Beyond RCP8.5: Marginal Mitigation Using Quasi-Representative Concentration Pathways

William A. Brock¹,² and J. Isaac Miller², *

¹ University of Wisconsin-Madison, Madison, Wisconsin, USA
² University of Missouri, Columbia, Missouri, USA
* milleriisaac@missouri.edu

Abstract

Assessments of decreases in economic damages from climate change mitigation typically rely on climate output from computationally expensive precomputed runs of general circulation models (GCMs) under a handful of scenarios with discretely varying targets, such as the four representative concentration pathways (RCPs) for CO₂ and other anthropogenically emitted gases. Although such analyses are extremely valuable in informing scientists and policymakers about specific, well-known, and massive mitigation goals, we add to the literature by considering potential outcomes from more modest policy changes that may not be represented by any concentration pathway or GCM output. We construct computationally efficient Quasi-representative Concentration Pathways (QCPs) in order to leverage existing scenarios featuring plausible concentration pathways. Computational efficiency allows for common statistical methods for assessing model uncertainty based on iterative replication, such as bootstrapping. We illustrate by feeding two QCPs through a computationally efficient statistical emulator and dose response functions extrapolated from estimates in the recent literature in order to gauge effects of mitigation on the relative risk of heat stress mortality.

Introduction

Scenarios that consider future emissions of CO₂ and other well-mixed greenhouse gases and their associated climate projections are central tools in assessing the impacts of future climate change. In particular, the Representative Concentration Pathways (RCPs) and especially RCP8.5 are often used to assess risks and possible costs associated with climate change. There are four RCPs: RCP2.6, RCP4.5, RCP6.0, and RCP8.5, where the “w.z” in RCPw.z stands for Watts per meter squared (W/m²) of radiative forcing in 2100.

The IPCC Fifth Assessment Report (AR5) database¹ provides an additional 1,184 scenarios from 31 models, 807 of which pertain to CO₂ concentrations. Pierrehumbert (2014)² among others have argued that the much longer lifetime of CO₂ compared to other greenhouse gases justifies an emphasis on CO₂ for long-run projections, and emulators such as that of Castruccio et al. (2014)³ focus on CO₂ accordingly.

Deryugina and Hsiang (2017)⁴ recently conceptualize the “marginal product of climate” with respect to the production of economic output. An equally important concept involves the marginal product of CO₂ with respect to the “production” of climate change. Of course, this concept is not new, but we emphasize the usefulness of examining benefits from marginal or incremental changes to concentration pathways in contrast to more ambitious policy changes along the lines of Burke et al. (2018)⁵ or Monier et al. (2018).⁶

In order to consider marginal changes from a policy, ideally we would run a single general circulation model (GCM) embedded in an integrated assessment model under two scenarios holding technology, socio-economic drivers, and other policy measures constant, differing only in the one key dimension of the proposed policy change. The model would take into account not only the properly discounted benefits from the policy change, but also the costs, and it would further allow for parameter instability from changes in technology, adaption, etc.
The two GCM model runs would be replicated a large number of times using a bootstrap or similar procedure to assess uncertainty in the model parameters and data. Why? Palmer (2019) recently emphasized the need for stochasticity in climate models. He writes:

“… the spread generated in ensembles that only have initial perturbations is typically too small, particularly in the tropics, implying that the observed values fall outside the range of the ensemble too often. This implies that there is a second source of uncertainty, not represented in purely initial condition ensembles: model uncertainty.”

The bootstrap is a well-known statistical tool that draws models from a common statistical distribution. In other words, the bootstrap generates an arbitrarily large ensemble of models that take into account not only model uncertainty, but any other type of uncertainty that can be programmed into the replications. The main obstacle is computational time.

Given the state of technology, the approach just outlined would be prohibitively time-intensive because the amount of computational time that goes into a single model run is enormous. The pathways themselves are only starting points, and the associated output from GCMs is massive in spatial resolution and temporal span and frequency. Not only does computational time prohibit replication of output, it also complicates traditional statistical measures of uncertainty that rely on iterative replications, such as the bootstrap. Two natural alternatives are provided by statistical emulation and pattern scaling of selected output variables, such as temperature and precipitation. These allow for quickly implemented approximations to comprehensive GCM model output.

In this paper, we take a step towards estimating and computing the “marginal product of CO₂” of a hypothetical policy to move 2100 radiative forcing due to CO₂ from $F = x.y$ to $F - dF = (x-a,y-b)$ for $a,b > 0$ for a given component of projected future climate change, temperature, and ultimately for a given component of damages, relative risk of heat stress death. We see our methodology as a route towards informing incremental changes in policies to combat a very specific cost of climate change, such as increased mortality due to heat stress. However, ours is not an integrated assessment model and is not intended to provide a full cost-benefit analysis of such a change.

To this end, we construct what we call Quasi-Representative Concentration Pathways (QCPs) for values of radiative forcing $w.z$ at 2100, other than the values 2.6, 4.5, 6.0, and 8.5 of the four existing RCPs and those of the 807 AR5 scenarios. We use the QCPs to project spatially and temporally disaggregated temperature changes for the period 2031-2100 from changes in radiative forcing due to CO₂ from $F = x.y$ to $F - dF = (x-a,y-b)$ at 2100 using an emulator of bias-corrected GCM runs.

Both $w.z$ and $x.y$ are forcings in W/m² in 2100, but the former represents forcings from all sources, corresponding to the RCP and our QCP labels, while the latter represents forcing due to CO₂ concentrations. In fact, our analysis uses CO₂ concentrations in ppm rather than W/m². By equation (1) below, $x.y$ follows directly from a given concentration, but $w.z$ does not follow directly and may be viewed simply as a label.

A deficiency of the state of modeling technology noted by Monier et al. (2018) is the use of models that span different modeling groups, policy and technology assumptions, etc. The deficiency is especially salient in studies that use different modeling assumptions to generate output for the same scenario, such as RCP8.5, in order to assess uncertainty about those outputs. It seems unavoidable given the computational intensity involved, but we address the deficiency in two ways.

First, the use of an emulator allows high-speed model replication to assess uncertainty, obviating the need for climate outputs from disparate modeling groups for this purpose. The transmission of CO₂ concentrations to climate change is accomplished by emulating an ensemble of models from the same family, CMIP5 HadGEM2-ES, also used by Liu et al. (2017) to assess mortality risk from climate change. This component of the model is trained not on hypothetical future emissions paths, but on...
historical model output that is bias-corrected to match the time series characteristics of the instrumental record.

Second, we do use scenarios from a large number of modeling groups, but only to generate a single QCP as model input. CO₂ concentrations may be output from models to which CO₂ emissions are input and an infinite number of emissions paths could result in the same concentration in 2100. Instead, we configure the QCPs to take into account information from all of the 811 scenarios (4 RCPs plus 807 AR5 scenarios), weighted by proximity. We use the QCPs to evaluate marginal changes that do not align with any of the 811 scenarios.

As an example of the usefulness of our approach in contributing to an important policy discussion, we compare a path resulting in 3.5°C above pre-industrial temperatures in 2100 along the lines of Monier et al. (2018)⁸ with a hypothetical path resulting in 2.5°C, which is similar to but not exactly the same as some of the AR5 scenarios. Gasparrini et al. (2015)¹² estimate “dose response” functions that map relative risk (RR) of mortality from heat exposure to temperature using data on over 74 million deaths at a variety of different locations between 1985 and 2012. We utilize these functions to project RR of mortality from potentially increased heat exposure due to climate change to the end of the century and from a hypothetical policy that reduces the temperature in 2100 from 3.5°C to 2.5°C.

Why such a narrow focus on heat stress deaths since there are so many components of potential climate change damages? Morgan et al. (2017)¹³ and Pezzey (2019)¹⁴ recently make powerful arguments against trying to compute and project an aggregate measure of total climate change. They argue that attempts to monetize the full set of potential damages needed to create an aggregate “damage function” will always be so controversial that an alternative approach, such as finding an efficient pathway to implement a target (a carbon budget, e.g.) is a preferable. By focusing on heat stress deaths, we take a more modest approach with the virtue of a very tight focus on projecting the impact of a component of climate damage on a component of damages that is supported by dose response functions backed up by extremely plentiful data.

The BRACE¹⁵ (Benefits of Reduced Anthropogenic Climate ChangE) project of the National Center for Atmospheric Research can be viewed in the same light. For example, Tebaldi and Lobell (2018)¹⁶ compare outcomes for maize and wheat under RCP8.5 and RCP4.5. Our work can be viewed as an analogous exercise for heat stress deaths in humans in units of RR for a collection of cities made possible by the detailed collection of data and estimation of dose response functions by Gasparrini et al. (2015)¹² by city, but our innovation allows plausible alternative scenarios beyond the RCPs and AR5 scenarios.

In a recent study, Carleton et al. (2018)¹⁷ carefully examine the relationship between carbon emissions and mortality. Their study goes further than ours in a few key dimensions. They use raw mortality data to estimate heterogeneous dose response functions resulting in a projected number of deaths resulting from climate change. In doing so, they predict adaption to climate change. Moreover, they embed their dose response functions into an integrated assessment model, estimating a social cost of carbon specific to heat stress. Yet, like Tebaldi and Lobell (2018)¹⁶ and others, they consider only a small number of scenarios, RCP8.5 and RCP4.5.

Our QCPs complement comprehensive studies like that of Carleton et al. (2018)¹⁷ by approximating results for mitigation targets less ambitious than a reduction of 4 W/m², the difference between RCP8.5 and RCP4.5. Results like theirs could be extended along our QCP framework to investigate marginal policy questions. Such an approach – especially using RR rather than actual mortalities – would have the advantage of avoiding controversial assumptions about population projections, projections of the value of lives, and projections of other socioeconomic dynamics, which are not subject to stable physical or biological laws as the climate and physiological systems are.
Results

Marginal Climate and QCPs

Moss et al.\textsuperscript{18} write, “The RCPs will be used to initiate climate model simulations for developing climate scenarios for use in a broad range of climate-change related research and assessment,” and they were requested to be “compatible with the full range of stabilization, mitigation and baseline emissions scenarios available in the current scientific literature.” We want to be able to project RR corresponding to a forcing from CO\textsubscript{2} concentration in 2100 of \( x.y \) W/m\(^2\) for a set of latitudes and longitudes over the period 2031-2100. As we vary \( x.y \) continuously, we want our QCPs to satisfy criteria similar to those imposed on the RCPs and AR5 scenarios.

Suppose that we consider policies aimed at two climate scenarios by 2100: 2.5°C and 3.5°C in 2100. These temperatures correspond to point estimates using forcings in 2100 of 579 and 752 ppm CO\textsubscript{2} concentrations by our model described below. Monier et al. (2018)\textsuperscript{6} dubbed the 3.5°C scenario Paris Forever, with Paris Agreement commitments for emissions reductions but no climate policy after 2030. The 2.5°C scenario represents a more aggressive effort to curb emissions than Paris Forever, in the sense that a substantive 1°C reduction is achieved, but one that fails to limit temperature to 2°C. We label the 2.5°C scenario Paris Plus.

Importantly, Paris Plus does not correspond to any of the RCPs, nor does it correspond to any of the scenarios of Monier et al. (2018).\textsuperscript{6} One of the AR5 scenarios yields close to 579 ppm 2100. Such may not be the case with, say, 3.4°C or 3.3°C, where the temperature reduction from the Paris Forever scenario would be more marginal.

In order to analyze scenarios such as Paris Plus on a policy-relevant time scale, one might simply use the AR5 scenario that is closest in some sense, or one might interpolate between the nearest AR5 scenarios, RCPs, or paths published elsewhere. Our QCPs generalizes these approaches in two ways. First, and most importantly, we consider a weighted average of up to 811 scenarios: 4 RCPs plus 807 AR5 scenarios, with weights declining in distance from the target. In the case of Paris Plus, the AR5 scenario resulting a CO\textsubscript{2} concentration closest to 579 ppm receives the most weight, while one resulting in more than 1230 ppm or less than 382ppm, the two most extreme, receives no weight. Second, by weighting each of the parameters of a quadratic approximation of the log of the concentration pathways, we allow the curvature of the QCPs to approximate that of pathways resulting in forcing in 2100 near the target. Refer to the Methods section for more details.

We do not suggest constructing QCPs that would result in forcings more extreme than available scenarios – less than 382 ppm or greater than 1230 ppm in the case of AR5 scenarios – for two reasons. First, the weighting scheme will take into account only scenarios strictly greater than or strictly less than the QCP, which may bias the results, akin to the well-known empty bin problem associated with this type of nonparametric estimation. Second, Castruccio et al. (2014)\textsuperscript{3} suggest that the emulator on which ours is based performs best when the prediction set – i.e., the QCP – is not extreme.

A potential concern is that QCPs are not generated by a single modeling group under a single set of assumptions about technology, policy, etc., but aggregate multiple pathways. However, model averaging is a widely accepted method used across many academic disciplines precisely because averages of forecasts made by multiple models can be more robust than those from a single model if the models are weighted by skill and the procedure is properly implemented.\textsuperscript{19,20,21,22}

Paris Forever and Paris Plus correspond to 752 and 579 ppm CO\textsubscript{2} concentrations in 2100 respectively, while RCP2.6, RCP4.5, RCP6.0, and RCP8.5 correspond to 420.90, 538.36, 669.72, and 935.87 ppm respectively. We label Paris Forever as QCP6.77, simply finding \( w.z \) that interpolates 6.0 and 8.5 as 752 interpolates 669.72 and 935.87, and label Paris Plus as QCP4.96 similarly. Figure 1 shows these two QCPs along with the QCPs corresponding to the four RCPs, RCP8.5, RCP6.0, RCP4.5, and RCP2.6, which suggest the close approximation of the nonparametric weighting.

\textsuperscript{3}Castruccio et al., \textit{Clim. Past}, 2014

\textsuperscript{6}Monier et al., \textit{Clim. Past}, 2018
Emulation of Monthly Spatially Disaggregated Temperatures

We use bias-corrected historical monthly CMIP5 HadGEM2-ES model ensemble output for the Northern Hemisphere over 1860-2004, which is available monthly at increments of 1.25° latitude by 1.875° longitude to implement the emulator. Once the emulator is trained, we use it to forecast monthly temperatures over 2031-2100 under a QCP. After Castruccio et al. (2014) thoroughly train and evaluate their emulator over multiple runs, multiple scenarios, and multiple centuries, together using more than 10,000 model years of data, they conclude that their approach “permits emulation with a relatively small set of precomputed runs” and that the “choice of training set is not especially crucial if prediction and training scenarios are similar.” Their thorough testing justifies our use of the emulator over the historical record and a QCP less extreme than the AR5 scenarios for which we have data.

We motivate the emulator in the Methods section using a simple model of Arrhenius’s relationship. We improve on that of Castruccio et al. (2014) by neglecting short-run effects, which allows much faster estimation by linear regression rather than by numerical optimization and thus analysis of more finely disaggregated spatial data and facilitates bootstrap methods used to account for uncertainty. We explain the bias correction methodology and give a formal proof of the emulator’s statistical consistency in the Supplementary Online Materials (SOM).

Figures 2 and 3 show the results of the emulator. The top panel of Figure 2 shows bias-corrected historical average June-August temperatures over a base period, 1971-2000, for most of the Northern Hemisphere. The middle panel shows the increase from the baseline in average temperatures June-August temperatures over 2071-2100 under Paris Plus. The bottom panel shows the increase from Paris Plus in average temperatures June-August temperatures over 2071-2100 under Paris Forever.

Not surprisingly, the largest increases under either scenario are projected around Greenland, the Northern Atlantic, Kamchatka, and the Northern Pacific. Polar amplification is thus evident and appears to be conveyed mainly by ocean currents. Noticeable increases are projected for the Eastern US, US-Canadian High Plains, US-Mexican Southwest, much of northwestern Europe and northwestern Africa, and parts of south and southeastern Asia. In contrast, June-August temperatures are projected to decrease in some areas, notably in Turkey, Central Asia, and the vicinity of northeastern China and the Korean peninsula.

Figure 3 shows the bias-corrected historical annual average temperatures and annual average temperatures projected conditionally on the two scenarios for twelve cities for which Gasparrini et al. (2015) estimated dose response functions for heat exposure. Rome, Madrid, Stockholm, Bangkok, London, and New York City exhibit similar and expected patterns of projected warming with faster warming under the scenario in which the globe warms faster. There is a major difference in magnitudes, however, with Stockholm and London gaining several degrees under either scenario and the remaining four warming considerably less. The difference is easily explained both by the high latitudes of these two cities (hence polar amplification) and their proximity to the North Atlantic.

In contrast, Toronto, Beijing, Tokyo, Taipei, and Chicago show very little and slightly negative projected change, while Seoul shows a noticeable negative projected change. One explanation for the declines is that these are annual averages. The emulator allows winters to get colder while summers get warmer, so that mortality risk from extreme heat may increase from an increasing seasonal variance even as the annual average temperature decreases.

Another possible explanation for the cooling in East Asia is offered by the use of CO₂ concentrations rather than concentrations of CO₂ equivalents, which would include sulfur dioxide. As Kaufmann et al. (2011) point out, sulfur emissions have increased enormously in China in recent decades. Although a role for sulfur emissions in slowing down global warming as suggested by Kaufmann et al. (2011) and hinted by Storelvo et al. (2016) is controversial, a role for sulfur emissions in regional cooling near eastern China seems plausible.

---

1. Arrhenius’s relationship.
2. Castruccio et al. (2014).
5. Supplementary Online Materials.
7. CO₂ equivalents.
Annual Average Daily Relative Risks

Estimated from those of Gasparrini et al. (2015),\textsuperscript{12} our dose response functions are spatially heterogeneous like those the Liu et al. (2017),\textsuperscript{11} yet nonlinear in temperature like those of Hsiang et al. (2017).\textsuperscript{12} We acknowledge, however, that these dose response functions do not account for relevant omitted variables, such as those reflecting health and other correlated dangers from climate change – including malnourishment, food insecurity, tropical disease, and respiratory illnesses from pollutants (Dear and Wang, 2015; Crowley, 2016)\textsuperscript{26,27} – and those reflecting wealth and demographics, which could facilitate adaption, as it has in the US (Barreca et al., 2016).\textsuperscript{28} Recent progress has been made in this area by Carleton et al. (2018).\textsuperscript{17}

Figure 4 shows the output of the annual average daily dose response functions for the twelve cities discussed above. Specifically, the solid black lines show the expected annual average relative risks under Paris Forever projected over 2031-2100. Risk here is relative to no heat exposure. The solid grey lines show annual average risks under Paris Plus relative to those under Paris Forever. Thus, a negative value shows a benefit from the more aggressive mitigation policy.

The dashed lines represent 95% bootstrap forecast intervals. We defer a detailed discussion of the bootstrap to the Methods section, but note here that it is intended to capture uncertainty from three important sources: statistical estimation of the emulator, daily variations in temperature over the course of each month in the projection, and uncertainty of the estimated dose response functions. Note that the bounds of the intervals are for a given point in time, but should not be interpreted pathwise.

The most salient aspect of Figure 4 is the heterogeneity in risk patterns. We expect patterns like that of Rome, Stockholm, London, and New York, in the sense that mortality risk increases under Paris Forever, but that the increase is mitigated under Paris Plus (negative relative risk). Toronto, Taipei, and Bangkok are similar, but the risk is already elevated. Madrid and Chicago are also similar, but the conditional expectations are flat, meaning that we do not expect any change in risk but that the possibility of a given risk level is increasing. Both conditional expectations and 95% intervals show zero risk relative to a day with no heat stress for Seoul.

Risks in Tokyo and Beijing are difficult to interpret. Tokyo has a flat conditional expectation, but the possibility of a given risk level is projected to decrease under Paris Forever. Paris Plus may actually increase the risk relative to Paris Forever. The figure for Beijing suggests the possibility that Paris Forever actually decreases risk over time, and the sign of the effect of the mitigating Paris Plus scenario vs. Paris Forever is ambiguous.

Also heterogeneous are the levels of risk, given by the units of the vertical axes. Expected average daily relative risk increases by only 0.0007, 0.0009, and 0.0039 per day in Toronto, New York, and Rome respectively over the entire 70 year period. However, it increases by 0.0521, 0.0440, and 0.0230 in London, Bangkok, and Stockholm respectively over the same period. Fortunately, under the more aggressive Paris Plus, the latter are mitigated by -0.0377, -0.0193, and -0.0143. Paris Plus has a concrete, measurable effect on mortality risk in these cities.

Risks under both scenarios are ameliorated by adaption, as highlighted by Barreca et al. (2016)\textsuperscript{26} for the US. In relatively wealthy and relatively temperature locations like London and Stockholm, adaption is both economically and technologically feasible. In poorer and hotter locations like Bangkok, adaption may be more difficult. A hotter future London may be no hotter than, say, Texas or Florida today, which have already learned to deal with heat. A hotter future Bangkok may be hotter than any developed city on Earth today, so there is no current technological model for dealing with that degree of heat.
Discussion

This paper puts forth a novel method to evaluate what might be called the “marginal product of CO$_2$” on a component of damages – mortality from heat stress – from a component of climate change – spatially and temporally disaggregated temperature changes. Our contribution is primarily methodological. We show how one can leverage the massive amount of work done by large climate modeling groups to produce the RCPs and AR5 scenarios and ensuing output in a computationally efficient manner. Computational efficiency allows one to generate paths that reflect marginal changes in concentrations beyond RCP8.5 and other wisely used scenarios.

To this end, we optimally weight a large number of these scenarios by proximity to a target outcome in, say, 2100. These so-called Quasi–representative Concentration Pathways (QCPs) are easy and fast to construct, so that, along with a computationally efficient climate model emulator, the methodology lends itself readily to bootstrapping or other methods commonly used to assess statistical uncertainty in multi-step modeling procedures.

We modestly improve on the emulator of Castruccio et al. (2014) by showing that an equivalent reduced-form emulator may be estimated consistently using linear least squares. Further, we extend the dose response functions of Gasparrini et al. (2015) to show how the QCPs may be used to address simple policy-relevant issues of incremental changes. Alternative emulators or alternative dose response functions, such as those of Deschênes and Greenstone (2011) and Carleton et al. (2018), could be used with our QCPs instead. Our choice of emulator and dose response function improve the speed with which we can obtain uncertainty intervals.

An open area in incremental policy evaluation under climate change and mitigation is to provide better measures of uncertainty that are disciplined by state-of-the-art climate models as well as state-of-the-art dose response functions at the smaller scales of key policy interest. We have made a start here but much remains to be done when more complete data sets become available and as progress is made by GCM emulation teams in downscaling climate change effects to yet smaller scales where much of policy concern lies.

Methods

The main contribution of this paper is the conceptualization and promotion of using data disciplined objects like our QCP($w,z$) to project relative risk (RR) for a continuum of values $w,z$ for a range of geographic locations. In this way, marginal RRs can be projected for an interval of dates, such as 2031-2100, so that one can explore impacts on RR by a marginal policy change that moves radiative forcing from CO$_2$ in 2100 from $F = x.y$ to $F–dF = (x–a.y–b)$, for $a,b > 0$. The use of an emulator allows the use of computationally efficient numerical methods for assessing uncertainty, such as the bootstrap, while the QCP itself is generated by an ensemble of future concentration pathways developed by a large number of disparate models and modeling groups. We do not believe that there is any computationally efficient way to capture temperature and risk uncertainty under any given scenario other than such a bootstrapped emulator.

Marginal Climate and QCPs

The main innovation of our analysis lies in constructing QCPs, so we enumerate the procedure here:

1. Fit the natural log of each of the 811 pathways, 4 RCPs and 807 AR5 scenarios, to quadratic functions of time over 2031-2100 using nonlinear least squares, restricting the quadratic to pass through two points: the log CO$_2$ concentration in 2065, the midpoint, and the log CO$_2$ concentration in 2100 under the RCP. In order to identify the curvature of the quadratic, the log CO$_2$ concentration in 2031 is a free parameter chosen to maximize the fit of the quadratic over 2031-2100. We can think of these 811 approximations as “baseline” log QCPs. For example, log QCP8.5 approximates the
natural log of the pathway described by RCP8.5. We label the resulting fitted quadratic function as:
$$\gamma_{1,i}(t/T)^2 + \gamma_{2,i}(t/T) + \gamma_{3,i}$$ for year \(t\) and scenario \(i\), setting 2030 to be \(t=0\) and 2100 to be \(t=T=70\).

2. Determine a value of forcing due to CO\(_2\) concentrations in ppm in 2100, with its natural log denoted by \(h_{2100}\), corresponding to the desired value of total forcings \(w.z\) in W/m\(^2\) for the desired QCP. Or, conversely, determine the value \(w.z\) corresponding to forcing due to CO\(_2\). As described above, we take the latter approach, labeling by QCP6.77 (Paris Forever) and QCP4.96 (Paris Plus) the scenarios that yield values that interpolates the CO\(_2\) concentrations, 752 and 579 respectively, between the nearest RCPs. This interpolation is merely for labeling and is not part of the construction of the QCP.

3. Using a local regression with Nadaraya-Watson kernel estimator,\(^3\) regress \(\gamma_{j,i}\) for \(j=1,2,3\) (three local regressions) onto the baseline log QCPs in 2100. The baseline log QCPs in 2100 are given by \(\gamma_{1,i} (T/T)^2 + \gamma_{2,i}(T/T) + \gamma_{3,i} = \gamma_{1,i} + \gamma_{2,i} + \gamma_{3,i}\) with parameters estimated in Step 1. We use a Gaussian kernel with Silverman bandwidth, and we implement this step by evaluating the weights over a range of values of log concentrations in 2100 to create a universe of log QCPs. For each point of evaluation – i.e., each log concentration in 2100 – the local regressions average each of the three parameters of the quadratic functions of the 811 baseline log QCPs weighted by proximity of the log concentration in 2100 of those baselines to the log concentration at the point of evaluation.

4. Finally, choose the (log) QCP with parameters such that \(\gamma_{1,0} + \gamma_{2,0} + \gamma_{3,0}\) best matches the target \(h_{2100}\). The QCP is given by the exponent of \(\gamma_{1,0}(T/T)^2 + \gamma_{2,0}(T/T) + \gamma_{3,0}\) for each year \(t\).

For simplicity, steps 3 and 4 can by combined by evaluating the Nadaraya-Watson kernel estimator local only to \(h_{2100}\), rather than generating a universe of log QCPs over a range of log concentrations.

**Emulation of Monthly Spatially Disaggregated Temperatures**

Castruccio et al. (2014)\(^3\) justify statistical emulation of climate models (GCMs) as follows:

“We propose … [an] emulation approach based on a collection of precomputed climate model runs that allows us to capture rate dependencies in regional climate evolution. This collection of runs, or training set, is used to obtain estimates of the parameters in simple statistical models that describe temperature and precipitation as a function of past trajectories of radiative forcing due to CO\(_2\). The resulting tool allows us to reproduce (emulate) the output of [a GCM] under a large range of forcing scenarios. Once the emulator is constructed, emulation of a climate scenario is effectively instantaneous. … In contrast, climate projection from a state-of-the-art model can still take days to weeks even on the most powerful platforms. … The simplicity and robustness of statistical emulation … makes it a promising tool for impacts assessment.”

In this light, emulation is an excellent and efficient tool for translating QCPs, which are based on radiated forcing scenarios fed into state-of-the-art models, into climate projections that may be used for an impact assessment.

Pierrehumbert (2014, equation 1)\(^1\) gives radiative forcing \(\Delta F\) at date \(t\) due to CO\(_2\), expressed in W/m\(^2\), as

\[(1) \ \Delta F = 5.35 \ln(c_i / c_b) = 5.35(h_i - h_b),\]

where \(c_i\) and \(c_b\) are atmospheric CO\(_2\) concentrations at date \(t\) and a base period \(b\) and \(h_i\) and \(h_b\) are defined as their natural logs. Suppose policymakers want to move radiative forcing due to CO\(_2\) at 2100 from \(x.y\) W/m\(^2\) to \((x-a,y-b)\), for \(a,b>0\). Then the change

\[(2) \ \Delta F = 5.35 \ln[c_{2100}(x-a,y-b) / c_{2100}(x,y)] = 5.35[h_{2100}(x-a,y-b) - h_{2100}(x,y)]\]

would result.

At first blush one would think that in order to compute the net benefit in reduction of RR of mortality from heat exposure that the world is getting in enacting this change, we would need to know how much of
the world’s emission budget is being reduced and how this reduces GDP and utility of material consumption as well as how much the temperature increase is reduced. But this route involves so many “unknown unknowns” about the future that it seems fruitless. We take an alternative route here. Our route starts with using the reduction in the atmospheric concentration of CO₂ measured in parts per million (ppm) resulting from reducing radiative forcing x.y at 2100 to \((x-a.y-b)\) at 2100 which is given by equation (2) above.

We use our QCPs to leverage information about the path of atmospheric concentration of CO₂ from the work on developing the RCPs and AR5 scenarios where the change in total concentration at 2100 is pinned down by (1) and (2) for each QCP\(w.z\). In other words, we produce a time path of concentrations at each date \(t\) for each QCP\(w.z\).

The time path of concentrations is then fed into estimated emulator equations derived from those of Castruccio et al. (2014).\(^3\) The reason for not using exactly the same emulator is that their focus on both the short and long run leads to a nonlinearity such that the parameters may not be numerically identified, as explained in the supplement to their paper. Our emphasis on long-run emulation allows us to ignore this issue and avoid numerical optimization of the emulator parameters, improving computational stability and efficiency.

Denote by \(\Delta F((x-a.y-b) | x.y)\) the reduction in 2100 radiative forcing from a policy to move 2100 radiative forcing from \(x.y\) to \((x-a.y-b)\), for \(a,b > 0\). The ratio,

\[
(3) \quad \frac{dT_{2100}}{\Delta F((x-a.y-b) | x.y)}
\]

represents a degree Celsius temperature reduction in 2100 at a given location from a single \(W/m^2\) reduction in radiative forcing due to CO₂ in 2100.

We need to construct high-frequency temperatures concentrations for a set of high-resolution latitudes and longitudes in order to project RR using the city-specific data of Gasparrini et al. (2015).\(^1\) Liu et al. (2017)\(^1\) use daily temperatures from CMIP5 HadGEM2-ES model output under the four future RCP scenarios. Like those authors, we use HadGEM2-ES model output temperature at a high-frequency resolution, but we use bias-corrected monthly temperatures over the historical record from an ensemble of HadGEM2-ES runs.

Using model output allows disaggregation to a level not possible with the instrumental record. However, in order to force the dose response functions, we must bias-correct the model output over a subset of the data using the instrumental record. Once we condition high-resolution temperatures on historical CO₂ concentrations, which we do using historical data published by NASA\(^3\) and NOAA\(^3\) (Mauna Loa series) with 1959-2011 values given by an average of the two series, we can emulate temperature distributions for any hypothetical concentration pathway – in particular, for any QCP.

We explain the bias correction and emulation procedures in detail in the SOM, including a proof of the statistical consistency under empirically tested statistical assumptions.

**Annual Average Daily Relative Risks**

Once we have approximations like (1)-(3) to the change in temperature at each date \(t\) induced by moving 2100 radiative forcing due to CO₂ from \(x.y\) to \((x-a.y-b)\), for \(a,b > 0\) using bias-corrected temperature data, we can run the change in temperature through “dose response” functions to approximate the reduction in RR for any specific location for which we have dose response functions.

Gasparrini et al. (2015)\(^1\) estimate dose response functions for 363 cities in the Northern Hemisphere and another 21 in the Southern Hemisphere. The geographic distribution of the cities in this data set is non-uniform, with good coverage in the Northern Hemisphere of Italy, Japan, Spain, Thailand, the UK, and the more heavily populated parts of China and the US. Notably absent from the dataset are Africa, France,
India, the Middle East, and Russia, which are prone to deadly heat events or have experienced them in the 21st Century.

We employ these dose response functions, but we modify them in two important ways. First, we fit the right “hot” tails to quadratic functions. The dose response functions were estimated by those authors based on the historical record, but approximating the hot tail is necessary in order to extend it to levels of heat unprecedented in the respective cities. We do this by setting the vertex of the quadratic at the values published by those authors. Mathematically, for the quadratic $c_1 z^2 + c_2 z + c_3 = 0$, we choose $c_1$ using nonlinear least squares subject to the constraints that $c_2 = -2c_1 MMT$ and $c_3 = 1 + c_1 MMT^2$ where $MMT$ denotes the minimum mortality temperature with unit relative risk. Second, we flatten the left “cold” tails, simply because of our focus on mortality from heat exposure. Table S1 of the SOM shows our parameters estimates.

Recall that our temperature projections have been downscaled to a monthly frequency. We believe that the monthly frequency is high enough to capture predictable seasonality but low enough for reasonable projections: we can predict that July 2100 will be hotter than June 2100 for a given location but not that July 2, 2100 will be hotter than July 1, 2100. We treat monthly relative risk as a geometric mean of 30 days per month of relative risks. In order to calculate an annual average, we simply take a geometric mean of the monthly relative risk projections for each year.

Having said that, we acknowledge that monthly average temperatures may underestimate the risk of exposure relative to daily averages because the dose response functions are convex. For this reason, we use daily data for assessing forecast uncertainty.

**Model and Forecast Uncertainty**

We face uncertainty not only in estimation, but also in the underlying data. It is a difficult but important task to account for the main sources of uncertainty, a task which is considerably facilitated by numerical statistical methods, such as the bootstrap, which cannot be run by iterating state-or-the-art climate models due to computational constraints. This is where emulation is particularly useful.

We do not explicitly account for uncertainty in the forcings data, but this is ameliorated by using CO$_2$ rather than CO$_2$ plus harder-to-measure concentrations (Myhre et al., 2013). Because our modeling goal is long-term projections and the long-run trend in temperature seems to be driven by well-mixed and temporally persistent CO$_2$ concentrations, we accept the possibility of regional biases that may result from not projecting short-run, local, and highly uncertain forcings from, say, aerosols. Once we estimate a QCP, we treat that pathway as a fixed scenario and condition our climate projections on it.

Temperature uncertainty is dealt with in the bias-correction procedure using an ensemble of HadGEM2-ES model runs to match both the first moments (bias-correction) and second moments (uncertainty and temporal variation) of the disaggregated historical temperature data.

We use a bootstrap to account explicitly for further uncertainty. Specifically, we bootstrap uncertainty in three areas: (a) statistical uncertainty in the training of the emulator and therefore in the temperature projections, (b) daily temperature variations within a month, and (c) statistical uncertainty in the estimation by Gasparrini et al. (2015) of the dose response functions. Our methodologies treat these uncertainties as being mutually but not necessarily serially independent.

In order to bootstrap the emulator, we forecast temperatures for each month in each region of dimension 1.25° latitude by 1.875° longitude in the NH, which amounts to 12 x 72 x 192 = 165,888 regressions using annual data. We employ an AR(1) bootstrap to preserve serial correlation in the errors, in the spirit of Poppick et al. (2017). We redraw contemporaneously from the AR(1) whitened annual residuals from the 165,888 regressions, thus preserving any spatial correlations or seasonality in the residuals. We then reconstruct the AR(1) residuals and temperatures for the bootstrap sample and reiterate the point
forecasts, adding the AR(1) residuals to them to account for forecast uncertainty. This creates monthly temperature forecasts that fluctuate from year to year around the point forecasts for a specific month.

Although the procedure just described takes into account seasonality, it does not take into account daily variations with a month. These variations are projected to be zero, but non-zero variations are critical in assessing the risk from heat exposure: the average of today and tomorrow could be temperate if either both days are temperate (entailing no risk) or if today is cold and yesterday will be hot (entailing some risk).

Daily variations are bootstrapped as follows. We calculate the standard deviation of the first differences of bias-corrected daily model output over a subsample, 2000-2004, for each month and location. We then add a 30-day random walk with increments having that standard deviation to each monthly temperature forecast. The random walk should not be taken literally. It simply adds persistence to the daily variations, so that heat waves or cold snaps do not dissipate immediately unless they occur at the end of the month. Although the random walk creates discontinuities in the sample path at the end of each month, such discontinuities are averaged out over months and ultimately over years.

Finally, we capture uncertainty in the dose response functions from the uncertainty bands estimated by Gasparrini et al. (2015). While we do not have those authors’ original data, we have data for each city’s estimated dose response function as well as 95% confidence intervals for those functions of temperature. Just as we fit the dose response functions to quadratics, we fit the functions describing the upper and lower end of the interval to quadratic functions. For each city, this yields a set of two coefficients $c_1$ on the squared term of the quadratic corresponding to the upper and lower ends. We generate Gaussian distributions with means given by the average of the two coefficients and standard deviation given by the difference of the two coefficients divided by 1.96. In implementing the bootstrap, we simply generate a quadratic function with coefficient $c_1$ drawn from this distribution subject to the same constraints as discussed above for the point forecast with the additional constraint that RR may not fall below unity.

The bootstrap is repeated 9,999 times, and for each day we pick 97.5 and 2.5 quantiles from the 10,000 iterations and the point forecast of RR for each city. We report the geometric mean of 360 days for each year. The intervals are illustrated by the dashed lines in Figure 4.
References

Acknowledgements

The authors are extremely grateful to Antonio Gasparrini for sharing his estimated dose response functions and to participants of the 2018 Econometric Models of Climate Change Conference (Rome) for useful comments.

Author Contributions Statement

Both authors conceived the study, discussed the methodology, results, and implications, and contributed to writing the manuscript at all stages. J.I.M. managed the data and designed and performed the statistical analyses.

Additional Information

Competing Financial Interests

The authors declare no competing financial interests.
Figure 1: RCPs and QCPs. Four RCPs with corresponding QCPs constructed by weighting RCPs and AR5 scenarios. QCP6.77 and QCP4.96 represent baseline Paris Forever (3.5°C) and Paris Plus (2.5°C).

Figure 2: Average June-August Temperatures and Temperature Increases. Top panel: Average June-August temperatures over 1971-2000. Middle panel: Projected temperature increases in the average June-August temperatures from 1971-2000 to 2071-2100 under Paris Plus (2.5°C/QCP4.96). Bottom panel: Projected temperature increases in the average June-August temperatures over 2071-2100 between Paris Plus (2.5°C/QCP4.96) and Paris Forever (3.5°C/QCP6.77).
Figure 3: Average Annual Temperatures. HadGEM2-ES ensemble bias-corrected historical temperatures and 2031-2100 trajectories under Paris Forever (3.5°C/QCP6.77) and Paris Plus (2.5°C/QCP4.96) for 1.25° by 1.875° grid boxes corresponding Toronto, Beijing, Rome, Tokyo, Seoul, Madrid, Stockholm, Taipei, Bangkok, London, New York City, and Chicago.
Figure 4: Average Annual Temperatures. Average daily risk from heat exposure over 2031-2100 under Paris Forever (3.5°C/QCP6.77) relative to no heat exposure and average daily risk from heat exposure over 2031-2100 under Paris Plus (2.5°C/QCP4.96) relative to that under Paris Forever (3.5°C/QCP6.77) for Toronto, Beijing, Rome, Tokyo, Seoul, Madrid, Stockholm, Taipei, Bangkok, London, New York City, and Chicago. Dashed lines show 95% forecast intervals, calculated as described in the Methods section.
Bias Correction Procedure

Ivanov et al.\(^1\) and other authors emphasize the need to bias-correct global climate model output data, such as that which we employ. Moreover, Knutti et al.\(^2\) suggest using model ensembles yet criticize the common practice of using equal weights for each model in the ensemble. We employ a bias-correction method that addresses both of these issues while maintaining the variability of the instrumental record.\(^3\)

Our bias-correction methodology is accomplished as follows. For each 5° latitude band in the Northern Hemisphere, we regress monthly instrumental HadCRUT4 temperatures over 1961-1990, its base period, onto an ensemble of CMIP5 HadGEM2-ES model runs over the same period and a set of monthly binary variables. Coefficients on the binary variables reflect seasonality that may differ across latitude bands. Prior to this regression, we aggregate across longitudes for two reasons: the longitudes of the instrumental record and model output do not line up evenly and aggregating the base period of the HadCRUT4 anomalies across longitudes should help to smooth out uncertainties in absolute temperature measurements over the base period.

The regression just described accomplishes not only a bias correction to each member of the model ensemble but a projection based on the regression is a weighted average of the different ensemble members that optimally approximates the temperature record over the base period in the spirit of Knutti et al.\(^2\) Using an average creates a new problem, however. Even with the variability from seasonality preserved by the binary variables, averaging the ensemble members artificially smooths the variability in the data, which may severely underestimate the probability of observing extreme temperatures.

To correct for the smoothness, we first calculate the \(R^2\) from a regression of the seasonally adjusted instrumental record onto the seasonally adjusted model ensemble members. The \(R^2\) captures the proportion of nonseasonal variation in the instrumental record that is explained by the weighted average of the model ensemble members. We seasonally adjust the weighted average – i.e., the projection from the first regression, then we divide it by the square root of the \(R^2\) from the second regression, and finally we “re-seasonalize” by adding back in the seasonal adjustment. The inflation by \(1/R\) matches the nonseasonal variation
in the bias-corrected model output to that in the instrumental record as suggested by Otto et al.\textsuperscript{3}

Once the weights (regression coefficients), inflation factor $1/R$, and seasonal adjustment are calculated, the bias correction may be conducted over any time period and spatial resolution for which we have model output and is not limited to that for which we have historical data.

Our bias correction procedure yields model output with variability roughly matching that of the instrumental record over the base period on both annual and monthly time scales, as illustrated by Figure S1. The left panels compare monthly seasonally adjusted bias-corrected model output with the monthly seasonally adjusted instrumental record averaged across longitudinal within four latitude bands in the Northern Hemisphere. The mean and variance of the data appear to be similar. The right panels similarly compare monthly “reseasonalized” bias-corrected model output with the monthly instrumental record. The fits are excellent, but deteriorate close to the equator where seasonal temperature differences are not as sharp.

**Emulation Procedure**

The emulation procedure takes annual global QCPs as inputs and projects monthly spatially disaggregated temperatures. Castruccio et al.\textsuperscript{4} (equation 1) model the relationship between (disaggregated) temperature and CO$_2$ concentrations as

$$T_t = \beta_0 + (\beta_1/2)[(h_t - h_{pre}) + (h_{t-1} - h_{pre})] + \beta_2 \sum_{k=2}^{\infty} \rho^{-2}(1-\rho)\rho^k (h_{t-k} - h_{pre}) + e_t$$

which may be rewritten as

$$T_t = \beta_0 - (\beta_1 + \beta_2)h_{pre} + \sum_{k=0}^{\infty} \varphi_k h_{t-k} + e_t$$

with $\varphi_0 = \varphi_1 = \beta_1/2$ and $\varphi_k = \beta_2\rho^{-2}(1-\rho)\rho^k$ for $k \geq 2$. The linear process may be decomposed as

$$T_t = \beta_0 - (\beta_1 + \beta_2)h_{pre} + \left(\sum_{k=0}^{\infty} \varphi_k\right) h_{t} - \sum_{k=0}^{\infty} \left(\sum_{s=k+1}^{\infty} \varphi_s\right) \Delta h_{t-k} + e_t \quad (1)$$

following Phillips and Solo.\textsuperscript{5}

We assume that the series of CO$_2$ concentrations follows an ARIMA(1,1,0) with drift, so that

$$\Delta h_t = \tilde{\tau} + \sum_{s=0}^{\infty} \tau^s u_{t-s} \quad (2)$$
where $\bar{\tau}$ is the historical growth rate of about 0.04% per year and the second term captures stochastic and serially dependent fluctuations around the historical growth rate. A strand of the literature\textsuperscript{6,7} debates the deterministic vs. stochastic nature of long-run trends in temperatures and CO\textsubscript{2} concentrations. The proof presented below considers both cases: $\bar{\tau} = 0$ (stochastic trend dominates) and $\bar{\tau} \neq 0$ (deterministic trend dominates).

Substituting equation (2) into (1) allows

$$T_t = \alpha_0 + \alpha_1 h_t + \eta_t$$

where

$$\eta_t = e_t - \sum_{k=0}^{\infty} \left( \sum_{s=k+1}^{\infty} \varphi_s \right) \left( \sum_{s=0}^{\infty} \tau^s u_{t-k-s} \right),$$

$$\alpha_1 = \sum_{k=0}^{\infty} \varphi_k = \beta_1 + \beta_2,$$

and

$$\alpha_0 = \beta_0 - (\beta_1 + \beta_2) h_{pre} - \bar{\tau} \left( \frac{1}{2} \beta_1 + \frac{2}{1 - \rho} \beta_2 \right)$$

after some algebra.

**Proposition.** Assume that $(e_t)$ and $(u_t)$ are martingale difference sequences, that $|\tau|$ and $|\rho|$ are less than unity, and that $h_0 = O_p(1)$. Let $\kappa_T = T$ if $\bar{\tau} = 0$ and $\kappa_T = T^{3/2}$ if $\bar{\tau} \neq 0$. The least squares estimators $\hat{\alpha}_0$ and $\hat{\alpha}_1$ of the regression coefficients in (3) are consistent with

$$T^{1/2}(\hat{\alpha}_0 - \alpha_0) = O_p(1) \quad \text{and} \quad \kappa_T(\hat{\alpha}_1 - \alpha_1) = O_p(1)$$

as the sample size $T \to \infty$.

**Proof.** Using (2) and defining $z_t = \sum_{s=1}^{t} \sum_{k=0}^{\infty} \tau^k u_{t-k} - \sum_{s=1}^{t} \tau^s u_{t-s} - \sum_{s=1}^{t} \tau^s u_{t-s}$, write $h_t = h_0 + \bar{\tau} t + z_t$. Under the given assumptions $\sum_k k |\rho^k| < \infty$, so that the law of large numbers (LLN) and central limit theorem for linear processes applies to $(\eta_t)$ and the functional central limit theorem for linear processes applies to $(z_t)$\textsuperscript{5}.

Start with $\hat{\alpha}_1$ and note that $\hat{\alpha}_1 - \alpha_1 = B_T^{-1} A_T$, where

$$A_T = \sum_t h_t \eta_t - T^{-1} \sum_t h_t \sum_t \eta_t$$

and

$$B_T = \sum_t h_t^2 - T^{-1} \left( \sum_t h_t \right)^2.$$

Right away, we can see that $\sum_t \eta_t = O_p(T^{1/2})$ due to the central limit theorem.

Suppose first that $\bar{\tau} = 0$. It follows from the functional central limit theorem and fairly
standard asymptotic theory that

\[ [a] \sum_t h_t \eta_t = \sum_t (h_0 + z_t) \eta_t = O_p(T), \]

\[ [b] \sum_t h_t = \sum_t (h_0 + z_t) = O_p(T^{3/2}), \]

\[ [c] \sum_t h_t^2 = \sum_t (h_0 + z_t)^2 = O_p(T^2). \]

Consequently, \( A_T = O_p(T) \) and \( B_T = O_p(T^2) \), so that \( \hat{\alpha}_1 - \alpha_1 = O_p(T) \), and the stated result holds.

Suppose instead that \( \bar{\tau} \neq 0 \), so that the drift dominates the stochastic trend. Then

\[ [a] \sum_t h_t \eta_t = \sum_t (h_0 + \bar{\tau} t + z_t) \eta_t = O_p(T^{3/2}), \]

\[ [b] \sum_t h_t = \sum_t (h_0 + \bar{\tau} t + z_t) = O_p(T^2), \]

\[ [c] \sum_t h_t^2 = \sum_t (h_0 + \bar{\tau} t + z_t)^2 = O_p(T^3), \]

so that \( A_T = O_p(T^{3/2}) \) and \( B_T = O_p(T^3) \). The stated result clearly holds for this case.

Now, for the intercept,

\[
T^{1/2}(\hat{\alpha}_0 - \alpha_0) = T^{-1/2} \sum_t (T_t - \hat{\alpha}_1 h_t) = -\kappa_T (\hat{\alpha}_1 - \alpha_1) \left( \kappa_T^{-1} T^{-1/2} \sum_t h_t \right) + T^{-1/2} \sum_t \eta_t
\]

holds. In either case, \( \bar{\tau} = 0 \) or not, this is \( O_p(1) \) by the asymptotics already shown. \( \square \)

The proposition establishes the consistency of the estimators of the long-run parameters \( \alpha_0 \) and \( \alpha_1 \). This strategy does not identify the parameters of the model of Castruccio et al., but it identifies the long-run relationship between log CO\textsubscript{2} concentrations and temperature, which is sufficient for our purposes. The assumptions that \((e_t)\) and \((u_t)\) are martingale difference sequences are sufficient, but not necessary. They could just as easily be defined by applying absolutely summable linear filters to martingale difference sequences to allow for (additional) serial correlation.

We can stack locations indexed by latitude and longitude into a long vector and run so-called seemingly unrelated regressions. It is well known\textsuperscript{8} that with a common regressor \( h_t \), the regressions can be estimated separately or as a system, even though the errors are correlated across equations – i.e., spatially correlated. This strategy is not unlike pattern scaling,\textsuperscript{9} except disaggregated temperatures are mapped to log CO\textsubscript{2} concentrations rather than to global temperatures.
We take the regression one step further, stacking $T_t$ over both locations on the globe and months within a year to create $\bar{T}_t$. We can then estimate

$$\bar{T}_t = \bar{\alpha}_0 + \bar{\alpha}_1 h_t + \bar{\eta}_t$$

(4)

where $\bar{\eta}_t$ is a similarly stacked error term and $\bar{\alpha}_0$ and $\bar{\alpha}_1$ are similarly stacked parameter vectors. This regression maps an annual global log CO$_2$ concentration $h_t$ to a vector of spatially and temporally disaggregated temperatures.

To estimate (4), we employ CMIP5 HadGEM2-ES monthly historical temperatures similarly to Liu et al.$^{10}$ but with the bias correction described above and RCP CO$_2$ concentrations over 1860-2004. An important consideration is the legitimacy of the regression in (3) for these data. Our empirical evidence supports the notion that $h_t$ follows an ARIMA with drift, so that $T_t$ for each of the 165,888 [$=(192 \times 72 \text{ locations}) \times (12 \text{ months})$] elements must also follow an ARIMA with drift and be cointegrated with $h_t$.

Are they? A conceptually straightforward approach to answer this question would be to run a test for the number of cointegrating vectors in $(h_t, T_t')$. However, because this approach would involve inverting a 165,889 $\times$ 165,889 matrix, it is not practical. A more practical approach would be a panel cointegration test of the regression in (4), but the obvious presence of cross-sectional dependence of the residuals defeats the most common panel unit root tests. The residual-based test of Chang and Nguyen$^{11}$ is robust to this type of cross-sectional dependence. However, they note a deficiency when the panel has common stochastic or deterministic trends across cross-sectional units, as is the case here with a single global $h_t$ due to well-mixed CO$_2$ concentrations. In this case, they suggest creating instruments using odd power Hermite polynomials. This, too, would be impractical with 165,888 cross-sectional units.

A feasible way to get at the issue of cointegration, which is similar in spirit to the residual-based panel tests of the literature is to calculate 165,888 residual-based cointegration tests. The maximum of these Engle-Granger-type tests turns out to be $-38.43$ from an autoregressive coefficient of 0.73. Using a standard 5% critical value given by $-27.07^{12}$ we easily reject the null of no cointegration for every single month at every single location. Substituting individual tests for a joint test in this way is crude, but given that such a joint test does not seem feasible, the preponderance of evidence from the individual tests suggests that the regression in (4) is reasonable.

Given point estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ and bootstrap estimates $\tilde{\alpha}_0$ and $\tilde{\alpha}_1$ of $\bar{\alpha}_0$ and $\bar{\alpha}_1$, temperature forecasts are given by $\tilde{T}_{T+k} = \tilde{\alpha}_0 + \tilde{\alpha}_1 q_{T+k}$, where $q_{T+k}$ is the QCP in year $T + k$. See the Methods section for a discussion of the bootstrap.
SOM References


Additional Table and Figure

<table>
<thead>
<tr>
<th>City</th>
<th>$MMT$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toronto</td>
<td>18.9</td>
<td>0.002</td>
<td>-0.071</td>
<td>1.667</td>
</tr>
<tr>
<td>Beijing</td>
<td>25.5</td>
<td>0.012</td>
<td>-0.627</td>
<td>8.989</td>
</tr>
<tr>
<td>Rome</td>
<td>22.2</td>
<td>0.010</td>
<td>-0.445</td>
<td>5.939</td>
</tr>
<tr>
<td>Tokyo</td>
<td>26.5</td>
<td>0.008</td>
<td>-0.407</td>
<td>6.392</td>
</tr>
<tr>
<td>Seoul</td>
<td>24.5</td>
<td>0.008</td>
<td>-0.382</td>
<td>5.684</td>
</tr>
<tr>
<td>Madrid</td>
<td>21.9</td>
<td>0.008</td>
<td>-0.347</td>
<td>4.803</td>
</tr>
<tr>
<td>Stockholm</td>
<td>18.9</td>
<td>0.004</td>
<td>-0.169</td>
<td>2.594</td>
</tr>
<tr>
<td>Taipei</td>
<td>26.0</td>
<td>0.005</td>
<td>-0.239</td>
<td>4.113</td>
</tr>
<tr>
<td>Bangkok</td>
<td>29.9</td>
<td>0.018</td>
<td>-1.054</td>
<td>16.764</td>
</tr>
<tr>
<td>London</td>
<td>19.5</td>
<td>0.014</td>
<td>-0.552</td>
<td>6.385</td>
</tr>
<tr>
<td>New York City</td>
<td>23.1</td>
<td>0.006</td>
<td>-0.272</td>
<td>4.143</td>
</tr>
<tr>
<td>Chicago</td>
<td>24.7</td>
<td>0.011</td>
<td>-0.539</td>
<td>7.655</td>
</tr>
</tbody>
</table>

Table S1: Estimated Dose Response Functions. The quadratic function $c_1 z^2 + c_2 z + c_3 = 0$ with $c_2 = -2c_1 MMT$ and $c_3 = 1 + c_1 MMT^2$ is fit to the right tail of the function estimated by Gasparrini et al.\textsuperscript{13} for each city. $MMT$ is given by those authors and the table shows parameters estimated by the present authors using nonlinear least squares. Using these parameter estimates, the dose response function for heat stress at each location is given by $c_1 z^2 + c_2 z + c_3$ if $z \geq MMT$ and 1 otherwise, which may be extended for any temperature.
Figure S1: Instrumental and Bias-Corrected Temperatures by Latitude 1961-1990.
Solid blue: bias-corrected model temperatures; dotted red: instrumental temperature record. Left panels: seasonally adjusted; right panels: re-seasonalized. From top to bottom, 1.25° latitude bands centered on 68.175°N, 45.675°N, 23.175°N, and 0.675°N.