MME with values of r arising between the observations and an isospectral test ensemble, i.e. an ensemble generated by a random process sharing the same power spectrum as the true model ensemble. The test ensemble will have zero mean trend but individual members will look rather like historical series but for trends occurring at different (random) times. A regional warm season fails this test if we cannot rule out that $\langle r \rangle$ could be generated by this random process with high probability.

H_0: The mean value of the correlation coefficient, $\langle r \rangle$, between 15-year means of the observed time series and each historical ensemble member can be generated with high probability, p, by a random process isospectral to the true CMIP5 MME.

We seek to rule out H_0 at the 5% level and fail a regional warm season if we cannot. The isospectral test ensemble contains 10,000 members and upon regeneration produces p values that are robust to within a percent.

Secondly (Test 2) we assess the total time series variance of 20$^{\text{th}}$C simulations (also 1905 - 2004) across the ensemble by comparison with the total 20$^{\text{th}}$C variance of the observed time series (without any smoothing). A regional warm season fails this test if the observed total variance falls outside the full model range of total variances.

Members of the isospectral test ensemble are created as follows. Each member is constructed as a Fourier series in which the phases are selected from a uniform random distribution $\mathrm{U}[0, 2\pi]$ while the coefficients at each frequency are selected at random from among the empirical distribution of values obtained from a Fourier decomposition of each member of the true MME. At a given frequency each empirical coefficient will likely be selected many times over in the construction of many test members but the same combination of coefficients across the frequencies is very unlikely to be selected in an ensemble of only 10,000 test members so that the spectrum of each member is very likely unique in addition to possessing random phases. The resulting test ensemble Fourier spectrum will be virtually indistinguishable from the CMIP5 empirical spectrum represented by shaded colours in SI Figure 7.

Test 2 assesses the overall scale of variability to which Test 1 is blind.

Note that model series are not detrended prior to examining the correlation with observations, power spectra or total variance. This is because we wish to test the relative magnitude and timing of forced responses rather than remove them; we are not attempting to isolate internal variability.

SI Figure 6 displays the warm season historical model and observational time series together with Test 1 results. Out of 26 SREX regions 21 pass this test as well as all 6 continents and the global combination. Most regions have $\langle r \rangle \geq 0.6$, many with p values of 1% or better. The continental and GLO region have $\langle r \rangle$ values higher still, all with significance values of 1% or better. While a failure such as 15. WAF (MAM) may be suspected from casual examination of the time series (poor correlation will arise from differing trends in each half of the century) other failures would not be clear, emphasising the need for an objective criterion.

SI Figure 7 displays the warm season historical model and observational Fourier spectra with Test 2 results. Two regions fail the total variance test in their warm season: 7. AMZ (SON) and 23. SAS (JJA), both of which have observed total variances that fall below every member of the MME. We note that the observed total variance falls close to the bottom of the model range in more than half of the SREX regions and 3 out of 6 continents however.

Data availability The model data supporting this study is available in a public repository, for example at http://cmip-pcmdi.llnl.gov/cmip5/. The Met Office observational dataset CRUTEM4 is publically available from http://www.metoffice.gov.uk/hadobs/.

References

[24] PD Jones, DH Lister, TJ Osborn, C Harpham, M Salmon, and CP Morice. Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D5), 2012.

[25] Colin P Morice, John J Kennedy, Nick A Rayner, and Phil D Jones. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 117(D8), 2012.

[26] Gareth S Jones, Peter A Stott, and Nikolaos Christidis. Attribution of observed historical near–surface temperature variations to anthropogenic and natural causes using cmip5 simulations. *Journal of Geophysical Research: Atmospheres*, 118(10):4001–4024, 2013.

[27] Mark New, David Lister, Mike Hulme, and Ian Makin. A high-resolution data set of surface climate over global land areas. *Climate research*, 21(1):1–25, 2002.

[28] Ed Hawkins, Bruce Anderson, Noah Diffenbaugh, Irina Mahlstein, Richard Betts, Gabi Hegerl, Manoj Joshi, Reto Knutti, Doug McNeall, Susan Solomon, et al. Uncertainties in the timing of unprecedented climates. *Nature*, 511(7507):E3–E5, 2014.

[29] Piers M Forster, Timothy Andrews, Peter Good, Jonathan M Gregory, Lawrence S Jackson, and Mark Zelinka. Evaluating adjusted forcing and model spread for historical and future scenarios in the cmip5 generation of climate models. *Journal of Geophysical Research: Atmospheres*, 118(3):1139–1150, 2013.