Peer Review File

Manuscript Title: The economic commitment of climate change

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

The manuscript used the gross-regional product per capita (GRPpc) data in 1600+ subnational regions worldwide to first establish the empirical response functions to regional climate variations during 1979-2019 and then estimate future economic damages from CMIP6 climate projections toward 2100. The authors concluded that the world economy is projected for 21% income reductions in the next 25 years due to anthropogenic climate impacts and these damages already outweigh the mitigation costs to limit global warming in the coming decades by large factors. They further showed that the largest losses are projected in regions with lower cumulative historical emissions and lower present-day income. As they discussed, these estimates are larger than previous studies because of taking first-difference of climate variables with sufficient lags and more components. The result is new and should be interesting to the community of climate impact on economy and Nature readers in general.

However, I have a major concern on the uncertainty and validity of the empirical climate-economic response functions (hereafter simply called "model") they built and used for projections. The authors claimed that their model represents "empirically validated" impacts on economic output, which is very misleading since validation against actual observations (not available) of such climate-economic relationships is not possible and was not done in this manuscript. Regressions built upon correlations between two sets of variables are not "validated", since one can find spurious correlations between anything.

I do appreciate the authors took extra steps (than previous studies) in choosing the climate variables with additional components and more lags. But doing so may not necessarily lead to a more realistic model. Here, overfitting becomes a critical problem. Table S1 lists the regression results, from which I interpret the numbers without parenthesis (they called "effect") as the regression coefficients α_(i,L) in equation (4). [It would be more appreciated if this definition is made clearly, and each variable's name matches the mathematical notion in that equation without lengthy redefinition in the table description.] Many regression coefficients [or the "effects" defined] are not statistically significant, especially for those key components such as average temperature, whose change is concluded as the "predominant" factor for committed damages. In addition, it is essential to identify how multilinearities among the climate variables affect the result, since auto and cross correlations are often present. Lacking significance in the estimated coefficients and including multilinearities in the overall regression model will invalid the model, especially for identification of relative contributions among the variables.

I also appreciate the authors included model's fitting statistical measures (R2, WR2, AIC, and BIC) in Table S1. Ionically, these important measures have not been discussed in any text, even without noting their actual meaning in the table caption. The R2 value is 0.272, which indicates that the model can explain 27.2% of the total GRPpc variance. The authors did not clarify what this variance is since the model is built at each subnational region. If it is referred to the total spatiotemporal panel or plus the linear trend term (k_r y), the variance explained by the climate components would be quite small at each region. I am not quite sure what WR2 (within-region R2) will tell us. I guess AIC is Akaike Information Criterion, while BIC is Bayesian Information Criterion. Both AIC and BIC values are important scores indicative of model overfitting information.

But they are meaningful only in the relative term, comparing different possible realizations of the regression model. That is, a model realization with a lower AIC or BIC value is a better choice for less overfitting. Therefore, they should be listed for all models individually built through stepwise regressions by adding one term a time in equation (4). These AIC and BIC values for individual realizations will help select the best model that explains the largest GRPpc variance using the smallest number of climate variables. It is also critical to separate the interannual from spatial variances since the final model is used at each region. Thus, the credibility of each regional model lies in its ability to capture the interannual variability, from which the climate change may be used to project future economic damage.

Extended Data Figure 1 clearly indicates a serious problem of the model built. The existence of large oscillations in the panel regression effects among different lags is especially a concern. Among others, it may indicate strong autocorrelations built into the model. Using the first-difference ($\Delta F_y=F_y-F_{(y-1)}$) may further amplify the effect of autocorrelation. [Note that I don't object the use of ΔF_y, which is a good way to remove technology-related trends (for example).] This is why it is vital to consider stepwise regressions and discriminate the models that have larger possibility of overfitting as discussed above. [Minor comments: in equation (4), the seasonal temperature difference $\hat{T}_r$ is not defined, and the formulation for specific nonlinear terms such as $\bar{T}_r \cdot \Delta Pext_{(r,y-L)}$ is not explained.]

Since regional characteristics are one of the main focuses in this study, it is important to show the geographical distributions of R2 as well as AIC and its increment to the next competitor δAIC. These scores will quantify the model performance (how much the select climate components explain the total economic variance) and the model validity (capturing the largest amount of variance with the least number of climate variables). Without considering such information, the model so built may likely suffer from overfitting problems and contain large uncertainties.

The calculation for future projection is not clear. Lines 435-460 describes the procedure. The question is how $\delta_{(r,y)}$ and $\pi_{(r,y)}$ in equation (5) are related to $g_{(r,y)}$ in equation (4) and how ΔF_y is calculated in climate change scenarios. Typically, climate changes are calculated as the differences of the future-scenario minus present-day simulations form a same GCM averaged in a specified duration. This difference method removes, among others, the effect of the GCM's systematic climate biases (which usually are substantial regionally). In this case, the calculation for ΔF_y is important – is it defined still as the first-difference in both the present-day and future-scenario climate simulations (then how to count their change) or directly the change between the two simulations (then how to justify the replacement of the yearly tendency ΔF_y with a much larger step climate change)? Another method could be to calculate $g_{(r,y)}$ as in equation (4) separately for the present-day and future-scenario climate simulations from the same GCM and then calculate their difference. In this case, the performance of the present-day $g_{(r,y)}$ distribution over the globe should be quantified by comparing those based on individual GCMs' climate simulations and their ensemble against the baseline calculated from observed climate conditions. In either case, the applicability of the model for future projection should be addressed.

The rest of the manuscript is focused on discussing the results from the regression $g_{(r,y)}$ model calculations based on CMIP6 SSP2-RCP2.6 and SSP5-RCP8.5 climate projections made by 21 GCMs. These discussions are mostly show-and-tell, except for two interesting points: 1) the projected damages are compared with the currently committed mitigation costs; and 2) the injustice of committed climate damages is highlighted (Figure 3). Given that the regression model so built is questionable (as commented above) for its validity in capturing the "observed" economic-climate relationships and its applicability in projecting the climate change impacts, I cannot see how substantial uncertainties their projected economic damages are involved. As such, I cannot evaluate the robustness, validity, and reliability of the conclusions by this study.

Referee #2 (Remarks to the Author):

Comments on The economic commitment of climate change

This is an interesting paper and one of a growing literature that explores historic responses of the economy to climate change and then uses the estimated coefficients to project likely future impacts. As such the paper is well written and provides a useful and interesting addition to this literature. The authors suggest that there are two key contributions – one is the statistical innovation of using lags in a more complex econometric specification than is available in the literature and the other is in the findings that the near term cumulative effects are significant and so warrant immediate attention. To many informed readers that is a conclusion, which does not require reemphasizing.

It is hoped that the following comments are of utility to the authors in revising this paper.

A useful contribution of this paper is to include lags of the GDP-pc growth term to determine whether there is a reversal from the initial effects of climate change variables. Unsurprisingly the finding is that lags suggest a reversion to previous growth rates and hence results are not as exaggerated as in some specifications that do not include lags. However the authors may wish to revisit their explanation of the rationale that they provide. Especially in lines 59 to 68, it is unclear how a specification that begins with first differences without lags, and then tests for lags of first differences implies a baseline "…null hypothesis of purely level effects….[with] … instantaneous effect on …growth…"(line 62). Then by line 68 there is a correction with the recognition that the baseline without lags "… has an infinitely persistent effect on .. growth…". It cannot be both and the former assertion seems to be plainly wrong.

Secondly, it is somewhat difficult to comprehend the full rationale for the particular econometric specification that is used. Since economic theory does not dictate the form of the specification, it is necessary for empirical work to show that the preferred estimating equation is robust and makes sense. Hence for instance, it is unclear why the regressions are of the form $\alpha_t \Delta x_{(i,t-L)}+\beta x_t \Delta x_{(i,t-L)}$ instead of say $\alpha_t \Delta x_{(i,t-L)}+\beta \Delta x_t \Delta x_{(i,t-L)}$ which also seems plausible. More generally it is notable that there are no regression tables provided or robustness tests shown to determine the sensitivity of the results to plausible changes in specification. IN sum its impossible to tell how robust these results are to small or large changes in specification. Nor is the current specification and the interaction terms well justified.

The authors may also wish to revisit the interpretation of their findings and present further tests to support some of the conclusions. If, as is often alleged, the climate variables for heat and precipitation are correlated in this data set then multicollinearity would suggest that the estimated coefficients no longer provide an accurate estimate of the marginal impacts of each of the weather variables used. While collective impacts would be accurate – one cannot unambiguously identify the marginal effects of any one of the correlated explanatory variables. This issue could be addressed by providing a correlation matrix to facilitate interpretation.

Finally it may be helpful for this paper not to accompany the often hyperbolic narratives in some of the literature. The results (if robust) suggest a 21% reduction in gdp over 25 – 30 years. This amounts to a rather modest 0.7% reduction in gdp each year, which would be well within the margin of statistical error for gdp revisions conducted by statistical authorities. In other words though not a trivial impact, the effects are not as large as the tone of some of the text suggests.

Referee #3 (Remarks to the Author):

Review of manuscript #2023-01-01329 entitled 'The economic commitment of climate change'.

A. Summary of the key results:
Kotz et al. use several metrics of historical and simulated climate change along with socioeconomic data to make projections of the economic damages associated with future climate change. In particular, they assess the economic damages at the time when projections from a low-emissions scenario diverge from that of a high-emissions scenario. The authors find that by this mid-century point of divergence, the world economy will experience a 21±8% reduction in population-weighted income, vastly outweighing projected mitigation costs. Lastly, they break down the results by a variety of climate metrics at sub-national levels to better understand the drivers of change.

B. Originality and significance:
This study differs from past work through a variety of methodological changes, which lead to much larger estimates of economic damages. The work seems novel, but I am not overly familiar with past research on the economics of climate change. If proven to be robust, the conclusions of this work have broad significance that would warrant publication in a high-profile journal such as this one.

C. Data & methodology:
The data seems appropriate for this type of study. I would like to see the climate data described in slightly more detail.

D. Appropriate use of statistics and treatment of uncertainties:
There are a couple of instances (discussed below) where the use of statistics can be improved.

E. Conclusions:
The conclusions seem robust, but it is difficult to fully gauge how sensitive the results are to a variety of methodological choices at this time.

F. Suggested improvements:
There are several minor changes that can be made to improve the manuscript as discussed below.

G. References:
The references are mostly appropriate.

H. Clarity and context:
The paper is very well written, but I found several instances where aspects of the analysis could be better described.

Major comments

Sensitivity of results to methods: It is somewhat difficult to fully gauge the robustness of the results when there are several seemingly arbitrary methodological choices being made. It would be helpful to show how sensitive the results are to some of these choices. For example:
L438-440: How sensitive are the results (e.g. time of divergence between the emissions scenarios) to the use of 30-year moving averages. This seems like a conservative choice, which would delay the point when the two scenarios are distinguishable from one another.
Fig S4 (L158-160): I am also having a difficult time grasping why the changes need to be framed in the context of their historical interannual(?) variability. Please clarify. Additionally, 2005-2035 seems like an odd choice of baseline climate to calculate change metrics from. I recommend comparing with the historical period from which the variability metrics are calculated across.

Fig 3 Analysis: This is the weakest component of the manuscript in my opinion. It is only briefly touched on, which begs the question whether it warrants a place in the main text. I recommend adding a bit more detail here. Moreover, there should be at least some basic statistics shown to support the statements made in this paragraph (e.g. correlation between metrics).

Wording: the use of “committed” throughout the manuscript could be misconstrued when interpreting the meaning of these results. This seems to imply the economic costs that are already “baked in” because of past emissions, when in reality it is the minimum cost associated with continued emissions (at a lower rate: RCP2.6). I think this needs to be clearly stated early in the text to limit confusion.

Minor comments:
L54: Cite and briefly discuss the findings of past assessments (e.g, Extended Data Table 1).
L57-59: One more sentence here briefly describing persistence would help non-specialists.
L93: Include a supplemental table showing the models used.
L94: need a citation for CMIP6 (e.g. Eyring et al. 2016).
L104: change to “without climate change impacts”
L105: You should still define what the IPCC likelihood means.
L108: Missing period after (Fig 1)
L109: define pure growth or pure level effects in the methods.
Fig 1: Make the 0 y-axis line solid.
L134: remove ‘already’
L140: state an approximate dollar cost for mitigation.
L142: Why does this IAM seem like such an outlier?
L155: Can these changes be roughly quantified? For example, state the average contribution from each of the metrics assessed.
L155: Add a companion supplemental figure that shows the regular climate change metrics (unlike Fig S4). This would really help the reader to understand for example, how much warming or precipitation change is associated with this level of economic damage (e.g. the change metrics by 2048).
L161: Add some measure of uncertainty to this estimate.
L163: State which variable’s inclusion has the biggest impact as this seems important to know for future studies.
L165: Briefly explain why this is.
L174: I would think that the change in frequency of events exceeding an extreme precipitation threshold would be more relevant than the magnitude change above said threshold. Why is magnitude change a preferable metric here?
L177- Frequently citing the wrong figure in this paragraph.
L180: change to “Damages due to increasing mean temperature…”
L182: This is despite increases in temperature being much larger across high latitudes, which should be briefly noted.
L185: Why is there such a contrast between Europe and the US (at a lower latitude)?
L184-186: rephrase this sentence. Unclear if referring to the temperature variability change or damages.
L190: this is very difficult to see given the size of panels d-f.
L193: could add another reference here.
L210: this should say “national cumulative emissions per capita”
L219: Put their estimate in brackets here?
L334-335: Precipitation from reanalysis can be poor, state what bias correction is done to this product over land.
L393-397: It doesn’t seem like this would account for very large administrative boundaries where the climate changes where people live may not reflect the area average (e.g. much of Canada).

Author Rebuttals to Initial Comments:

Response to referee #1

The manuscript used the gross-regional product per capita (GRPpc) data in 1600+ subnational regions worldwide to first establish the empirical response functions to regional climate variations during 1979-2019 and then estimate future economic damages from CMIP6 climate projections toward 2100. The authors concluded that the world economy is projected for 21% income reductions in the next 25 years due to anthropogenic climate impacts and these damages already outweigh the mitigation costs to limit global warming in the coming decades by large factors. They further showed that the largest losses are projected in regions with lower cumulative historical emissions and lower present-day income. As they discussed, these estimates are larger than previous studies because of taking the first-difference of climate variables with sufficient lags and more components. The result is new and should be interesting to the community of climate impact on economy and Nature readers in general.

We thank the reviewer for their comprehensive assessment of our work, both their positive comments and highly constructive criticisms.

However, I have a major concern on the uncertainty and validity of the empirical climate-economic response functions (hereafter simply called "model") they built and used for projections. The authors claimed that their model represents "empirically validated" impacts on economic output, which is very misleading since validation against actual observations (not available) of such climate-economic relationships is not possible and was not done in this manuscript. Regressions built upon correlations between two sets of variables are not "validated", since one can find spurious correlations between anything.

We appreciate the detailed comments which the reviewer provided regarding our empirical model. In revising the manuscript, we have undertaken extensive robustness tests focussed in particular on addressing issues relating to auto-correlation, multicollinearity and lag selection while avoiding overfitting the data. These tests are outlined below in reference to the reviewer's specific concerns. Moreover, we agree that our description of these empirical models as "validated" is misleading and have since altered this wording throughout the updated manuscript.

I do appreciate the authors took extra steps (than previous studies) in choosing the climate variables with additional components and more lags. But doing so may not necessarily lead to a more realistic model. Here, overfitting becomes a critical problem. Table S1 lists the regression results, from which I interpret the numbers without parenthesis (they called "effect") as the regression coefficients α_(i,L) in equation (4). [It would be more appreciated if this definition is made clearly, and each variable's name matches the mathematical notion in that equation without lengthy redefinition in the table description.] Many regression coefficients [or the "effects" defined] are not statistically significant, especially for those key components such as average temperature, whose change is concluded as the "predominant" factor for committed damages. In addition, it is essential to identify how multilinearities among the climate variables affect the result, since auto and cross correlations are often present. Lacking significance in the estimated coefficients and including multilinearities in the overall regression model will invalid the model, especially for identification of relative contributions among the variables.

We appreciate the concerns related to the selection of the number of lags, in particular in relation to imperfect multi-collinearity arising from auto- and cross-correlated climate variables and also in relation to overfitting in general. We now provide a more detailed lag selection procedure and set of robustness tests in the updated manuscript, including extensive additional Monte-Carlo simulations to assess the robustness of the models to (imperfect) multicollinearity arising from auto- and cross-correlation in the lagged climate variables, and an incremental model selection based on observed significance of terms as well as Information Criteria to avoid overfitting (see newly added Supplementary Methods Sections S1-S3 as well as newly added figures S1-S7).

To answer the questions directly: First, with regards to the reviewer's concerns that the regression coefficients for many terms are not significant in Table S1 of the SI: We note that when interpreting the significance of these coefficients it is important to consider together both, the term relating to the main independent climate variable as well as that of its interaction term (e.g. of annual mean temperature and the interaction of annual mean temperature with its historical average; in tables S1-4 the relevant terms are shown in adjacent rows). Extended Data Figure 1 of the main manuscript provides the opportunity for an assessment of the combined significance of these terms by plotting the estimated marginal effects of the main independent variables and their confidence intervals. This shows that significant marginal effects are present at all lags at least up to 8 lags for temperature variables and up to 4 years for precipitation variables. Indeed it was observing the significance of such effects that led us to select a model with 8 and 4 lags for temperature and precipitation variables in the original manuscript. A note on the necessity of this form of interpretation is now made more explicit in the caption of Table S1 in reference to ED Fig. 1, as well as having updated the variable names to reflect those in the equations of the methods section, and explicitly referred to the regression coefficients to which the numbers in the table refer.

With regards to the issues of (imperfect) multicollinearity arising from auto- and cross-correlations, we have provided extensive robustness tests which address these concerns. First, in Fig. S1 we explicitly assess the extent of such auto- and cross-correlations in the climate data, finding stronger auto-correlations in all climate variables at a lag of one year and weaker auto-correlations for more distant lags, as well as strong cross-correlations between precipitation variables in particular.

We proceed to test whether the presence of such correlations influences the accuracy and precision of our empirical models using two sets of Monte Carlo simulations. These are explained in detail in the new supplementary Methods, specifically sections S1 and S2, and shown in Figs. S2-S4. We refer to these additional tests in the updated main manuscript on L.90-103.

These simulations follow and develop upon the methodology introduced in the supplementary appendix of Dell et al., 2012. Ensembles of artificial data-sets are generated by randomly re-assigning climate data to different regions, hence preserving the auto- and cross-correlative structure of the climate data. Artificial effects of known size and persistence are then added to the data and the empirical models are then tested for their ability to detect and quantify the known effects. Results indicate the following conclusions:

1. Auto-correlations introduce negligible systematic bias or imprecision in estimates of the effects (Figs. S2 & S3).
2. Including an insufficient number of lags can systematically underestimate the effect of a climatic change (a consequence of the conservative nature of our specification using first differences, which avoids systematic overestimation when too few lags are included, see L59-85 of the manuscript for a discussion of this choice which we now also frame as a “robust-lower bound on impact persistence”).
3. Including more lags than necessary can however increase imprecision. (Fig. S2 & S3)
4. Information criteria only provide helpful guidance in lag selection when starting from a large number of lags and decreasing. (Fig. S3)
5. Cross-correlations only introduce systematic bias in the estimated effects when climate variables are not included simultaneously in the regressions. Accounting for all climate variables simultaneously is necessary to recover the true effect and relative contributions of individual variables. (Fig. S4).

In combination with further robustness tests and model selection procedures outlined below in response to your additional comments, we believe these new results strongly support our empirical approach and overall results.

I also appreciate the authors included model’s fitting statistical measures (R2, WR2, AIC, and BIC) in Table S1. Ionically, these important measures have not been discussed in any text, even without noting their actual meaning in the table caption. The R2 value is 0.272, which indicates that the model can explain 27.2% of the total GRPpc variance. The authors did not clarify what this variance is since the model is built at each subnational region. If it is referred to the total spatiotemporal panel or plus the linear trend term (k_r y), the variance explained by the climate components would be quite small at each region. I am not quite sure what WR2 (within-region R2) will tell us. I guess AIC is Akaike Information Criterion, while BIC is Bayesian Information Criterion. Both AIC and BIC values are important scores indicative of model overfitting information. But they are meaningful only in the relative term, comparing different possible realizations of the regression model. That is, a model realization with a lower AIC or BIC value is a better choice for less overfitting. Therefore, they should be listed for all models individually built through stepwise regressions by adding one term a time in equation (4). These AIC and BIC values for individual realizations will help select the best model that explains the largest GRPpc variance using the smallest number of climate variables. It is also critical to separate the interannual from spatial variances since the final model is used at each region. Thus, the credibility of each regional model lies in its ability to capture the interannual variability, from which the climate change may be used to project future economic damage.

We thank the reviewer for these useful comments and apologize for some lack of clarity regarding terms in the original manuscript. The definitions of within-region R2 has now been made clearer in the regression table S1, noting that it describes the variance explained along the temporal dimension of interest. One should not be surprised that these values are small (<5%) given that a large portion of temporal variance in economic output is influenced by non-climatic factors (see also the within region R2 in previous publications such as Dell 2012, and Burke 2015). Furthermore, it is clear from our Monte Carlo simulations that a small R-squared does not preclude our models from estimating a true exogenous effect accurately (see the within-region R-squared indicated in Figs. S2 and S4).

With regards to the use of Information Criteria for model selection, we now clearly state in the respective captions of Extended Data Figure 1, Figs. S3-S4 and Table S1 that these are the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), respectively. Our Monte-Carlo simulations indicate that Information Criteria may be helpful for an incremental lag selection procedure which avoids overfitting when starting from a large initial number of lags and reducing them. In Fig. S5 we investigate such an incremental lag selection criteria by starting from a model with ten lags for all climate variables and incrementally reducing the number of lags in one variable at a time. Results indicate that for precipitation variables four lags is an appropriate number of lags which optimizes the tradeoff between over- and underfitting (approximate minimum of both BIC and AIC). For temperature variables, eight to ten lags do not result in overfitting (reducing the number of lags increases both BIC and AIC). This additional robustness test is discussed in the main text L.90-95.

Finally, to account for any uncertainties that arise from the precise choice of number of lags, we include empirical models with different numbers of lags in the error sampling procedure of the damage projections. This samples from models with eight, nine or ten lags for temperature terms (models shown in Figs. S8-10 and Table S2-4) to reflect any changes in the parameter estimates that come from these small adjustments to the model. See L.99-103 and L.592-597 of the updated manuscript and for these details. The main results of the projections are largely unchanged when including these different models (a 19% reduction in income over the next 26 years in the updated manuscript rather than a 21% reduction over the next 25 years in the original manuscript).

We thank the reviewer for the prompt to conduct these additional analyses which we believe have further justified the robustness of our empirical models with respect to overfitting and the validity of our projections with respect to small changes in the empirical model.

Extended Data Figure 1 clearly indicates a serious problem of the model built. The existence of large oscillations in the panel regression effects among different lags is especially a concern. Among others, it may indicate strong autocorrelations built into the model. Using the first-difference ($\Delta F_y=F_y-F_{(y-1)}$) may further amplify the effect of autocorrelation. [Note that I don't object the use of ΔF_y, which is a good way to remove technology-related trends (for example).] This is why it is vital to consider stepwise regressions and discriminate the models that have larger possibility of overfitting as discussed above. [Minor comments: in equation (4), the seasonal temperature difference $\hat{T}_r$ is not defined, and the formulation for specific nonlinear terms such as $\bar{T}_r \cdot \Delta Pext_{(r,y-L)}$ is not explained.]

We thank the reviewer for raising this potential issue. We agree that oscillations in parameter magnitudes can sometimes indicate the influence of autocorrelation. With regard to taking first-differences, it is our understanding that this procedure typically reduces the strength of auto-correlation (at least any introduced by a trend), and we are not aware of mechanisms by which it may amplify it. Our use of first-differences is mainly motivated by the aim to avoid baseline assumptions of infinite impact persistence which may then upwards bias projections of future damages if the number of lags included is insufficient to adequately capture the time-scales over which impacts may persist. Please see L. 59-85 of the manuscript for further details on the motivation for this choice.

Although the results of our Monte-Carlo simulations shown in Figs. S2 and S3 indicate that the presence of auto-correlation in our climate variables does not introduce significant bias or imprecision, we take further steps to address the potential issue you raise. We use a restricted distributed lag model (the "Almon-technique", Almon 1995) to increase the degrees of freedom by describing the lag distribution as a quadratic function. Such an approach reflects the apriori intuition you express that oscillations in the lagged coefficients are unrealistic, and hence restricts the lagged parameters to lie on a smooth curve. The methods are outlined and discussed in the additional supplementary Methods Section S3, and the results of this model are shown in Fig. S6. Reference to these results in the main text is given on L.95-99.

The restricted lag model reduces oscillations in the lagged parameters as expected (Fig. S6). However, it continues to provide cumulative marginal effects of a similar magnitude, at least for annual mean temperature, the variable for which oscillations were initially observed (Fig. S7). As discussed in the supplementary methods section pointed to above, this likely reflects the fact that imperfect multicollinearity leads to correlated parameter estimates (i.e. an upward bias in one parameter will correlate with a downward bias in the other, (X. Basagana & J. Barrera-Gomez 2022)). This is why even when multicollinearity is severe, it does not lead to biases in out of sample prediction (J. Neter et al. 1996). In our context, it also means that any effects of multicollinearity due to autocorrelation in the lagged variables leads to correlated estimates of the lagged parameters, which ultimately cause negligible differences in the cumulative marginal effects (if one lag is biased up, the other is biased down). Given that it is these cumulative marginal effects which are of primary importance for projections of the damage caused by future climate change, this suggests that even were it present (which our Monte-Carlo simulations suggest is not the case), bias due to auto-correlation between lagged variables would not cause biases in the cumulative marginal effects or damage projections.

Given that a restricted lag model places considerable constraints on the functional form of the lag distribution, which appear to have difficulty matching the distribution for other variables (in particular for temperature variability in which there is quite considerable heterogeneity in the lag distribution across different values of the moderating variable, see different coloured curves in ED Fig. 1 vs those in Fig. S6), we choose to continue to use the un-restricted lag model for the future projections. This choice is outlined in the supplementary Methods Section S3 as noted above. It may be worth noting here also that with regard to the magnitude of cumulative effects between the restricted vs unrestricted model, the choice to use the unrestricted model leads to smaller cumulative marginal effects of annual mean temperature (Fig. S7) and is hence again in the spirit of a conservative approach to projecting damages which we take throughout the manuscript.

Since regional characteristics are one of the main focuses in this study, it is important to show the geographical distributions of R2 as well as AIC and its increment to the next competitor δAIC. These scores will quantify the model performance (how much the select climate components explain the total economic variance) and the model validity (capturing the largest amount of variance with the least number of climate variables). Without considering such information, the model so built may likely suffer from overfitting problems and contain large uncertainties.

In the context of the panel fixed-effects model, R2 and AIC are obtained for an overall model for all regions. There is hence no geographical distribution of R2 or AIC to be obtained or discussed. As outlined in response to previous reviewer comments, we have highlighted the within-region R2 which describes the proportion of variance explained along the temporal dimension of interest whenever referenced (see captions of Fig. S2 and Tables S1-4). Moreover, we have provided considerable additional robustness tests which demonstrate that the empirical models neither suffer from overfitting (Fig. S5) or large systematic or random uncertainties (Fig. S2-4) and explain the largest portion of variance with the least number of variables (Fig. S5).

The calculation for future projection is not clear. Lines 435-460 describes the procedure. The question is how δ_(r,y) and π_(r,y) in equation (5) are related to g_(r,y) in equation (4) and how ∆F_y is calculated in climate change scenarios. Typically, climate changes are calculated as the differences of the future-scenario minus present-day simulations form a same GCM averaged in a specified duration. This difference method removes, among others, the effect of the GCM's systematic climate biases (which usually are substantial regionally). In this case, the calculation for ∆F_y is important – is it defined still as the first-difference in both the present-day and future-scenario climate simulations (then how to count their change) or directly the change between the two simulations (then how to justify the replacement of the yearly tendency ∆F_y with a much larger step climate change)? Another method could be to calculate g_(r,y) as in equation (4) separately for the present-day and future-scenario climate simulations from the same GCM and then calculate their difference. In this case, the performance of the present-day g_(r,y) distribution over the globe should be quantified by comparing those based on individual GCMs' climate simulations and their ensemble against the baseline calculated from observed climate conditions. In either case, the applicability of the model for future projection should be addressed.

We thank the reviewer for their detailed attention to the manuscript and for identifying a lack of clarity in this part of the methods. We have now updated the methods section to make clearer how the future projection is calculated (please see L560-597 of the updated manuscript). In particular, we have clarified that changes in the primary climate variables of interest as projected by the climate models are calculated as year-on-year changes, to reflect the way in which these variables were used for identification in the empirical models. The moderating variables of the interaction terms are calculated under the future projections using 30-year moving averages to reflect the long-term averages used in the empirical models, but to further account for the changing vulnerability to climate shocks as long-term climate conditions evolve (30-years is a common timeframe to define long-term climatic conditions, but we show robustness tests that the projections are robust to this choice in Figs S11 to S12 at the recommendation of another reviewer). These changes are then used to evaluate equation (4) to calculate a time series of growth impacts, δ_(r,y).

We take year-on-year changes to match the variables which were used in the empirical model, and to avoid the application of a larger step change in climate which you mention. We do not take any difference between present-day and future-scenarios, because when interested in the effects of climate change on the economy we should consider the difference between a future *climate-change* scenario and a future *no-climate-change* scenario. Given the first-differenced form of our climate variables, a future *no-climate-change* scenario constitutes one in which all first-differenced climate variables have a value of zero (other than their random year to year fluctuations) and in which the time-averaged evaluation of equation (4) would therefore simply be zero. We outline this approach in further details on L575-580 of the updated manuscript. Furthermore, in our projections of future damages we use a Monte-Carlo procedure to evaluate uncertainty from the different climate models, different empirical models with different numbers of lags as well as the sampling uncertainty regarding the empirical model parameters (see L.592-597 of the manuscript for further details).

The rest of the manuscript is focused on discussing the results from the regression g_(r,y) model calculations based on CMIP6 SSP2-RCP2.6 and SSP5-RCP8.5 climate projections made by 21 GCMs. These discussions are mostly show-and-tell, except for two interesting points: 1) the projected damages are compared with the currently committed mitigation costs; and 2) the injustice of committed climate damages is highlighted (Figure 3). Given that the regression model so built is questionable (as commented above) for its validity in capturing the "observed" economic-climate relationships and its applicability in projecting the climate change impacts, I cannot see how substantial uncertainties their projected economic damages are involved. As such, I cannot evaluate the robustness, validity, and reliability of the conclusions by this study.

We thank the reviewer for their attentive and constructive comments, by addressing which we have strengthened the robustness of our empirical models and the validity of our projections.

We note that in addressing the concerns of other reviewers, we have also provided additional explicit analysis of the distribution of committed damages across regions by income and historical cumulative emissions which can be found on L246-261 of the updated manuscript.

References.

M. Dell et al 2012. Temperature Shocks and Economic Growth: Evidence from the Last Half Century. *American Economic Journal: Macroeconomics*. https://www.aeaweb.org/articles?id=10.1257/mac.4.3.66

M. Burke et al 2015. Global non-linear effect of temperature on economic production. *Nature.* https://www.nature.com/articles/nature15725

S. Almon 1995. The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*.

X. Basagãna & J. Barrera-Gomez 2022. Reflection on modern methods: visualizing the effects of collinearity in distributed lag models in distributed lag models. *International Journal of Epidemiology.*

J. Neter et al. 1996. Applied linear statistical models.

Response to referee #2

Comments on The economic commitment of climate change

This is an interesting paper and one of a growing literature that explores historic responses of the economy to climate change and then uses the estimated coefficients to project likely future impacts. As such the paper is well written and provides a useful and interesting addition to this literature. The authors suggest that there are two key contributions – one is the statistical innovation of using lags in a more complex econometric specification than is available in the literature and the other is in the findings that the near term cumulative effects are significant and so warrant immediate attention. To many informed readers that is a conclusion, which does not require reemphasizing.

We thank the reviewer for their positive assessment of our work and its contributions and also for their critical but constructive comments. We would like to point out a possible misunderstanding here: While we agree (and clearly state this in the manuscript) that recent work has highlighted the economic benefits of climate change mitigation, demonstrating that the Paris Climate Agreement is also economically optimal (e.g. L. Drouet 2022, M. Burke 2018), our finding provides an additional perspective which was lacking in previous literature: we find that the committed damages, i.e. the near-term damages that cannot be avoided anymore from a mitigation perspective, already outweigh the costs required to stick to the Paris Climate Agreement six-fold. This is in contrast to previous work (based on cost-benefit analyses) that has found that the net-benefits of mitigation only emerge later in the century. We now emphasize this subtle but important difference in the heading of the respective section.

It is hoped that the following comments are of utility to the authors in revising this paper.

A useful contribution of this paper is to include lags of the GDP-pc growth term to determine whether there is a reversal from the initial effects of climate change variables. Unsurprisingly the finding is that lags suggest a reversion to previous growth rates and hence results are not as exaggerated as in some specifications that do not include lags. However the authors may wish to revisit their explanation of the rationale that they provide. Especially in lines 59 to 68, it is unclear how a specification that begins with first differences without lags, and then tests for lags of first differences implies a baseline "…null hypothesis of purely level effects….[with] … instantaneous effect on …growth…"(line 62). Then by line 68 there is a correction with the recognition that the baseline without lags "… has an infinitely persistent effect on .. growth…". It cannot be both and the former assertion seems to be plainly wrong.

We thank the reviewer for their detailed reading of the manuscript. In this case we believe the comment results from a mis-reading of our original text. We do indeed first state that when using the first differenced climate variables without lags, the null hypothesis is one of level effects with only an instantaneous effect on the growth rate (since this implies a dependence of the growth rate on a change in the climate, such that a step change increase in the climate would lead only to a on- off effect on the growth rate). We then go on to state that in contrast to this case, when "climate variables are used ***without taking the first difference***… the baseline specification without any lags constitutes a null hypothesis of pure growth effects, in which a change in climate has an infinitely persistent effect on the growth rate." The confusion appears to have arisen by having missed the fact that the second statement refers to a case where climate variables are included without taking their first difference. We have added additional text to clarify these two cases in the updated manuscript on L59-85 of the main text, as well as L.528-539 in the methods section.

Secondly, it is somewhat difficult to comprehend the full rationale for the particular econometric specification that is used. Since economic theory does not dictate the form of the specification, it is necessary for empirical work to show that the preferred estimating equation is robust and makes sense. Hence for instance, it is unclear why the regressions are of the form $\alpha_t \Delta x_{(i,t-L)}+\beta x_t \Delta x_{(i,t-L)}$ instead of say $\alpha_t \Delta x_{(i,t-L)}+\beta \Delta x_t \Delta x_{(i,t-L)}$ which also seems plausible. More generally it is notable that there are no regression tables provided or robustness tests shown to determine the sensitivity of the results to plausible changes in specification. IN sum its impossible to tell how robust these results are to small or large changes in specification. Nor is the current specification and the interaction terms well justified.

We thank the reviewer for their detailed attention to the econometric specification. In the updated manuscript we have provided extensive additional robustness tests which assess the validity of the particular empirical specification. Most of these address the robustness of the empirical models to auto and cross-correlations in the climate variables (as raised by multiple reviewers) as well as the choice of the number of lags. Other components of the empirical specification such as the specific choice of climate variables and their interaction terms were explored comprehensively in previous publications. In the main manuscript we now provide more detailed direction to these previous publications (e.g. L.53-55) which provide extensive motivation for the use of these particular climate variables and interactions as well as extensive robustness tests of their validity.

Given that in the present context, the main empirical challenge is to assess the persistence of climate impacts on growth, we have introduced a number of additional robustness tests which address this point. Please see L86-106 of the updated manuscript, the new supplementary methods Sections S1 and S2, and the new figures Figs. S1-7 and Tables S1-4 for full details of the results. To summarise, we provide:

- Monte-Carlo simulations which demonstrate the robustness of our results to auto-correlation in the climate variables which may cause multicollinearity in the lagged variables (Figs. S1-S3).
- Monte-Carlo simulations which demonstrate the robustness of our results to correlations between climate variables which may also cause multicollinearity (Figs. S1 and S4).
- Incremental lag selection procedure using information criteria to avoid overfitting (Fig. S5).
- A restricted distributed lag model to investigate and limit potential parameter oscillation caused by autocorrelation (Figs. S6 and 7).

We find that our original empirical results are robust to these different tests.

Moreover, in the error sampling procedure of our damage projections we now include a sampling from empirical models with slightly different numbers of lags (see L.99-103 of the updated manuscript and L.592-597 of the methods for these details). With this method we aim to explicitly account for any outstanding uncertainty arising from the precise choice of the number of lags in the empirical model. The main results of the projections are largely unchanged when accounting for these small changes in the model specification (a 19% reduction in income over the next 26 years in the updated manuscript rather than a 21% reduction over the next 25 years in the original manuscript).

Finally, we also provide robustness tests regarding the timescale with which changes in the moderating variable of the empirical models are estimated under future projections (10 years and 20 years compared to 30 years in the main specification; Figs. S11-S12) as well as robustness tests with respect to the choice of method used for accounting for sub-national price changes (Fig. S13).

In addition to the regression table corresponding to the empirical model including 10 lags for each climate variable with which we begin our empirical analysis (Table S1), we now also provide regression tables for our preferred specifications including eight to ten lags for the temperature variables and four lags for the precipitation variables (Tables S2-S4).

The authors may also wish to revisit the interpretation of their findings and present further tests to support some of the conclusions. If, as is often alleged, the climate variables for heat and precipitation are correlated in this data set then multicollinearity would suggest that the estimated coefficients no longer provide an accurate estimate of the marginal impacts of each of the weather variables used. While collective impacts would be accurate – one cannot unambiguously identify the marginal effects of any one of the correlated explanatory variables. This issue could be addressed by providing a correlation matrix to facilitate interpretation.

We thank the reviewer for this helpful comment and suggestion. As outlined above, we have now included extensive additional robustness tests which include correlation matrices to assess the extent of cross (as well as auto-) correlation between the climate variables (Fig. S1 of the updated manuscript), as well as Monte-Carlo simulations which demonstrate that the empirical results are robust to these correlations (Figs. S2-S4). To summarise:
We do find considerable cross-correlations between precipitation variables in particular (Fig. S1f). The Monte-Carlo simulations outlined in supplementary methods Section S2 assess whether the empirical models are robust to such cross-correlations. We randomly reassign real climate data to different economic regions (preserving the correlation

structure of the climate data), and add effects of known size from each climate variable into the data to generate ensembles of artificial data with known effects. We then test whether our empirical models are able to recover these effects in the presence of the cross-correlations between the climate variables. We find that we are able to do so accurately only when including all climate variables simultaneously in the regressions (Fig. S4f-j). When including only individual climate variables, the empirical models consistently underestimate the true effect sizes, particularly for the precipitation variables which exhibit strong cross-correlation (Fig. S4a-e). These results indicate that accounting for all climate variables simultaneously in the regressions is actually the necessary procedure in order to recover their separate effects in the presence of cross-correlations. Moreover, these results provide further support for our finding that the impacts of annual mean temperature intensify when accounting for other correlated variables (see. L.193-198 of the updated manuscript and Fig. S16).

Finally it may be helpful for this paper not to accompany the often hyperbolic narratives in some of the literature. The results (if robust) suggest a 21% reduction in gdp over 25 – 30 years. This amounts to a rather modest 0.7% reduction in gdp each year, which would be well within the margin of statistical error for gdp revisions conducted by statistical authorities. In other words though not a trivial impact, the effects are not as large as the tone of some of the text suggests.

We believe there may be a misunderstanding here with respect to the nature and magnitude of our results. The 19% reduction in GDP projected to occur by mid-century is a permanent reduction in annual GDP. While the changes in the annual growth rates which lead to these reductions may appear modest (0.5-0.7%), and may perhaps be within the margin of statistical error of GDP measurement, their cumulative effect over time is permanent and certainly larger than this margin of statistical error.

Nevertheless, we appreciate the reviewer's comment with respect to framing. In the manuscript we refrain from making any comments which constitute value judgements on the magnitude of the damages. We predominantly make relative statements regarding the magnitude of damages in comparison to the magnitude of mitigation costs from IPCC Integrated Assessment Models (abstract, L.164-173), the magnitudes due to different variables (L.179-210) the magnitudes between regions (abstract, 223-263), and the magnitudes with respect to other studies (278-297). As such we believe that the tone of the majority of the text conveys a judgment-free, factual and comparative narrative.

The only line which (in our opinion) makes a value judgment on the magnitude of damages is L. 124-127 of the updated manuscript which reads:

Even though levels of income per capita generally still increase relative to those today, this constitutes a substantial income reduction for the majority of regions, including North America and Europe, with South Asia and Africa being the most strongly affected (Fig. 1).

This line is intended to avoid an over-statement of the results by emphasizing that levels of income generally still increase relative to today's values (something which could be missed when seeing a %-reduction), and in our opinion this statement agrees well with the reviewer's comments that the impacts are still not trivial. We hope this addresses the concerns of the reviewer and if there are still any outstanding we would happily be pointed to more specific instances. Moreover, we note that in numerous instances we have added damages estimates in monetary rather than % reduction form at the request of another reviewer.

Response to referee #3

Review of manuscript #2023-01-01329 entitled ‘The economic commitment of climate change’.

A. Summary of the key results:
Kotz et al. use several metrics of historical and simulated climate change along with socioeconomic data to make projections of the economic damages associated with future climate change. In particular, they assess the economic damages at the time when projections from a low-emissions scenario diverge from that of a high-emissions scenario. The authors find that by this mid-century point of divergence, the world economy will experience a 21±8% reduction in population-weighted income, vastly outweighing projected mitigation costs. Lastly, they break down the results by a variety of climate metrics at sub-national levels to better understand the drivers of change.

We thank the reviewer for their highly positive assessment of our work and for their constructive feedback on certain methodological issues.

B. Originality and significance:
This study differs from past work through a variety of methodological changes, which lead to much larger estimates of economic damages. The work seems novel, but I am not overly familiar with past research on the economics of climate change. If proven to be robust, the conclusions of this work have broad significance that would warrant publication in a high-profile journal such as this one.

C. Data & methodology:
The data seems appropriate for this type of study. I would like to see the climate data described in slightly more detail.

D. Appropriate use of statistics and treatment of uncertainties:
There are a couple of instances (discussed below) where the use of statistics can be improved.

E. Conclusions:
The conclusions seem robust, but it is difficult to fully gauge how sensitive the results are to a variety of methodological choices at this time.

In the updated manuscript we provide extensive robustness tests of our results. These include robustness tests of the empirical models using Monte-Carlo simulations, Information Criteria, and restricted distributed lag models (see supplementary Methods Section S1-S3 and Figs. S1-S7), as well as robustness tests of the choice of time-frame for estimating moving averages of the moderating climate variables under future projections (Figs. S11 & S12). See further details below.

F. Suggested improvements:
There are several minor changes that can be made to improve the manuscript as discussed below.

G. References:
The references are mostly appropriate.

H. Clarity and context:
The paper is very well written, but I found several instances where aspects of the analysis could be better described.

Major comments

Sensitivity of results to methods: It is somewhat difficult to fully gauge the robustness of the results when there are several seemingly arbitrary methodological choices being made. It would be helpful to show how sensitive the results are to some of these choices. For example:

L438-440: How sensitive are the results (e.g. time of divergence between the emissions scenarios) to the use of 30-year moving averages. This seems like a conservative choice, which would delay the point when the two scenarios are distinguishable from one another.

We thank the reviewer for their concerns regarding the sensitivity of our results to certain methodological choices. In addressing the concerns of all reviewers we have updated the manuscript to include extensive additional tests of the robustness of our results. In particular, these focus on the robustness of our empirical models to:

- Autocorrelation in the climate variables, (see supplementary Methods Section S1 and Figs. S1-3, S6 and S7)
- Cross-correlation in the climate variables (see supplementary Methods Section S2 Figs. S1 and S4)
- The selection of the number of lags in the empirical models (see Fig. S5)

Please see L86-100 of the updated manuscript, and the new supplementary discussion and methods sections for further details of these additional tests.

Additionally, we also provide robustness tests to address the specific point raised by the reviewer regarding the use of 30-year moving averages to evaluate the moderating climate variables, referred to on L173-177 of the updated manuscript. Figs. S11 and S12 show results using 10 and 20-year moving averages instead and show that the results are very robust to this choice. The magnitude of damages and global time of divergence of damages are very insensitive to this choice because most of the damages are driven by changes in the primary climate variables which are used as independent variables in the empirical models, for which year-on-year changes are calculated without any moving averages. Moving averages are only used to assess the moderating variables of the interaction terms in the empirical models (see L560-568 of the manuscript in the methods section), which reflect slowly evolving vulnerability of regional economic growth to changes in the climate (i.e. the moderating variables control the marginal effects of a given change in climate, shown by the different coloured curves in Extended Data Figure 1 and Figs. S8-10). We have made some changes to the methods section to make this distinction clearer, please see L. 560-568 of the updated manuscript for these details.

Fig S4 (L158-160): I am also having a difficult time grasping why the changes need to be framed in the context of their historical interannual(?) variability. Please clarify. Additionally, 2005-2035 seems like an odd choice of baseline climate to calculate change metrics from. I recommend comparing with the historical period from which the variability metrics are calculated across.

We provide a framing of the future climatic changes in terms of historical interannual variability in the context of assessing the relative importance of each climate variable for future damages. This is because the empirical models which quantify the vulnerability of economic growth to climatic changes (ED Fig. 1 and Figs. S8-S10) are estimated on historical data i.e. on historical interannual variability. As such, when applying these empirical models to future climate change simulations, climate variables which exhibit a large change in comparison to the interannual variability are likely to cause more damage in a purely statistical sense. E.g. imagine that in the historical period annual precipitation changes by 500mm from one year to the next whereas annual mean temperature by 0.5C. Our empirical models fit changes of these magnitudes to changes in economic growth. If then looking at future climate change scenarios where regional temperatures increase by 1C and precipitation by only 100mm, it is likely that the change in temperature will cause a larger impact than those of precipitation. We provide additional discussion of this phenomena including a delineation between its statistical and potential mechanistic interpretation on L. 179-188 of the updated manuscript. Moreover, we clarify that historical variability refers to interannual in both this text and in the caption of Fig. S14. Finally, we also follow your suggestion and express changes with respect to the historical period in which the empirical model was fit, namely 1979-2019 (newly added Fig. S15). Moreover, we also add further discussion regarding the fact that future climate projections typically underestimate the change in temperature variability and precipitation extremes compared to those observed historically, suggesting that this could also be a factor in these variables causing less damage in our projections (see L.207-210 of the updated manuscript).

Fig 3 Analysis: This is the weakest component of the manuscript in my opinion. It is only briefly touched on, which begs the question whether it warrants a place in the main text. I recommend adding a bit more detail here. Moreover, there should be at least some basic statistics shown to support the statements made in this paragraph (e.g. correlation between metrics).

We thank the reviewer for this helpful suggestion, and consequently we provide additional analysis with regards to the injustice of committed damages on L244-263 of the updated manuscript. In this analysis we provide Spearman's rank correlation statistics of the relation between committed damages and both historical cumulative emissions and present day income. These indicate the presence of significant injustice (p-values<0.01 for correlation across both dimensions). Moreover, we quantify this relationship by assessing the difference in committed damages between the upper and lower quartile of regions when ranked by present day income or historical emissions (see text referred to above and the new Fig. S17).

Wording: the use of "committed" throughout the manuscript could be misconstrued when interpreting the meaning of these results. This seems to imply the economic costs that are already "baked in" because of past emissions, when in reality it is the minimum cost associated with continued emissions (at a lower rate: RCP2.6). I think this needs to be clearly stated early in the text to limit confusion.

We thank the reviewer for their concern regarding this wording. In the original manuscript we used the terms committed due to "historical emissions and socioeconomic inertia" to refer to both of the effects which the reviewer mentions. We now elaborate further on what we mean by the committed damages on L. 34-35 and L.117-120 of the updated manuscript. These discussions, as well as that in the caption of Fig. 2, make clear our definition in terms of the statistical indistinguishability between the two most extreme emission scenarios.

Minor comments:
L54: Cite and briefly discuss the findings of past assessments (e.g, Extended Data Table 1).

Done in the updated manuscript with further discussion L41-55.

L57-59: One more sentence here briefly describing persistence would help non-specialists.

Done in the updated manuscript on Ls.59-79, including more detailed descriptions of growth and level effects.

L93: Include a supplemental table showing the models used.

Done. See Table S5.

L94: need a citation for CMIP6 (e.g. Eyring et al. 2016).

Done.

L104: change to "without climate change impacts"

Done.

L105: You should still define what the IPCC likelihood means.

We now refer to the caption of Fig. 1 where this definition is given.

L108: Missing period after (Fig 1)

Thanks!

L109: define pure growth or pure level effects in the methods.

See L.59-79 and also L.505-511 of the updated methods.

Fig 1: Make the 0 y-axis line solid.

Done.

L134: remove ‘already’

We think the framing that near-term damages before 2050 already outweigh mitigation costs is an important part of the narrative of these results, complementing the fact that cost-benefit analyses typically only find benefits of mitigation to emerge in the second part of the century as outlined in the discussion of results on L.167-169. We see that “already” is perhaps not the best way to make this point since it constitutes some form of “value judgment” of the results, and we therefore have updated the title of this section accordingly to instead reference that we would like to explicitly compare damages to mitigation costs before 2050.

L140: state an approximate dollar cost for mitigation.

The main reason we presented impacts and costs in terms of % change to a baseline is that this makes results independent of the choice of baseline socio economic projection. In the updated manuscript we now additionally provide dollar costs under the middle-of-the-road socio economic projection SSP2, see Ls. 127-129 and 165.

L142: Why does this IAM seem like such an outlier?

Prompted by the reviewer’s question, we revisited the IPCC AR6 database from which we found and selected IAMs providing appropriate data for estimating mitigation costs. Doing so, we realised that two of the five original models we used had actually failed the IPCC vetting procedure used for selecting IAMs, despite their results being included in the database. We missed this detail because information regarding the vetting procedure is only provided within a meta-data file. The two models which failed the vetting procedure provided estimates of mitigation costs which differed significantly from the other three, one producing estimates much lower and the other much higher (the one which was mentioned in the text of the previous manuscript). The vetting procedure is designed to select models which have accurately reproduced historical trends in emissions and climate (see Annex to AR6 WGIII), and this is likely the reason that these models produced cost estimates which were much larger or smaller (e.g. perhaps having too high a climate sensitivity, emission intensity of GDP, although we cannot clarify the precise reason for vetting given the available information). In the updated version of the manuscript, we have only included the three IAMs which both provide the appropriate data needed for our analysis and which passed the IPCC vetting procedure. We provide some additional text describing this additional selection criteria in the updated methods section on L. 582-584. We thank the reviewer for making this detailed inquiry!

L155: Can these changes be roughly quantified? For example, state the average contribution from each of the metrics assessed.

In the updated manuscript we now provide additional details quantifying the contribution to overall damages from each of the climate variables. Please see L. 199-210 of the updated manuscript and the additional content of Fig. S16a. We find that the biggest additional contribution to overall damages arises from the inclusion of daily temperature variability in addition to annual average temperature.

L155: Add a companion supplemental figure that shows the regular climate change metrics (unlike Fig S4). This would really help the reader to understand for example, how much warming or precipitation change is associated with this level of economic damage (e.g. the change metrics by 2048).

We have done so, this is now Figure S15 of the supplementary which is referenced in the caption of Figure S14 and on L. 229-231.

L161: Add some measure of uncertainty to this estimate.

Done.

L163: State which variable's inclusion has the biggest impact as this seems important to know for future studies.

As our answer to the above point, in the updated manuscript we now provide additional details quantifying the contribution to overall damages from each of the climate variables. Please see L. 199-210 of the updated manuscript and the additional content of Fig. S16a. We find that the biggest additional contribution to overall damages arises from the inclusion of daily temperature variability in addition to annual average temperature.

L165: Briefly explain why this is.

Done in the updated manuscript. Please see L. 195-198 which refers to the Monte-Carlo simulations presented in Fig. S4 which demonstrate that insufficiently accounting for other correlated climate variables can limit the ability to recover the true magnitude of the effects.

L174: I would think that the change in frequency of events exceeding an extreme precipitation threshold would be more relevant than the magnitude change above said threshold. Why is magnitude change a preferable metric here?

This measure is designed to capture both changes in the frequency and intensity of events by counting the amount of rainfall on days exceeding the threshold. (e.g. this means that a day that exceeds the threshold by 20mm is weighted more strongly than one that exceeds it by 1mm). This question is explored in the preceding publications which investigated the impacts of different precipitation measures on economic output (see https://www.nature.com/articles/s41586-021-04283-8$) . We make more explicit reference to this publication in terms of the choice of specific measures in L. 53-55 of the updated manuscript.

L177- Frequently citing the wrong figure in this paragraph.

Thank you. This has been updated.

L180: change to "Damages due to increasing mean temperature…"

Done. See L225 of the updated manuscript.

L182: This is despite increases in temperature being much larger across high latitudes, which should be briefly noted.

An interesting point! We have now added this on L. 229-231 of the updated manuscript.

L185: Why is there such a contrast between Europe and the US (at a lower latitude)?

We believe that technical discussion of the mechanisms behind specific regional climate change is beyond the scope of the manuscript, but we refer the reviewer to the following studies which may provide some additional information in this regard: https://www.pnas.org/doi/full/10.1073/pnas.2103294118, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2012GL052730. These differences between the US and Europe appear robust in both CMIP6, historical data, and regional climate projections. Regional mechanisms such as soil moisture may play an important role, but a specific analysis of why these are different for the US and Europe seems lacking in the literature.

L184-186: rephrase this sentence. Unclear if referring to the temperature variability change or damages.

We have rephrased this sentence aiming to distinguish more clearly between physical changes and economic damages. See L. 231-235 of the updated manuscript.

L190: this is very difficult to see given the size of panels d-f.

In the updated manuscript we have reformatted Figure 2 to give more space to the precipitation variables.

L193: could add another reference here.

Done.

L210: this should say "national cumulative emissions per capita"

Done.

L219: Put their estimate in brackets here?

Given that a straightforward comparison is dependent on the date of comparison (e.g. end of century is usually what is described in other studies rather than mid-century as mainly discussed here) we believe it is easier to continue to refer to Extended Data Table 1 which provides comparisons across multiple different studies at a comparable time-frame and to which we now make a more explicit reference on L.280 and 290.

L334-335: Precipitation from reanalysis can be poor, state what bias correction is done to this product over land.

Done on L. 410-413 of the updated manuscript. Moreover on L.419-420 we have made explicit reference to preceding studies in which robustness tests with different reanalysis data were conducted.

L393-397: It doesn't seem like this would account for very large administrative boundaries where the climate changes where people live may not reflect the area average (e.g. much of Canada).

Given that agricultural (and sometimes also manufacturing) activity is not always correlated with population density, we do not know a priori that a population weighting produces a more accurate reflection of economic exposure to climatic changes. While this may indeed be the case in the high latitudes of Canada, we cannot guarantee this to be the case in all regions. Given that in the two crucial studies preceding this publication on which these results are built (refs. 15 and 16 of the main manuscript), there is negligible difference found regarding the effect of a population or area weighting, we continue to use area weighting as the main specification because we consider this a choice with less a priori assumptions. Given the results of these previous publications, we strongly suspect that it will not influence the results in any meaningful way.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

This is a revised manuscript from an earlier submission, which I reviewed. My review this time will only focus on new issues with the the new submission.

Referee #2 (Remarks to the Author):

The attempts to address the referee comments and the detailed replies are greatly appreciated. However upon re-reading the paper it is apparent that the very same problems persist.

One issue is terminological and could have been fixed, but has not. In economics (please note that this is not a paper in physics), we talk of growth effects to imply that a change (or shock) leads to a permanent change in growth. For instance, a new road (or rise in temperature from a to b) increases growth from 4% to 5%. Formally, this is captured by a first difference on the LHS of a regression, such as:
$\Delta Y_t = \alpha + \beta \Delta X_t + \varepsilon_t$ (1)
where ΔY_t is the change in the dependent variable Y between time t and t-1, ΔX_t is the change in the independent variable X between time t and t-1, which is similar to the specification in this paper

Or the specification may be of the form as in many other papers in econometrics:
$\Delta Y_t = \alpha + \beta X_t + \varepsilon_t$ (2)
where ΔY_t is the change in the dependent variable Y between time t and t-1, X_t is the level of the independent variable X between time t and t-1, and ε_t is the error term.
This also generates a growth effect since a one time extension of the road (or new temperature level) improves growth forever. Statistically whether or not to difference the LHS and RHS will depend on whether unit root issues emerge (i.e. if these are cointegrating vectors and the order of cointegration). The choice ought to be determined by the statistical properties of the data aimed at minimizing spurious regression problems. This is why there is a literature on cointegration in panel and time series data.

Adding lags allows one to examine if the growth rate reverses or accelerates – but it is still a growth effect (unless annulled).
By contrast a level effect is when there is a one-off increase in the dependent variable that does not persist. Thus the above might become $Y_t = \alpha + \beta X_t + \varepsilon_t$ or $Y_t = \alpha + \beta \Delta X_t + \varepsilon_t$
The authors seem to think that if there are no lags in equation (1) it becomes a level equation – this is not the case in economics. Perhaps the authors are assuming that a constant growth rate over time (β) is a level effect while varying rates are called "growth effects". This is not how the terms are conventionally used – including in the economics references in the paper (eg Newell et al 2021). But the rationale is unclear for claims such as Line 68, 69 and supp material.

Next I had requested robustness tests of the specification used since there is no economic theory used to guide the specification. This of course is standard practice. Instead the authors reply with reference to autocorrelation tests, (rightly) brought up by another referee. And cite their own papers to justify this specification. Neither of these responses answers the question raised. Autocorrelation is about memory in the data over time that impacts regression coefficients, muticollinearity which is addressed is about correlation between variables. This is quite different from knowing if a plausible change of specification (with or without autocorrelation and

multicollinearity) will alter the results. With over 800 specifications examined in the literature it would be good to know if this one is robust. There is good reason to be cautious as metanalyses have shown that growth (as opposed to level of GDP) models are fragile. To quote the standard reference here: "Across just those growth models that specify a non-linear temperature function, the combined model and sampling uncertainty yield a standard deviation of predicted impacts equal to 132% of GDP, ……. specifying impacts on GDP levels, not growth, yield far less uncertainty in climate impacts; the standard deviation is equal to less than 3% of GDP …….” Newell et al 2021)
Hence a single specification will not make the case for those familiar with the literature.

Finally, whichever way the authors wish to cut things a 19% drop in GDP over 25 to 30 years remains within margins of routine statistical corrections – this is unarguably true and suggests that inflated interpretations will only serve to undermine the credibility of the paper among those that are familiar with macroeconomic statistics.

Referee #3 (Remarks to the Author):

The authors have done an excellent job of addressing the previous set of reviewer comments. In particular, the added robustness testing and methodological detail strengthens the paper and will help to improve reproducibility. I have no further suggested improvements.

Author Rebuttals to First Revision:

Response to Referee #1

This is a revised manuscript from an earlier submission, which I reviewed. My review this time will only focus on new issues with the new submission.

We thank the reviewer for the comments and suggestions. We would like to point out that we had addressed some of the concerns that are reiterated here with regards to our empirical model already in the previous round of revisions. In particular we had included several additional analyses to address overfitting during model selection based on the use of Information Criteria (Fig. S1) as well as robustness tests regarding oscillations in lagged parameter estimates (SI Section S4, Figs. S9 & S10)).

We appreciate that given the extensive revisions it is quite possible to miss certain details, and have therefore provided more explicit discussion of the robustness of the empirical models to overfitting in the new Supplementary Methods Section S1.

In the revised manuscript, please find edits from the previous and current round of revisions in red and blue respectively. Line, figure and table numbers refer to the most recently revised version of the manuscript.

1. While the Monte-Carlo simulation analysis helps address the robustness of the region models to auto- and cross-correlations in climate variables, it does not answer the overfitting problem. The left figure illustrates examples of interannual correlations between observed crop yield and temperature variations at 0-10 lags in 40 years (the same length as used in this study), comparing the use of the first difference and the full level data. As expected, the correlations are totally different between using the full field and the first difference. This is not a major concern as it is reasonable to build a regression model with different predictors that are identified as most appropriate (Δ is a better choice in this study). The key issue here is that the correlations using the first difference oscillate substantially between even and odd lags and such oscillation significantly amplifies those using the full field. Such oscillation is expected since the same field f in any year k appears twice with a reverse sign between consecutive even and odd lags (e.g.,$\Delta f = f - f$ atlag=0and $\Delta f = f - f$ atlag=1).The amplification is especially serious when correlations are small between original fields or at zero lag; they can become substantially high and statistically significant values of both negative and positive signs. There is no physical reason to justify such oscillated and delayed climate impacts. The problem is then how such amplified oscillations among lags, which result purely from numerical data manipulation, affect the regression model outcome.

We appreciate the concerns of Referee #1 regarding oscillations in lagged parameter estimates, and thank them for providing an example of extreme oscillations in different data of a similar length. In response to having raised this issue in the previous round of reviews, we agreed that lagged parameter estimates may exhibit unrealistic oscillations due to (imperfect) multicollinearity between lagged parameters (L102-105 of main manuscript, SI Section S4 in the current manuscript). We addressed this concern explicitly and extensively in the previous round of reviews (in what is now SI Section S4). In this section we note that our Monte-Carlo simulations indicate that multicollinearity between lagged parameters appear not to introduce bias (as shown in SI Section S2), and demonstrate how using a constrained distributed lag model designed to limit oscillations leads to similar cumulative marginal effects (Figs. S9 & S10) and overall results. With regards to overfitting, please see our responses below.

I fully agree with the authors that including more variables at a greater number of lags would certainly (by design!) increase the regression model "skill" – explaining a larger fraction of the within-region R-squared variance. However, that skill is simply a measure of data training by the regression procedure and does not indicate any model ability of generalization in explaining or predicting the unseen data. The outcome cannot be used as the base to justify whether the built regression model is overfitting or underfitting. Without observational evidence or physical justification, selecting variables or lags based on such data training skills alone is purely subjective. For such data-driven problems, cross-validation is necessary and often a normal procedure for model parameter adjustment or feature selection. Holding out a segment of data for "validation", one may train the model on the remaining data with varying variables and lags and check the differences in the model representation of the holdout data to make an optimal selection of the variables and lags for the final model. This final model with so selected variables and lags may then be trained on all the (training + validation) data for the subsequent use (generalization) to predict the unseen test data. Ideally, we would like to split the entire data period into three independent sub-periods for training, validation, and test. Given that the 40-yr data period is relatively short, one may use the *K-fold* cross-validation procedure for building the regression model.

We appreciate the concerns of Referee #1 that the empirical model may be overfitting and that within-region R squared is not a useful metric to identify an appropriate model which can generalize well to other data. In general, we think it is important to note that our approach in this context is not one of prediction of regional GDP growth, but rather of inference of the impact of climatic conditions on GDP growth. Nevertheless, in our response to similar comments in the previous round of reviews, we provided a detailed assessment of overfitting using Information Criteria (as suggested by Referee #1 in their original review) to select the optimal model when incrementally removing lags (now in Fig. S1 and discussed on L93-96). Moreover, in the updated manuscript we have now also provided an additional analysis which similarly uses Information Criteria to demonstrate that including additional climate variables does not lead to overfitting (see Table S2, also provided below). These results demonstrate that our original main specification is optimal in terms of providing the best description of the data while avoiding overfitting (Fig. S1, L93-96 of main manuscript). For clarity, we have now included additional extensive discussion of these tests in the new SI Section S1 entitled "Avoiding overfitting".

Climate measure removed	None (full model)	Daily temp. variability	Total annual rainfall	Annual number of wet days	Extreme daily precip.
AIC	-34220	-33690	-34140	-34080	-34188
BIC	-5490	-5111	-5489	-5435	-5537

Table S2. Information criteria to assess model overfitting when removing additional climate variables. Akaike and Bayesian Information criteria to assess the relative strength of models which include either all climate variables or remove individual variables. The models here use eight lags for temperature and four for precipitation terms as indicated in Figure S5 to be optimal for avoiding overfitting in terms of lag selection. Lower information criteria indicate a better model in terms of explaining a greater amount of variance while avoiding overfitting by penalising additional terms. Both criteria indicate that including all climate variables provides the best model in terms of avoiding overfitting, except the more conservative BIC[4,5] measure when considering extreme daily precipitation.

We appreciate the suggestion of Referee # that cross-validation may provide a useful alternative method to address concerns regarding overfitting. Given that we have taken a number of steps to demonstrate that our empirical model does not suffer from overfitting (using Information Criteria as originally suggested by Referee #1 in their first review, as outlined above), we do not believe that such additional analysis is necessary to demonstrate that the model does not suffer from overfitting. A number of studies documents that cross-validation provides asymptotically equivalent guarantees on selecting the true model as compared to the use of Information Criteria such as that which we use in Fig. S1 and Table S2 (Shao 1993, Stone 1976). Moreover, cross-validation for model selection is arguably more appropriate for *prediction* problems rather than *causal inference* with which we are concerned in this context. We now provide detailed discussion of this issue in the new SI Section 1 entitled "Avoiding overfitting". In particular, we discuss how in the context of causal inference, model specification choices should be guided by logical reasoning in order to identify plausibly exogenous variation which maximize the confidence in a causal interpretation of results. For example, fixed-effects and time-trends are selected for this purpose (as documented in our methods section), whereas a model specification choice guided by cross-validation might choose an alternative set of fixed-effects which better fit the data but do not guarantee a causal interpretation. More specifically in this particular context our empirical model is also designed to provide inference of a lower-bound of the extent of climate impact persistence on growth rates, by using a first-differenced framework and by detecting statistically significant lags (see additional discussion in the methods section "Empirical framework - fixed-effects distributed lag model"). Again, the specification choice here is guided by logical reasoning for inference of a lower-bound of impact persistence, rather than simply a specification which optimally fits the data as might be detected by cross-validation. Further interesting discussion on the issue of cross-validation and inference vs prediction in the context of climate econometrics can be found in this blog post by Marshall Burke (http://www.g-feed.com/).

For these reasons, we do not think that it is helpful to apply cross-validation in this context. We have provided extensive additional tests using Information Criteria to demonstrate that our empirical models do not overfit, with regard to either the number of lags (Fig. S1) or the inclusion of different climate variables (Table S2), and provide additional discussion on these results and the use of cross-validation in the new SI Section S1 entitled "Avoiding overfitting".

2. Equation (4) contains interaction terms which include annual mean temperature T, seasonal temperature difference T, total annual precipitation Pr, and annual wet days, Pwd. All these climate variables are the "level" quantities that are subject to substantial systematic errors and uncertainties in future projections. [All variables must be given with specific units, which are currently missing]. As I mentioned earlier, the typical practice is to remove systematic errors by calculating the projected climate change as the difference of the future-scenario minus present-day simulations from each GCM. These errors are substantial at a regional scale. Hence, it is important to quantify how much g%,' uncertainty is due to these errors and what contributions are from the interaction terms relative to the other first-difference factors.

We appreciate the concerns of Referee #1 regarding systematic errors in Global Climate Model (GCM) projections of the moderating variables of the interaction terms which are expressed in levels. We agree that raw GCM projections exhibit regional biases in different climate variables. However, as outlined in our original manuscript, we use an ensemble of CMIP-6 climate models which have been bias-adjusted such that their distribution of regional daily precipitation and temperature accurately reflect those of historical observations (main Methods Section "Future climate data", L446-453). In fact, we take particular care to re-run regressions from previous studies using the exact data which were used to do this bias-correction (main Methods Section "Historical climate data", L433-444), in order to avoid introducing systematic biases such as those which Referee #1 cites.

In addition, in the revised manuscript we have now explicitly assessed the potential bias and uncertainties here, by evaluating the extent to which the bias-adjusted climate model output accurately reflects the observational data at the regional level, as well as the spread of regional projections across models. We find that the bias-adjusted climate model data reproduce the observed climatological patterns exceptionally well (Pearson correlations>0.998 and average absolute percentage error<3%), with limited spread (< 3%). This analysis is presented in the new Table S7 of the updated SI (which we also provide below) and is discussed on L119-122, L452 and L62-626 of the updated manuscript.

We thank the Referee for noting that some of the climate variables are lacking specific units. We have updated this in the revised manuscript which explicitly gives the units of the different variables on L433, 446 and 522-525.

Climate measure	Annual mean temperature	Seasonal temperature difference	Total annual rainfall	Annual number of wet days
Pearson correlation to observations	1.000	1.000	1.000	0.998
Average absolute percentage error to observations	0.2%	2.1%	1.2%	2.8%
Coefficient of variation across climate models	0.0038	0.018	0.018	0.030

Table S7. Evaluation of systematic bias and uncertainty in bias-adjusted climate model output over the historical period 1979-2015. The first row shows Pearson correlations between regional climate data from the mean of the bias-adjusted CMIP-6[15,16] ensemble and the W5E5 observational dataset[17] for the different climate variables used as moderating variables of the interaction terms of the empirical models and in the projections of future damages. The second row shows the absolute percentage difference between the climate data from the two sources, averaged across regions. The third row shows the coefficient of variation (standard deviation divided by the mean) of each climate measure across climate models, averaged across regions.

3. For prediction, the model overfitting is equally, if not more, important as the model underfitting. It is imperative that one should not include physically irrelevant (which may not be able to identify for this study) or statistically insignificant features (variables or lags) into the empirical model. Although these features may play a minor role in the model trained to explain the historical data, they may significantly alter the model's projection of the future when these features may have large variations.

In addition, the importance of the accumulative climate effect must be evaluated relative to the overall variance of the economic growth. While I fully agree with the authors that it is not surprising to see small (<5%) within-region R-squared variance values attributable to climate factors, it is difficult for me to understand how the future economic output due to the minor effect of climate change alone can be dropped by over 20-50% (which is related to total growth). I would strongly hope that the authors will more rigorously train the model with the select climate variables and lags at individual sub-national regions (via cross-validation as discussed in comment [1] and accounting the systematic error issue as discussed in comment [2]) and show the geographic map of the within-region R- squared variances explained by the climate factors (relative to the total economic output variance) and their statistical significances. Since a key finding of this study is based on the sub-national regions, such regional model training and performance information are critical. These will provide a more objective measure of the projection uncertainty due to the empirical model deficiency and from the small contribution relative to non-climate factors. I would also hope that the authors provide an interpretation of what it means for a projected dramatic drop in future economic output from a minor climate effect – an extremely unstable condition.

We appreciate the concerns of Referee #1 regarding overfitting and insignificant model parameters and their problems for out-of-sample prediction. We would first of all like to highlight again that our approach in this study (as in most climate econometric studies focused on identifying impacts) is one of causal inference of climate impacts by isolating plausibly exogenous variation, rather than prediction of regional GDP. In other words, we are not interested in explaining the maximum variation in regional GDP, but in identifying plausibly causal relationships between climate shocks and GDP which are globally generalizable for assessing the damages of future climate change.

For these reasons, we do not think it is appropriate or necessary to rely on metrics of fit such as within-region R-squared to evaluate our empirical models (as also emphasized by Referee #1 themselves in their first comment). Instead, our model-selection procedure is based on a) logical use of fixed-effects to isolate plausibly exogenous variation which allows for maximum confidence in a causal interpretation (see the now extended methods section "Empirical specification - fixed-effects distributed lag models"), b) the use of previous climate-econometric literature at both the sectoral and aggregate level to guide the choice of climate variables with relevance for economic growth (see Kalkuhl & Wenz 2020, Kotz et al 2021, Kotz et al 2022 for these analyses which are now summarized more explicitly in our main methods on L491-501), c) a logical identification strategy to distinguish a lower-bound of the persistence with which climate shocks impact growth (see main text "A robust lower-bound on the persistence of climate impacts on growth" on L.59-114 and the now extended methods section "Empirical specification - fixed-effects distributed lag models" on L. 538-592), and d) the use of statistical significance to assess that lower bound on impact persistence (see ED. Fig. 1 and Tables S1).

Nevertheless, as part of the last round of revisions, we provided detailed additional analyses using Information Criteria to assess whether our model does indeed suffer from overfitting in terms of the selection of an appropriate number of lags (currently Fig. S1). Expanding upon this analysis, we now also present a similar analysis using Information Criteria to assess overfitting with regards to the use of different climate variables (Table S2). These analyses demonstrate that the model does not suffer from overfitting in either way. We appreciate that some of these robustness tests from the previous round of reviews may have been obscured by the extensive updates, and have therefore pooled all discussion related to overfitting into the new SI Section S1 entitled "Avoiding overfitting".

Moreover, we explicitly remove parameters which are insignificant, as discussed in the caption to ED Fig. 1 and now mentioned on L95-96 of the updated manuscript. Please see Figs. S2-S4 and Tables S3-5 for the main models in which insignificant parameters have been removed. As such we believe that the model has been extensively assessed using different methods (significance of parameters, information criteria, as well as Monte-Carlo simulations) with regards to its robustness.

Finally, Referee #1 questions how a model which explains a small amount of economic variance in the historical period can reliably be used to project large future damages. The distinction between inference and prediction is again important here. Rather than trying to optimally predict regional GDP, our empirical models follow extensive climate-econometric literature (see e.g. the review by Auffhammer 2014) in trying to identify plausibly causal impacts on GDP growth rates. Given the fact that future climate change is large compared to the historical fluctuations available for causal inference, when projecting future damages with these empirical models, larger impacts than those observed in the historical period are inevitable. This is therefore a feature of all climate econometric studies which use historical fluctuations to infer plausibly causal climate impacts from data. Nevertheless, these models still constitute the most state-of-the-art methods for causal impact inference with which to project future damages from climate change. The alternative is to say that such an empirical assessment is simply not possible. In our study, we take steps to avoid out-of-sample extrapolation by limiting the projections of future moderating variables to the 95th percentile of what was observed historically, so that we do not extrapolate the marginal effects of our empirical models outside of the range of that in which they were identified (see L626-629). We appreciate that this is a nuanced issue, and as such have gladly provided additional discussion on this point on L. 310-322 of the updated manuscript.

References:

Shao, Jun. “Linear Model Selection by Cross-Validation.” *Journal of the American Statistical Association*, vol. 88, no. 422, 1993, pp. 486–94.

Stone, M. “An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike’s Criterion.” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, 1977, pp. 44–47.

Newell, Richard G., Brian C. Prest, and Steven E. Sexton. "The GDP-temperature relationship: implications for climate change damages." *Journal of Environmental Economics and Management* 108 (2021): 102445.

Kalkuhl, Matthias, and Leonie Wenz. "The impact of climate conditions on economic production. Evidence from a global panel of regions." *Journal of Environmental Economics and Management* 103 (2020): 102360.

Kotz, Maximilian, et al. "Day-to-day temperature variability reduces economic growth." *Nature Climate Change* 11.4 (2021): 319-325.

Kotz, Maximilian, Anders Levermann, and Leonie Wenz. "The effect of rainfall changes on economic production." *Nature* 601.7892 (2022): 223-227.

Auffhammer, Maximilian, et al. "Using weather data and climate model output in economic analyses of climate change." *Review of Environmental Economics and Policy* (2013).

Response to Referee #2

The attempts to address the referee comments and the detailed replies are greatly appreciated. However upon re-reading the paper it is apparent that the very same problems persist.

We thank the Referee for their detailed attention to our manuscript. However, it appears that a fundamental mis-understanding of the distinction between growth- and level-effects has occurred which we aim to clarify below in our point-by-point response, as well as in a new exposition of our empirical framework on L. 538-592 of the updated manuscript.

Please further find responses to the other points raised regarding the motivation and robustness of the empirical specification below.

In the revised manuscript, please find edits from the previous and current round of revisions in red and blue respectively. Line, figure and table numbers refer to the most recently revised version of the manuscript.

One issue is terminological and could have been fixed, but has not. In economics (please note that this is not a paper in physics), we talk of growth effects to imply that a change (or shock) leads to a permanent change in growth. For instance, a new road (or rise in temperature from a to b) increases growth from 4% to 5%. Formally, this is captured by a first difference on the LHS of a regression, such as:
$\Delta Y\_t = \alpha + \beta\Delta X\_t + \varepsilon\_t$ (1)
where ΔY_t is the change in the dependent variable Y between time t and t-1, ΔX_t is the change in the independent variable X between time t and t-1, which is similar to the specification in this paper.

We would like to point out that the above equation is simply a first-differenced form of a level-effects model, not a growth-effects model, as explained clearly on P7 of Newell et al 2021, in Olvera et al., 2022 (Fig. 1) and more indirectly in Dell et al. 2012 (P72-73, equation 3) and Kalkuhl & Wenz 2021 (equation 5). To see this, consider a level-model which is specified simply by a direct dependence of the level of output Y on an independent variable X,

$Y\_t = \alpha + \beta X\_t + \varepsilon\_t$ < (R1) basic model with level-effect dependence.

By simply taking the first difference of the above equation one arrives at the level-effect model in its first-differenced form:

$\Delta Y\_t = \alpha + \beta\Delta X\_t + \varepsilon\_t$ < (R2) level-effect model in first-differenced form.

The equivalence of these two equations is made clear in the study cited by Referee #2 (Newell et al. 2021) on P7. One can see that both of the above equations clearly imply that a permanent effect on the growth-rate (ΔY_t) would require the independent variable X_t to be changing *permanently*, not that a *single* change or shock in that independent variable X_t would lead to a permanent change in the growth rate as suggested by Referee #2.

In addition to Newell et al 2021, please also see Dell et al. 2012 (P72-73, equation 3), in which the level-effect of a model with economic growth rates as the dependent variable is specified as β in the following equation, whereas the growth-effect is specified as $\boldsymbol{\gamma}$:

$$\Delta Y_t = \alpha + (\beta + \boldsymbol{\gamma})X_t - \beta X_{t-1} = \alpha + \beta\Delta X_t + \boldsymbol{\gamma}X_t$$

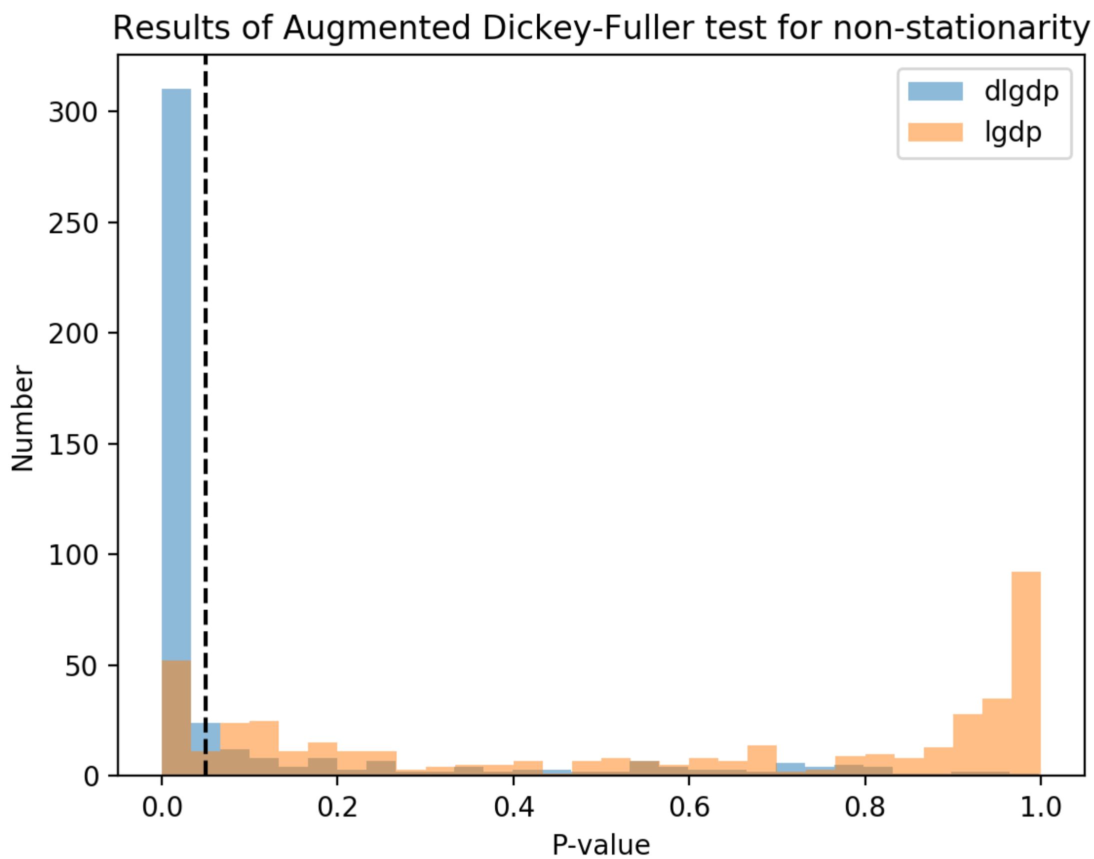
Again, this makes clear that a model with the economic growth rates as the dependent variable and the first difference of the independent variable is a level-effect model simply in first-differenced form.

Or the specification may be of the form as in many other papers in econometrics:
$\Delta Y_t = \alpha + \beta X_t + \varepsilon_t$ (2)
where ΔY_t is the change in the dependent variable Y between time t and t-1, X_t is the level of the independent variable X between time t and t-1, and ε_t is the error term.

We agree that the above equation is the correct specification of a growth-effect model, in which there is a direct dependence of the growth rate of output ΔY_t on an independent variable X_t. It should also be apparent that a change to that equation that replaces the level X_t by its first difference ΔX_t cannot denote the same growth specification (as suggested by Referee #2) but implies a level effect as explained above.

This also generates a growth effect since a one time extension of the road (or new temperature level) improves growth forever. Statistically whether or not to difference the LHS and RHS will depend on whether unit root issues emerge (i.e. if these are cointegrating vectors and the order of cointegration). The choice ought to be determined by the statistical properties of the data aimed at minimizing spurious regression problems. This is why there is a literature on cointegration in panel and time series data.

We agree that the choice of whether to first-difference or not should be determined by the statistical properties of the data. As is commonly known, economic output levels (Y_t) are non-stationary, which is precisely the reason that Newell et al 2021 argue on P7 that a level-effect model should be evaluated on the basis of its first-differenced form using equation (R2) above, rather than (R1). The figure below shows the results of an Augmented Dickey-Fuller test on regional time series of our economic output data, using both the level of output (lgdp) and the growth rate (dlgdp). Results show that the level of output fails the test of stationarity in most regions, whereas the growth rate does not. This justifies the use of the first-differenced version of the level-effects model (R2) when evaluating level effects.



Adding lags allows one to examine if the growth rate reverses or accelerates – but it is still a growth effect (unless annulled).

By contrast a level effect is when there is a one-off increase in the dependent variable that does not persist. Thus the above might become $Y_t = \alpha + \beta X_t + \varepsilon_t$ or $Y_t = \alpha + \beta \Delta X_t + \varepsilon_t$

The authors seem to think that if there are no lags in equation (1) it becomes a level equation – this is not the case in economics. Perhaps the authors are assuming that a constant growth rate over time (β) is a level effect while varying rates are called "growth effects". This is not how the terms are conventionally used – including in the economics references in the paper (eg Newell et al 2021). But the rationale is unclear for claims such as Line 68, 69 and supp material.

As outlined above and justified by a careful reading of Dell et al 2012 and Newell et al 2021, equation (1) without any lags is precisely a level-effect model in first-differenced form, in economics or in any other discipline. Including lags in equation (1) or (R2) still specifies a level-effect, but one in which an impact can occur in a number of years following the initial change in the independent variable X_t. It is for this reason that we consider our specification to be conservative in providing a lower-bound on the extent of the persistence of impacts on the growth-rate (by having a baseline specification of level-effects), but which allows some persistence based on what is observable (by including lags with significant effects which account for any delayed effects).

In our revised manuscript, we now provide a much more extensive exposition of our empirical specification which explains this distinction between growth and level-effects explicitly. Please see the updated methods section "Empirical specification - fixed-effects distributed lag models" on L. 548-592 for further details.

We would also like to point out that this is nothing new we have come up with or that stems from Physics but that this approach of distinguishing between growth and level effects is well established in the climate economics community and literature (compare e.g. Dell et al., QJE, 2012 (P73 equation 3), Newell et al., 2021 (P7), and Olvera et al., 2022 Figure 1 for a nice visualization to facilitate understanding).

Next I had requested robustness tests of the specification used since there is no economic theory used to guide the specification. This of course is standard practice. Instead the authors reply with reference to autocorrelation tests, (rightly) brought up by another referee. And cite their own papers to justify this specification. Neither of these responses answers the question raised. Autocorrelation is about memory in the data over time that impacts regression coefficients, muticollinearity which is addressed is about correlation between variables. This is quite different from knowing if a plausible change of specification (with or without autocorrelation and multicollinearity) will alter the results. With over 800 specifications examined in the literature it would be good to know if this one is robust. There is good reason to be cautious as metanalyses have shown that growth (as opposed to level of GDP) models are fragile. To quote the standard reference here: "Across just those growth models that specify a non-linear temperature function, the combined model and sampling uncertainty yield a standard deviation of predicted impacts equal to 132% of GDP, ……. specifying impacts on GDP levels, not growth, yield far less uncertainty in climate impacts; the standard deviation is equal to less than 3% of GDP ……." Newell et al 2021)
Hence a single specification will not make the case for those familiar with the literature.

We agree that the choice of model specification is a crucial part of our study and that it should be guided by economic theory as well as theory and evidence from other disciplines (as studying climate change and its economic impacts naturally requires knowledge from several disciplines). In this case, specification choices may refer to one of the following categories:

1. The choice of fixed-effects.
2. The choice of primary climate variables.
3. The choice of interaction terms.
4. The specification of growth- or level-effect models and the choice of the number of lags.

Point (1) is guided by extensive climate econometric literature which uses fixed-effects to isolate plausibly exogenous variation. Please see (Auffhammer 2013, Carleton & Hsiang 2016) for a summary, as well as L491-509 of our updated manuscript.

Points (2) and (3) are guided by extensive work in previous peer-reviewed studies in top journals (https://www.sciencedirect.com/science/article/pii/S0095069620300838, https://www.nature.com/articles/s41586-021-04283-8$, https://www.nature.com/articles/s41558-020-00985-5). In particular, the use of temperature variability and different precipitation characteristics are guided by sectoral level studies which identified impacts of such variables on important components of economic growth such as agriculture, health, labor outcomes and flood damages. The use of these variables is then checked with extensive robustness tests of alternative specifications in both papers. The choice of interaction terms is guided by intuition regarding potential adaptation mechanisms and numerous tests of different specifications. We referred to these papers in our justification of our model specification to avoid re-iterating previous results, but now provide more explicit recapitulation of the motivation and tests of these climate variables in L491-509 of the updated methods sections. We further show in the updated manuscript in Table S2 that Information Criteria indicate that the choice of climate variables is appropriate in optimizing the fit of the data while avoiding overfitting.

In this particular study, we follow the peer-reviewed literature referenced above in satisfying points (1-3) regarding the specification choices outlined above, and focus on addressing point (4) explicitly. Here, we follow a well-established literature identifying growth- or level-effects and the extent of impact persistence. As demonstrated above, we follow Newell et al. 2021 and Kalkuhl & Wenz (2020) in using a baseline specification of level-effects (meaning that our manuscript falls into the category of less "fragile" models which project damages with standard deviations equal to less than 3% of GDP according to Newell et al 2021), but include statistically significant lags to account for further delayed effects. This empirical specification is designed to provide a robust lower-bound on the extent of impact persistence. In the updated manuscript we now provide a more extensive exposition of this empirical framework in the methods section "Empirical specification - fixed-effects distributed lag model" on L538-592. Moreover, as outlined in the revised manuscript, we provide extensive tests of this part of the empirical specification choice, including:

- Assessments of whether including further lags causes overfitting (see the new Supplementary Methods Section S1 and Fig. S1)
- Assessments of whether including lags causes issues due to auto-correlation (Supplementary Methods Section S2, Figs. S5-S7).
- Assessment of whether including lags causes unrealistic oscillations in parameter estimates (Supplementary Methods Section S4, Figs. S9 & S10).

The results of all of these tests find our empirical specification to be robust.

Finally, whichever way the authors wish to cut things a 19% drop in GDP over 25 to 30 years remains within margins of routine statistical corrections – this is unarguably true and suggests that inflated interpretations will only serve to undermine the credibility of the paper among those that are familiar with macroeconomic statistics.

As noted in our response to the Referee's original comments, as far as we can tell, our manuscript contains no subjective interpretation of the magnitude of damages which we project. Only factual comparisons between values are provided such as:

- Between damages and mitigation costs
- Of damages between regions
- Of damages across different specifications.

We encourage the Referee to point out such inflated interpretations such that we can adjust them where they are present.

References:

Newell, Richard G., Brian C. Prest, and Steven E. Sexton. "The GDP-temperature relationship: implications for climate change damages." *Journal of Environmental Economics and Management* 108 (2021): 102445.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. "Temperature shocks and economic growth: Evidence from the last half century." *American Economic Journal: Macroeconomics* 4.3 (2012): 66-95.

Kalkuhl, Matthias, and Leonie Wenz. "The impact of climate conditions on economic production. Evidence from a global panel of regions." *Journal of Environmental Economics and Management* 103 (2020): 102360.

Bastien-Olvera, Bernado A., Francesco Granella, and Frances C. Moore. "Persistent effect of temperature on GDP identified from lower frequency temperature variability." *Environmental Research Letters* 17.8 (2022): 084038.

Auffhammer, Maximilian, et al. "Using weather data and climate model output in economic analyses of climate change." *Review of Environmental Economics and Policy* (2013).

Carleton, Tamma A., and Solomon M. Hsiang. "Social and economic impacts of climate." *Science* 353.6304 (2016)

Referee #3 (Remarks to the Author):

The authors have done an excellent job of addressing the previous set of reviewer comments. In particular, the added robustness testing and methodological detail strengthens the paper and will help to improve reproducibility. I have no further suggested improvements.

We are very happy that we satisfactorily addressed Referee #3's concerns and thank them for the very constructive and helpful feedback.

Reviewer Reports on the Second Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

I would like to express my gratitude to the authors for their excellent and comprehensive response to most of my previous comments. The statistics pertaining to the model selection have been significantly enhanced. However, further revisions are helpful to address the following relatively minor issues:

[1] I would suggest that the authors soften their overconfidence regarding model overfitting. Even using AIC and BIC to assess model selection, the outcome can only be considered in a relative sense. That is, the model with a lower AIC or BIC value is a better choice to minimize overfitting. However, this does not imply that the model selected completely "avoids" or "does not suffer from" overfitting.

[2] I do not entirely agree with the authors' argument regarding cross-validation. Specifically, using the study's emphasis on "causal inference rather than prediction" as a justification for not adopting the cross-validation approach is not a valid reason. This study not only focuses on "the inference of the plausibly causal impacts of climatic conditions on economic growth" but also utilizes the statistically inferred model to project future changes at the subnational level. If the selected statistical model fails to effectively capture the historical signals, such as the largest variance explained by climate variations when compared to other models, how can one be confident in its robustness for future projections?

[3] I would also suggest that the authors tone down their statement regarding the selection of specific climate variables and their maximum lags, which they claim to be supported by "robust evidence." It is worth noting that this so-called evidence relies solely on statistical data inference rather than established physical mechanisms. Additionally, some of the references cited by the authors to support this evidence are their own publications, while others do not fully account for all the variables and lags in question. This seems to create a circular argument. It might be beneficial for the authors to acknowledge the limitations of relying solely on statistical data inference, without disregarding the importance of physical mechanisms.

Referee #2 (Remarks to the Author):

The clarifications in the revised version address some of the issues that were raised. Overall the supplementary materials and the more technical material provide more detail and are much improved. However several concerns persist and the authors have not addressed some of these.
In particular what the authors call level effects seems to be a description of what is termed a steady state equilibrium in economics.
Further the discussion of the robustness of the results to alternative specifications in the main text seems to be - "other literature suggests the work is robust." . But we also know from previous work that robustness in one specification will not carry to another. So the reluctance to demonstrate that the results are robust to alternative specifications could be a concern.
Finally the rebuttal argues that ".. our manuscript contains no subjective interpretation of the magnitude of damages which we project." It may be observed that line 135 asserts ".., this constitutes a substantial reduction …" or line 311 ". …projections of reductions of income of 19% may appear large." And so on. A 20% reduction over (say) 50 years implies a 0.4% decline per annum - within the margin of statistical errors in GDP corrections in many a circumstance.

Author Rebuttals to Second Revision:

Referee #1 (Remarks to the Author):

I would like to express my gratitude to the authors for their excellent and comprehensive response to most of my previous comments. The statistics pertaining to the model selection have been significantly enhanced. However, further revisions are helpful to address the following relatively minor issues:

We are glad that the referee found our amendments to have comprehensively addressed their concerns regarding model selection and robustness.

We also appreciate their minor comments regarding the framing of our arguments, which we have addressed in the revised manuscript (highlighted in green in the updated manuscript whereas amendments from previous revisions are marked in red and blue).

[1] I would suggest that the authors soften their overconfidence regarding model overfitting. Even using AIC and BIC to assess model selection, the outcome can only be considered in a relative sense. That is, the model with a lower AIC or BIC value is a better choice to minimize overfitting. However, this does not imply that the model selected completely "avoids" or "does not suffer from" overfitting.

We thank the referee for this fair comment regarding wording. Throughout the text and SI we have made a number of changes to reflect the fact that Information Criteria only provides guidance on relative model selection, and that these methodological choices can therefore not *completely avoid* the possibility of overfitting. The phrasing "avoid overfitting" has been changed to "limit overfitting" throughout the manuscript and SI. Furthermore, on L97-101 of the main manuscript we now refer to these tests using the following language which we believe better reflects the strengths of the tests and the concerns of the referee regarding overstatements of confidence:

"Furthermore, evaluation by means of Information Criteria indicates that the inclusion of all five climate variables and the use of these numbers of lags provide a preferable trade-off between best-fitting the data and including additional terms which could cause overfitting, in comparison to model specifications excluding climate variables or including more or fewer lags (Supplementary Methods Section S1, Fig. S1 and Table S3)."

And in the supplementary Information Section S1 we provide additional description of the interpretation of these Information Criteria:

"BIC and AIC are evaluated using a trade-off between the maximized likelihood function and penalties for additional terms in the model which could result in overfitting. As such, they can be used to assess the relative strength of different models in terms of best describing the data and limiting the possibility of overfitting."

[2] I do not entirely agree with the authors' argument regarding cross-validation. Specifically, using the study's emphasis on "causal inference rather than prediction" as a justification for not adopting the cross-validation approach is not a valid reason. This study not only focuses on "the inference of the plausibly causal impacts of climatic conditions on economic growth" but also utilizes the statistically inferred model to project future changes at the subnational level. If the selected statistical model fails to effectively capture the historical signals, such as the largest variance explained by climate variations when compared to other models, how can one be confident in its robustness for future projections?

We completely agree that selecting models based on their ability to capture historical signals is of central importance, and that cross-validation could provide one way to do this, particularly in contexts where the primary objective is the prediction of economic growth. In our context, however, while our projections do look into the future, they should not be interpreted as a prediction of economic growth. This is because, following the literature, both our empirical models and our projections assume to be constant the many important non-climatic factors which contribute to changes in economic outcomes but are very difficult to predict (e.g. wars, pandemics, and even structural changes such as technology). As such, our projections should be considered an assessment of the expected exogenous impact of future climate conditions on the economy from a future baseline specified by the socioeconomic projections. It is primarily this point to which we aimed to refer when discussing the difference between inference and prediction. We have added additional text on this important point on L. 130-135 of the updated manuscript:

"Following a well-developed literature12,17,19, these projections do not aim to provide a prediction of future economic growth. Instead, they are a projection of the exogenous impact of future climate conditions on the economy relative to the baselines specified by socioeconomic projections, based on the plausibly causal relationships inferred by the empirical models, and assuming ceteris paribus. Other exogenous factors relevant for the prediction of economic output are purposefully assumed constant."

Given that the primary objective of our model is not one of prediction, and given that Information Criteria fulfill a very similar role to cross-validation in terms of selecting models which explain the largest variance while minimizing the possibility of overfitting (indeed we note that these techniques are in some cases asymptotically equivalent (Stone 1997)), we do not think it is necessary to use cross-validation as a further basis for model selection. Indeed, supplementary Section S1 documents how we use Information Criteria to select the combination of climate variables and number of lags which provide a preferable trade-off between explaining the maximum amount of variance and limiting overfitting, giving confidence in its robustness for future projections. Nevertheless, we believe that cross-validation may well provide a fruitful avenue for future research which in this context is beyond the scope of necessary steps to ensure the robustness of our empirical models. We have substantially amended the text discussing cross-validation to reflect these points, which we copy here from SI Section S1.3:

"AIC and BIC metrics support our choice of climate variables and number of lags, indicating that they provide a preferable trade-off between maximizing variance and limiting overfitting. Alternative methods exist which could fulfill similar functions in selecting models which optimize this trade-off. In particular, cross-validation provides an asymptotically equivalent approach6, which may be particularly attractive in the context of prediction problems. Cross-validation splits the available data into two parts, first training the empirical model with one set before testing it on the other. This yields a direct evaluation of the ability of the empirical model to predict new data. The aim of this paper, however, is not to accurately predict economic growth, but to project the exogenous impact of future climate conditions on the economy, based on robustly inferred causal relationships, and assuming ceteris paribus (compare previous climate-economy literature, e.g. refs. (1,11,13)). That is, factors important for predicting economic growth such as technological development, wars, pandemics and financial crises are assumed constant. As a consequence, the main objective of the model selection procedure is to provide a robust identification strategy for causal inference7–9. In particular, our empirical model is based on a careful selection of fixed-effects and regional time-trends to isolate variation in climate and economic growth which are plausibly exogenous, and a careful choice of climate variables in their first-differenced form with a number of lags to provide a lower-bound on the persistence of impacts on growth (see main text section "A robust lower bound on the persistence of climate impacts on growth" and methods section "Empirical models – fixed-effects distributed lag models"). Given this emphasis on inference rather than prediction in the identification of plausibly causal empirical models and the projection of exogenous impacts; the asymptotic equivalence of Information Criteria and cross-validation for model selection6; and the fact that AIC and BIC indicate that our empirical models already provide a preferable trade-off between maximizing variance and limiting overfitting, we do not pursue cross-validation as a further method for model selection. Cross-validation nevertheless offers an interesting avenue for further work on the prediction of economic growth in the context of climate impacts which is beyond the scope of this manuscript."

[3] I would also suggest that the authors tone down their statement regarding the selection of specific climate variables and their maximum lags, which they claim to be supported by "robust evidence." It is worth noting that this so-called evidence relies solely on statistical data inference rather than established physical mechanisms. Additionally, some of the references cited by the authors to support this evidence are their own publications, while others do not fully account for all the variables and lags in question. This seems to create a circular argument. It might be beneficial for the authors to acknowledge the limitations of relying solely on statistical data inference, without disregarding the importance of physical mechanisms.

We thank the referee for this comment regarding the phrasing of our manuscript. We first would like to note that the choice of climate variables is guided by physical mechanisms for which there is extensive empirical evidence outside of our own studies. For example, the impacts of temperature on agricultural (Lobell et al. 2013, Zhao et al 2017) and labor productivity (Dasgupta et al. 2021), of daily temperature variability on agricultural output (Wheeler et al. 2000, Rowhani et al. 2011, Ceglar et al. 2016) and human health (Shi et al. 2015, Xue et al. 2019), as well as of precipitation on agriculture, metropolitan labor outcomes and flood damages (Liant et al. 2017, Desbreaux et al. 2019, Damania et al. 2020, Davenport et al. 2021, Dave et al. 2021). These physical mechanisms are now listed explicitly with a number of references to empirical studies conducted by other authors in Table S1. Furthermore, we refer to this Table and summarize these mechanisms (while citing the studies on growth impacts which discuss and cite them in their introductions) on L.53-57 of the main manuscript, and L. 516 of the main methods section.

L53-57:
"The selection of these climate variables follows micro-level evidence for mechanisms related to the impacts of average temperatures on labor and agricultural productivity[17], of temperature variability on agricultural productivity and health[13], as well as of precipitation on agricultural, labor outcomes, and flood damages[14] (see Table S1 for an overview including more detailed references)."

L. 510-516:
"Assessments of daily temperature variability were motivated by evidence of impacts on agricultural output and human health, as well as macroeconomic literature on the impacts of volatility on growth when manifest in different dimensions such as government spending, exchange rates and even output itself[13]. Assessments of precipitation impacts were motivated by evidence of impacts on agricultural productivity, metropolitan labor outcomes and conflict, as well as damages caused by flash flooding[14]. See Table S1 for detailed references to empirical studies of these physical mechanisms."

Following the request of the reviewer, we have toned down our language by removing specific use of the phrase "robust evidence" from the main text. Instead, we allow the reader to assess the robustness of the results themselves by referring to the various robustness tests. These include robustness tests conducted in previous studies regarding the choice of climate variables as outlined on L.516-521, such as:

- using multiple climate data-sets,
- using different spatial-aggregation schemes,
- using different specifications of time-trends and error-clustering,

as well as those regarding the choice of climate variables and number of lags conducted in the present study as described on L. 93-122, such as:

- AIC/BIC to assess the inclusion of climate variables in Table S3;
- Statistical significance and AIC/BIC to assess the number of lags in Tables S2 & S4-6, SI section S1 and Fig. S1;
- Monte-Carlo simulations to demonstrate robustness to auto-correlation (SI Section S2, Figs. S6&S7) and cross-correlations (SI Section S3, Fig. S8);
- The use of restricted lag-models to limit oscillations in parameter estimates (SI Section S4, Figs. S9 & S10);

- Robustness tests of the extent to which physical climate models accurately reflect the climate variables of interest (Table S8);
- Robustness tests of the timescales at which the moderating variables of the empirical models are evaluated under future projections (Fig. S11-S12);
- Robustness tests of the method via which sub-national price changes are accounted for (Fig. S13).

References:

Stone, Mervyn. "An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion." *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977): 44-47.

Dasgupta, Shouro, et al. "Effects of climate change on combined labour productivity and supply: an empirical, multi-model study." *The Lancet Planetary Health* 5.7 (2021): e455-e465.

Lobell, David B., et al. "The critical role of extreme heat for maize production in the United States." *Nature climate change* 3.5 (2013): 497-501.

Zhao, Chuang, et al. "Temperature increase reduces global yields of major crops in four independent estimates." *Proceedings of the National Academy of sciences* 114.35 (2017): 9326-9331.

Wheeler, Timothy R., et al. "Temperature variability and the yield of annual crops." *Agriculture, Ecosystems & Environment* 82.1-3 (2000): 159-167.

Rowhani, Pedram, et al. "Climate variability and crop production in Tanzania." *Agricultural and forest meteorology* 151.4 (2011): 449-460.

Ceglar, Andrej, et al. "Impact of meteorological drivers on regional inter-annual crop yield variability in France." *Agricultural and forest meteorology* 216 (2016): 58-67.

Shi, Liuhua, et al. "Impacts of temperature and its variability on mortality in New England." *Nature climate change* 5.11 (2015): 988-991.

Xue, Tao, et al. "Declines in mental health associated with air pollution and temperature variability in China." *Nature communications* 10.1 (2019): 2165.

Liang, Xin-Zhong, et al. "Determining climate effects on US total agricultural productivity." *Proceedings of the National Academy of Sciences* 114.12 (2017): E2285-E2292.

Desbureaux, Sébastien, and Aude-Sophie Rodella. "Drought in the city: The economic impact of water scarcity in Latin American metropolitan areas." *World Development* 114 (2019): 13-27.

Damania, Richard. "The economics of water scarcity and variability." *Oxford Review of Economic Policy* 36.1 (2020): 24-44.

Davenport, Frances V., Marshall Burke, and Noah S. Diffenbaugh. "Contribution of historical precipitation change to US flood damages." *Proceedings of the National Academy of Sciences* 118.4 (2021): e2017524118.

Dave, Raviraj, Srikrishnan Siva Subramanian, and Udit Bhatia. "Extreme precipitation induced concurrent events trigger prolonged disruptions in regional road networks." *Environmental Research Letters* 16.10 (2021): 104050.

Referee #2 (Remarks to the Author):

The clarifications in the revised version address some of the issues that were raised. Overall the supplementary materials and the more technical material provide more detail and are much improved. However several concerns persist and the authors have not addressed some of these.
In particular what the authors call level effects seems to be a description of what is termed a steady state equilibrium in economics.
Further the discussion of the robustness of the results to alternative specifications in the main text seems to be - "other literature suggests the work is robust." . But we also know from previous work that robustness in one specification will not carry to another. So the reluctance to demonstrate that the results are robust to alternative specifications could be a concern.

As outlined in our previous response, the robustness of our results rests on a number of different lines of evidence. First, there is evidence from previous studies which use the same data and methods and conduct numerous robustness tests to identify specific climate variables with impacts on economic output. To the extent that these studies use exactly the same data and methods, they are highly relevant to this study. We explicitly list these robustness tests on L. 516-520 of the methods section, which include:

- using multiple climate data-sets,
- using different spatial-aggregation schemes,
- using different specifications of time-trends and error-clustering.

Second, we explicitly provide extensive new robustness tests which pertain to the specific empirical challenges faced in this context, described in detail on L.93-122 of the main manuscript, including:

- AIC/BIC to assess the inclusion of climate variables in Table S3;
- Statistical significance and AIC/BIC to assess the number of lags in Tables S2 & S4-6, SI section S1 and Fig. S1;
- Monte-Carlo simulations to demonstrate robustness to auto-correlation (SI Section S2, Figs. S6&S7) and cross-correlations (SI Section S3, Fig. S8);
- The use of restricted lag-models to limit oscillations in parameter estimates (SI Section S4, Figs. S9 & S10);
- Robustness tests of the extent to which physical climate models accurately reflect the climate variables of interest (Table S8);
- Robustness tests of the timescales at which the moderating variables of the empirical models are evaluated under future projections (Fig. S11-12);
- Robustness tests of the method via which sub-national price changes are accounted for (Fig. S13).

If there are any further specific robustness tests which the referee thinks are necessary and would therefore like to see, we would be happy to look into this. But so far, we think that these extensive tests encompass all considerations which are necessary to support our main conclusions.

Finally the rebuttal argues that ".. our manuscript contains no subjective interpretation of the magnitude of damages which we project." It may be observed that line 135 asserts ".., this constitutes a substantial reduction …" or line 311 ". …projections of reductions of income of 19% may appear large." And so on. A 20% reduction over (say) 50 years implies a 0.4% decline per annum - within the margin of statistical errors in GDP corrections in many a circumstance.

The referee appears to have misunderstood the nature of the damages we project. Damages do not refer to percentage changes from the present level of economic output, but from a baseline specified by socioeconomic projections. This is stated clearly in the axes' labels of all figures as well as in the main text on L. 146. As such, the 20% reduction refers to a permanent reduction in income levels compared to a scenario without future climate change, and should not be evaluated in terms of a year-on-year change. A 20% permanent reduction is therefore far outside of the margin of statistical errors in GDP corrections. We now emphasize the fact that this is a permanent reduction in income on L149 to which the referee referred. Regarding the wording, on L 150-152 we have removed the wording "substantial" and now simply quote the magnitude of projected impacts, such that it now reads:

"Even though levels of income per capita generally still increase relative to those today, this constitutes a permanent income reduction for the majority of regions, including North America and Europe (each with median income reductions of approximately 11%) and with South Asia and Africa being the most strongly affected (each with median income reductions of approximately 22%; Fig. 1)."

Moreover, we note that the statement on L325 that reductions "may appear large" is a relative statement comparing the magnitude of damages to the fraction of variance explained by the empirical models, and as such we have not changed this.

Reviewer Reports on the Third Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

The authors have made commendable efforts in thoroughly revising the manuscript to address the concerns raised in my previous review. The meticulous explanation given to model overfitting, cross-validation, and result robustness is truly appreciated.

I am pleased to confirm that the revisions have significantly bolstered the manuscript's quality, rendering it ready and accessible for publication. The substantial contributions it offers to the field make it a valuable addition to the existing literature.

Referee #4 (Remarks to the Author):

The paper addresses an important topic and contains a lot of sophisticated analysis. It is also well written and explained. The general problem with the climate econometrics literature, to which this paper belongs, is that the results, in particular any future projections, depend sensitively on model specification choices when trying to estimate the effect of historical climate fluctuations on economic outcomes. Some of these specification choices are quite obscure and the biases that can creep in are sometimes hard to see. In fairness, the authors are acutely aware of this and some of these specification choices are at the heart of the paper, particularly (i) the persistence of impacts of a climate shock on GDP (i.e., whether climate shocks affect the level or the growth rate of GDP) and (ii), at the prompting mainly of R1, serial correlation of the climate variables and multi-collinearity between the climate variables.

I find all of this well explained and fairly convincing, yet, purely subjectively, I have a hard time in believing the results, which seem unintuitively large given damages aren't perfectly persistent. In many ways this is an unfair critique because it is just an intuition. Yet, it is worth probing further the possible sources of bias, because we know from the experience with Burke, Hsiang and Miguel (2015 in Nature) that publishing numbers in high-impact journals, which subsequently essentially get discredited, can create a lot of confusion. This is what R2 is getting at, and yet in my view R2 hasn't found a 'smoking gun'. Rather, they are just raising a generic concern about model specification. It doesn't help that the paper relies on previously published work – it is unreasonable for referees to evaluate a whole history of published work and difficult to take it on faith that previous publications have always been carefully and properly refereed. But still, given how generic R2's complaint is here, it is difficult to know concretely what the authors could reasonably do to respond. Therefore, I side with the authors here. I also disagree with R2 that the estimates in the paper should be interpreted as small – as I said, I think they are big!

For my part, I wonder a lot about the spillovers from climate shocks across space. This is a concern in national-level studies, but it is much more of a concern in studies with sub-national resolution, because economies are densely interconnected both on the production and consumption side. A climate shock in one sub-national unit likely causes spillover effects in neighbouring units, which could amplify or dampen the local shock. We know from recent work, for example, that firms respond to climate damage to their production facilities by shifting production to their other non-affected facilities. From an econometric perspective, this is a threat to identification because it potentially violates the stable unit treatment value assumption. So, I really want to know what the authors have done in this and the previous work they leverage to explore this issue.

The other way the authors could allay my fears would be to justify how impacts of the size they estimate are consistent with the growth experience we actually observe across countries and time. In some sense, their approach guarantees this of course, because they estimate the economic impacts of climate shocks using historical data. But, due to the challenges of proving their model is not mis-specified, this is not enough. If rising temperatures and associated rainfall changes impact GDP so much, and also differently across countries, should we not be seeing this manifest itself in terms of regional patterns of convergence or non-convergence? Granted, future climate change is much larger than past change, but the impacts over the next couple of decades are already projected to be large, so I would like the authors to walk me through why their estimates are not inconsistent with observed *long-term* growth patterns across countries.

My other comment is that I am struggling to understand the sense in the cost/benefit comparison exercise. I may have got this wrong, but I think this compares the total cost of climate change relative to a no-climate-change baseline with the cost of mitigating climate change to bring it from a baseline/bau path to a lower (but not zero) emissions path. If so, then I find this an illegitimate comparison and liable to cause confusion. The only legitimate comparison is between the costs and benefits of moving from path A (bau) to path B (Paris), but that is not what is done here. Moreover, the basic result about the economic commitment of climate change (not the most appealing terminology by the way, but fine) seems to imply that there would no statistically significant economic benefit of reducing emissions until mid-century, because two emissions scenarios that proxy for bau and Paris respectively give strongly overlapping confidence intervals of damages in the period up to 2050. To be clear, I am not for one minute questioning the basic claim that mitigating climate change has benefits greater than costs, but the comparison in this paper is not the right way to demonstrate it. It needs fundamental rethinking, or just to be removed.

Author Rebuttals to Third Revision:

Referee #1 (Remarks to the Author):

The authors have made commendable efforts in thoroughly revising the manuscript to address the concerns raised in my previous review. The meticulous explanation given to model overfitting, cross-validation, and result robustness is truly appreciated.

I am pleased to confirm that the revisions have significantly bolstered the manuscript's quality, rendering it ready and accessible for publication. The substantial contributions it offers to the field make it a valuable addition to the existing literature.

We are glad that our additional alterations to the manuscript satisfactorily addressed the concerns of the referee, and appreciate their comments regarding the quality and contribution of our manuscript. We thank the referee for their detailed comments and contribution to the peer review process which substantially improved the manuscript. Best wishes!

Referee #4 (Remarks to the Author):

The paper addresses an important topic and contains a lot of sophisticated analysis. It is also well written and explained. The general problem with the climate econometrics literature, to which this paper belongs, is that the results, in particular any future projections, depend sensitively on model specification choices when trying to estimate the effect of historical climate fluctuations on economic outcomes. Some of these specification choices are quite obscure and the biases that can creep in are sometimes hard to see. In fairness, the authors are acutely aware of this and some of these specification choices are at the heart of the paper, particularly (i) the persistence of impacts of a climate shock on GDP (i.e., whether climate shocks affect the level or the growth rate of GDP) and (ii), at the prompting mainly of R1, serial correlation of the climate variables and multi-collinearity between the climate variables.

We thank the referee for their detailed reading of our analysis and appreciation of the nuances of the topic and the methodological approaches which we have taken to address them here.

I find all of this well explained and fairly convincing, yet, purely subjectively, I have a hard time in believing the results, which seem unintuitively large given damages aren't perfectly persistent. In many ways this is an unfair critique because it is just an intuition. Yet, it is worth probing further the possible sources of bias, because we know from the experience with Burke, Hsiang and Miguel (2015 in Nature) that publishing numbers in high-impact journals, which subsequently essentially get discredited, can create a lot of confusion. This is what R2 is getting at, and yet in my view R2 hasn't found a 'smoking gun'. Rather, they are just raising a generic concern about model specification. It doesn't help that the paper relies on previously published work – it is unreasonable for referees to evaluate a whole history of published work and difficult to take it on faith that previous publications have always been carefully and properly refereed. But still, given how generic R2's complaint is here, it is difficult to know concretely what the authors could reasonably do to respond. Therefore, I side with the authors here. I also disagree with R2 that the estimates in the paper should be interpreted as small – as I said, I think they are big!

We are glad that the referee appreciates our difficulties in addressing the concerns of the previous reviewer in a reasonable way. Furthermore, we agree that the wide ranging estimates of future damages in the climate-econometric literature can indeed create confusion. However, we believe that this is a natural part of the scientific process in which continual methodological developments bring new insights which have implications for our assessments of future climate damages. For example, in our opinion the major insight and contribution of Burke et al. 2015 was that temperature impacts on economic output are non-linear, i.e. that they vary with baseline climatic conditions. Their projections were essentially an exploration of the implications of that insight, and in our opinion therefore remain a solid contribution to the literature, despite more recent studies having updated their assessment of projected damages based on other methodological developments and having questioned their underlying assumption of growth effects.

In fact, diverging assumptions on level or growth effects (i.e. the question of damage persistence) have been one of the main reasons for the wide range of damage estimates in the literature (see Fig. 4 of Kikstra et al. 2021). We believe our work has shed important light on that question by developing a careful empirical framework which provides a robust lower bound on the persistence of climate impacts on economic growth. Our projections then explore the implications of having constrained it, as well as the implications of our previous work which demonstrated that other climate conditions such as temperature variability and precipitation extremes have important additional impacts to average temperature changes (Kotz et al. 2021 & 2022).

Naturally, there are other open questions in the climate-econometric literature which still need to be addressed such as adaptation and, as the referee rightly points out, spillover effects. We discuss those remaining open questions in the Discussion section, as well as below.

For my part, I wonder a lot about the spillovers from climate shocks across space. This is a concern in national-level studies, but it is much more of a concern in studies with sub-national resolution, because economies are densely interconnected both on the production and consumption side. A climate shock in one sub-national unit likely causes spillover effects in neighbouring units, which could amplify or dampen the local shock. We know from recent work, for example, that firms respond to climate damage to their production facilities by shifting production to their other non-affected facilities. From an econometric perspective, this is a threat to identification because it potentially violates the stable unit treatment value assumption. So, I really want to know what the authors have done in this and the previous work they leverage to explore this issue.

We agree that there are a number of potential mechanisms related to spillovers which may be important for assessments of overall climate damages. These include the possible relocation of production from one region to another (arguably an adaptation response to impacts which would most likely occur over longer-timescales e.g. Acharya et al. 2023), as well as adjustment to local shocks through trade e.g. by compensating local production shortages of intermediate products through imports from neighboring regions (more likely a short-term spillover). The second mechanism could mitigate the extent of local impacts in response to a local shock, and is likely already captured by our empirical analysis because the local GDP which we measure includes adjustments that have occurred through trade. Our projections implicitly assume that buffering mechanisms via trade with neighboring regions would continue under future climate change, even though trade-partnered regions can simultaneously be affected by adverse future conditions and could therefore be less able to play a buffering role.

However, the knock-on impacts via trade or spatial-connections for neighboring regions themselves are not identified within our main empirical analysis. That is, the question of whether economic impacts in a given year are caused not only by weather shocks in the region itself but also by shocks in neighboring or trade partner regions that happen at the same time.

Interestingly, the econometric literature on the effects of spillovers of climate impacts on aggregate productivity indicates that spatially-neighboring regions experience impacts of the same sign as the local region in response to a local climate shock (Deryugina & Hsiang 2014 (counties in the US), Schleypen et al. 2019 (subnational regions in Europe), Dasgupta et al 2022 (subnational data globally), Neal 2023 (national data globally)). These studies therefore imply that additionally accounting for spillover effects in neighboring regions would more likely raise damage estimates than reduce them. This echoes other work (some of which our own) which explicitly models spillovers in terms of the dynamic response through trade networks. These studies indicate that the repercussions of individual climate shocks can propagate through supply chains to impact partnered regions (Midelanis et al. 2021, Malik et al 2022), and can amplify overall impacts on welfare when shocks in multiple regions overlap spatially or temporally as expected under climate change (Kuhla 2021).

We have added further discussion on the role of spillovers on L. 356-375 of the updated manuscript, including potential mechanisms, previous literature and the results of an additional analysis we conducted which indicates that fully accounting for spillover effects would likely increase our damage estimates.

Specifically, we have employed a spatial lag model to explore the role of spillover effects. Even though such an approach only accounts for spillover effects from neighboring regions (and not from trade partners in more distant places which would require granular data on global trade linkages between subnational regions which are not readily available), it can help gain some first-order insights into the role of spillovers. In this vein, we also used a simplified model with regards to persistence, including no temporal lags of the climate variables. This avoids estimating a very complicated model with multiple climate variables, interactions, temporal and spatial lags. See methods section L. 680-700 for more details. Results shown in Fig. S18 (copied below) indicate that accounting for these spillovers can increase the magnitude, and also the heterogeneity, of overall impacts from a climate shock. Consistent with previous literature, this analysis indicates that accounting for the role of spillovers fully may raise estimates of damages in comparison to our main analysis.

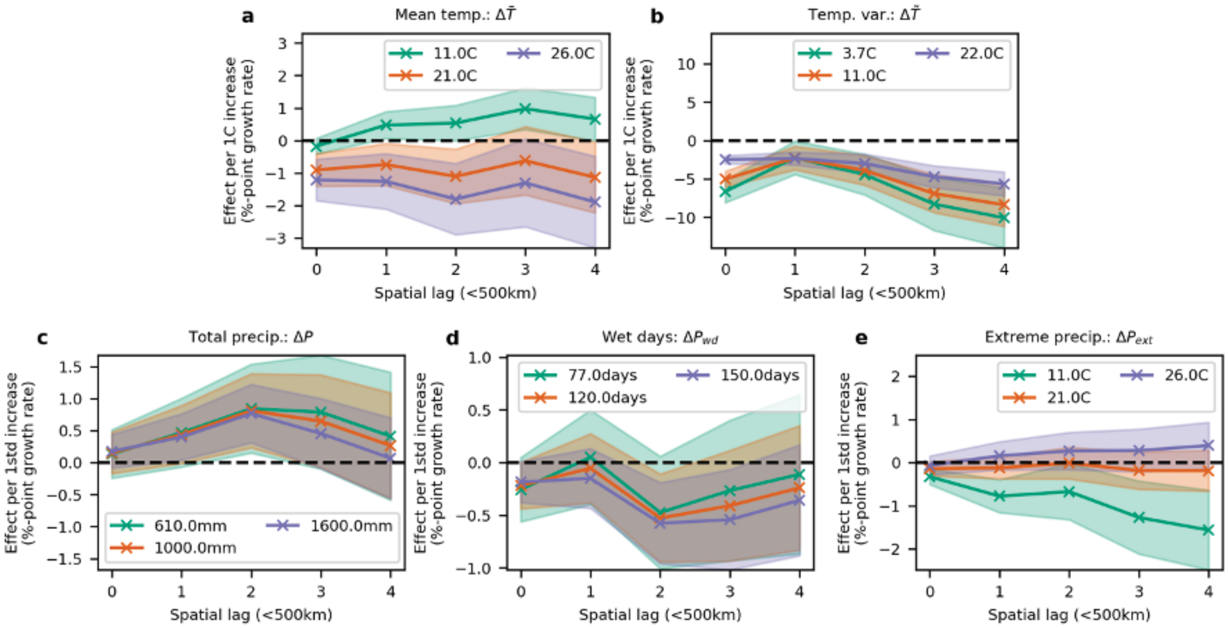


Fig. S18. Exploration of possible spill-over effects of contemporaneous climate impacts on spatially neighboring regions. Panels (a-e) show the cumulative impacts of different climate variables on economic growth rates when including the spatially lagged-effects of climate shocks in neighboring regions with centroids a distance of up to 500, 1000, 1500 and 2000km away (1, 2, 3 or 4 spatial lags, respectively). Spatial lags are constructed by taking the average of the first-differenced climate variables and their interaction terms over neighboring regions (see methods for detail). Due to data availability constraints, these models do not account for spill-overs which may occur via trade, and for simplicity they use no temporal lags of the climate variables, therefore only reflecting contemporaneous impacts. Error bars show the 95% confidence intervals having clustered standard errors by region.

However, this is just a first-order assessment. Further analysis which addresses both spatial- and trade-connected spill-overs, while also accounting for delayed impacts using temporal lags, will be necessary to adequately address this question fully. These approaches offer fruitful avenues for further research but are beyond the scope of the present manuscript which aims primarily to explore the impacts of different climate conditions and their persistence. We discuss this on l. 356-375 of the updated manuscript.

The first challenge to a comprehensive assessment of spillovers is that the construction of appropriate weights to assess which regions are “neighbors” is not a straightforward process (see for example the different approaches to do so in the econometric papers on spillovers which we reference above). Whether these weights should reflect spatial-metrics or trade relations is not clear, and if trade relations are important (as much literature demonstrates is the case) then constructing appropriate relations for sub-national regions is limited by data availability. Second, if climate change is likely to have major consequences for economies across the world with different magnitudes across regions (as our work and other work suggests), then it is almost certain that the structure of these dependencies will not remain constant in the future. Projecting future trade relations to construct inter-regional dependencies is subject to very large uncertainties (Beaufils & Wenz 2021) and doing so in order to assess spillovers would therefore add large additional uncertainty which would cloud the main insights of the present analysis which focuses primarily on the implications of impacts from different climate conditions, and the persistence of their impacts on growth. Third, assessing spill-overs while also adequately addressing the question of temporal persistence requires including both temporal-, spatial-, and perhaps cross spatio-temporal lags which would require extensive model testing.

We therefore leave this for future work which can address this important topic more explicitly using either dynamic models of spillovers via trade, or econometric methods to assess their effects. In the updated manuscript, we provide additional discussion on these issues on L356-375, making clear the types of spillovers which our empirical framework likely already captures, those which it does not, and how accounting for these might alter the overall magnitude of projected damages in future work.

The other way the authors could allay my fears would be to justify how impacts of the size they estimate are consistent with the growth experience we actually observe across countries and time. In some sense, their approach guarantees this of course, because they estimate the economic impacts of climate shocks using historical data. But, due to the challenges of proving their model is not mis-specified, this is not enough. If rising temperatures and associated rainfall changes impact GDP so much, and also differently across countries, should we not be seeing this manifest itself in terms of regional patterns of convergence or non-convergence? Granted, future climate change is much larger than past change, but the impacts over the next couple of decades are already projected to be large, so I would like the authors to walk me through why their estimates are not inconsistent with observed *long-term* growth patterns across countries.

We appreciate the reviewer's interest in contextualizing the magnitude of climate damages which we project in light of long-term historical economic development and climate change. While this is definitely interesting and worth exploring, we note that evaluating the overall role of historical climate change on economic development would require a counterfactual measurement of historical economic development in the absence of climate change, which simply does not exist. While we can compare long-term patterns of growth across countries, these differences are subject to considerable unobserved biases which are independent of rates of historical climate change, and therefore likely obscure any comparison which can be considered causal-evidence. It is precisely this reason that the climate-econometric literature has focussed on fixed-effects panel regressions which avoid these unobserved biases (Auffhammer 2018). Estimating counterfactual trajectories of historical economic development without climate change has been undertaken in the literature (and is becoming increasingly important, e.g. in the context of topical loss-and-damage debates), but essentially uses empirical estimates derived from panel fixed-effects models such as ours in combination with physical simulations of historical climate change to calculate them (see e.g. Diffenbaugh 2019 and Callahan et al 2021).

Nevertheless, the following "back-of-the-envelope" calculations should put the magnitude of damages which we project into perspective considering the long-term historical patterns of economic growth. We have experienced approximately 1C of global warming historically since 1970 (https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature), and CMIP6 climate models project approximately another 1C of global warming by 2050 (compared to 2020) under SSP585 (IPCC AR6 WG1, Fig. 4.2). This makes for a simple, approximate comparison of the future damages we project and those which we should have experienced historically since 1970, allowing a contextualisation against the background of historical economic development. We calculate an approximate 20% reduction in global GDP from the additional 1C of global warming projected under SSP585 (Fig. 1), with differences between the upper and lower quartile of the income distribution of approximately 10%-points (Fig. S17), meaning a maximal impact of 30% reduction in developing countries compared to 10% reduction in more wealthy countries. Let's assume that the historical 1C of global warming produced damages of similar magnitudes, although in reality they were likely smaller due to the non-linear response to average temperature which is more negative as regions warm (Extended Data Figure 1). We can then compare the magnitude of these damages to the background economic development which occurred between 1970 and 2020. Average growth rates of GDP per capita were approximately 1.8% over the past 50 years (https://ourworldindata.org/grapher/gdp-per-capita-growth?time=1995), implying an average growth in GDP per capita of over 140% since 1970. Taking the bottom quartile of countries by World Bank income per capita (using 2015 values) gives average growth rates of 0.84% annually over the past 50 years, whereas the upper quartile of countries gives average growth rates of 1.41% annually (note that this is consistent with evidence that absolute income convergence has not occurred historically, see Pritchett 1997, Gilles et al 2009, Kremer et al 2022). These imply overall income per capita growth of 52% and 101% in the lower and upper income quartiles respectively over the past 50 years (noting that the greatest income growth has occurred for countries in the middle quartiles).

Even given the approximate nature of these calculations, it becomes quite clear that while considerable, the implied damages of historical climate change (20%) are unlikely to have had consequences which are inconsistent with historical economic development (an increase in income per capita of 140%) or obviously noticeable without an appropriate no-climate-change counterfactual to which to compare. Moreover, we note that poorer regions have actually seen lower growth rates than richer regions historically. Our estimates indicate that climate change may have played a role in this, and that the gap between them would have been smaller (approx. 52+30=82% vs 101+10=111%) without climate change. We note again that the observation of lower growth rates in poor versus rich countries can in no way be interpreted as causal evidence of historical climate damages because of the large unobserved biases which influence differences across countries which are unrelated to climate. There is no counterfactual world without climate change from which we can measure whether poorer and richer countries *are actually* 30% and 10% worse off than they would have been without climate change. Therefore, we emphasize that we must rely on the empirical approach which we take to identify impacts with fixed-effects panel regressions which are plausibly causal.

Nevertheless, these “back-of-the-envelope” calculations demonstrate that the magnitude of damages which we project is consistent with historical developments, given that: a) historical economic development is much larger than the historical damages implied by our analysis, b) richer regions grew historically at faster rates than poorer regions, consistent with the pattern of climate damages we show, and in which historical climate change therefore potentially played a contributing role.

In the updated manuscript we have included discussion of these approximate calculations in a new Supplementary Discussion Section S5, which is referred to on L. 339-342 of the updated manuscript.

My other comment is that I am struggling to understand the sense in the cost/benefit comparison exercise. I may have got this wrong, but I think this compares the total cost of climate change relative to a no-climate-change baseline with the cost of mitigating climate change to bring it from a baseline/bau path to a lower (but not zero) emissions path. If so, then I find this an illegitimate comparison and liable to cause confusion. The only legitimate comparison is between the costs and benefits of moving from path A (bau) to path B (Paris), but that is not what is done here. Moreover, the basic result about the economic commitment of climate change (not the most appealing terminology by the way, but fine) seems to imply that there would no statistically significant economic benefit of reducing emissions until mid-century, because two emissions scenarios that proxy for bau and Paris respectively give strongly overlapping confidence intervals of damages in the period up to 2050. To be clear, I am not for one minute questioning the basic claim that mitigating climate change has benefits greater than costs, but the comparison in this paper is not the right way to demonstrate it. It needs fundamental rethinking, or just to be removed.

We thank the referee for their comments on the comparison of climate damages to mitigation costs, which we acknowledge is by no means a typical cost/benefit analysis. The referee is correct that we compare the damages caused by physical climate change (different from “the total cost of climate change” if the referee implies this to also include the costs of mitigation efforts) against the cost of mitigating climate change to bring emissions from a BAU to a “Paris-compatible” path (RCP2.6). We agree that for a formal cost-benefit analysis of different emission paths, the only legitimate comparison is between the overall costs involved in moving from path A (damages of BAU) to path B (damages of Paris + mitigation costs of Paris). However, we do not intend to conduct a formal cost-benefit analysis of different emission paths, but rather to very simply place the magnitude of projected damages from physical climate change in the context of the magnitude of estimates of mitigation costs. This is motivated by the fact that their relative magnitudes is a relevant factor in public perceptions of climate change (see L25-29 of the introduction), making this valuable information to convey separately from formal cost-benefit analyses which compare the total costs and benefits of moving from one emission path to another.

We also think that this comparison is of interest in this context given that we find physical damages from climate change to be statistically indistinguishable across different emission paths until mid-century. Formal cost-benefit analyses of different emission paths typically only find net benefits of mitigation to occur in the latter half of the century (e.g. Drouet 2021). Casual interpretation of these results might lead some to conclude that mitigation costs are just larger than damages until the latter half of the century, but our analysis clarifies that this is actually because physical climate trajectories and damage estimates will be indistinguishable across different emission paths until mid-century, and damage estimates are actually already much larger than mitigation costs before mid-century.

For these reasons, we think this is an important and novel implication of our analysis to convey. We have added additional text in the results (L186-188) and discussion (L379-388) to clarify our motivation for making this comparison, its distinction from formal cost-benefit analyses, and its implications. However, given that this is an implication of our analysis rather than an integral part of it, we are open to moving statements made in the abstract to the main text and the comparison to mitigation costs from Fig. 1 to the Supplementary Material if the referee continues to strongly disagree with the validity and relevance of this approach.

References

Kikstra, Jarmo S., et al. "The social cost of carbon dioxide under climate-economy feedbacks and temperature variability." *Environmental Research Letters* 16.9 (2021): 094037.

Kotz, Maximilian, et al. "Day-to-day temperature variability reduces economic growth." *Nature Climate Change* 11.4 (2021): 319-325.

Kotz, Maximilian, Anders Levermann, and Leonie Wenz. "The effect of rainfall changes on economic production." *Nature* 601.7892 (2022): 223-227.

Acharya, Viral et al. "Do Firms Mitigate Climate Impact on Employment? Evidence from US Heat Shocks." *NBER Working papers.* (2023). Available at: http://www.nber.org/papers/w31967

Deryugina, Tatyana, and Solomon M. Hsiang. *Does the environment still matter? Daily temperature and income in the United States*. No. w20750. National Bureau of Economic Research, 2014.

Schleypen, Jessie Ruth, et al. "Sharing the burden: quantifying climate change spillovers in the European Union under the Paris Agreement." *Spatial Economic Analysis* 17.1 (2022): 67-82.
Dasgupta, Shouro, et al. "Global temperature effects on economic activity and equity: a spatial analysis." *Italy: European Institute on Economics and the Environment* (2022).

Neal, Timothy, The Importance of External Weather Effects in Projecting the Economic Impacts of Climate Change (June 7, 2023). *UNSW Economics Working Paper 2023-09*, Available at SSRN: https://ssrn.com/abstract=4471379

Malik, Arunima, et al. "Impacts of climate change and extreme weather on food supply chains cascade across sectors and regions in Australia." *Nature Food* 3.8 (2022): 631-643.

Middelanis, Robin, et al. "Wave-like global economic ripple response to Hurricane Sandy." *Environmental Research Letters* 16.12 (2021): 124049.

Kuhla, Kilian, et al. "Ripple resonance amplifies economic welfare loss from weather extremes." *Environmental Research Letters* 16.11 (2021): 114010.

Beaufils, Timothé, and Leonie Wenz. "A scenario-based method for projecting multi-regional input–output tables." *Economic Systems Research* 34.4 (2022): 440-468.

Diffenbaugh, Noah S., and Marshall Burke. "Global warming has increased global economic inequality." *Proceedings of the National Academy of Sciences* 116.20 (2019): 9808-9813.

Callahan, Christopher W., and Justin S. Mankin. "Globally unequal effect of extreme heat on economic growth." *Science Advances* 8.43 (2022): eadd3726.

Pritchett, Lant. "Divergence, big time." *Journal of Economic perspectives* 11.3 (1997): 3-17.

Dufrénot, Gilles, Valérie Mignon, and Théo Naccache. *The slow convergence of per capita income between the developing countries:" growth resistance" and sometimes" growth tragedy"*. No. 09/03. CREDIT Research Paper, 2009.

Kremer, Michael, Jack Willis, and Yang You. "Converging to convergence." *NBER macroeconomics annual* 36.1 (2022): 337-412.

Drouet, Laurent, Valentina Bosetti, and Massimo Tavoni. "Net economic benefits of well-below 2° C scenarios and associated uncertainties." *Oxford Open Climate Change* 2.1 (2022): kgac003.

Reviewer Reports on the Fourth Revision:

Referees' comments:

Referee #4 (Remarks to the Author):

The authors have provided a thorough response to my comments, including two new pieces of analysis to respond to concerns I had, one on spatial spillovers and the other (more informally) on whether the paper's estimates are consistent with countries' growth experience over recent decades. I found both analyses reassuring.

The authors have also responded to my complaint about the cost/benefit comparison exercise by essentially proposing a compromise, adding some additional explanation of what the exercise is intended to do, and not do. I think this compromise is reasonable, but I think the authors could do a little more to make the point clearly. It is very well explained in the discussion, but not so clearly explained where the results are first presented. Thus, the authors should consider promoting some or all of the new discussion text to the section on "Committed damages...", or else consider rewording and expanding on the sentence on lines 195-6, which I find unclear. To my mind, the key points are (a) this is not a cost/benefit comparison, but (b) it is useful because it shows us the reason net benefits appear later in the 21st century is not that damages are small relative to costs earlier in the century, rather it is because damages are large but emissions reductions don't avoid many damages until later on.

Otherwise, I am happy to recommend this get published, and I don't need to see it again.

Author Rebuttals to Fourth Revision:
Response to Referees' comments:

Referee #4 (Remarks to the Author):

The authors have provided a thorough response to my comments, including two new pieces of analysis to respond to concerns I had, one on spatial spillovers and the other (more informally) on whether the paper's estimates are consistent with countries' growth experience over recent decades. I found both analyses reassuring.

The authors have also responded to my complaint about the cost/benefit comparison exercise by essentially proposing a compromise, adding some additional explanation of what the exercise is intended to do, and not do. I think this compromise is reasonable, but I think the authors could do a little more to make the point clearly. It is very well explained in the discussion, but not so clearly explained where the results are first presented. Thus, the authors should consider promoting some or all of the new discussion text to the section on "Committed damages...", or else consider rewording and expanding on the sentence on lines 195-6, which I find unclear. To my mind, the key points are (a) this is not a cost/benefit comparison, but (b) it is useful because it shows us the reason net benefits appear later in the 21st century is not that damages are small relative to costs earlier in the century, rather it is because damages are large but emissions reductions don't avoid many damages until later on.

Otherwise, I am happy to recommend this get published, and I don't need to see it again.

We are glad that our additional analysis and discussion satisfied the concerns of the referee. To address their final comment, we have taken their suggestion to incorporate text from the discussion into the earlier part of the results entitled "Damages already outweigh mitigation costs", shortening and adjusting the text in the discussion section accordingly. In doing so, we have focussed on the two key points which the referee highlights, which we agree are of particular importance.

These amendments can be found on L. 171-179 and L. 342-345 of the updated manuscript, copied below:

L171-179
"This comparison aims simply to compare the magnitude of future damages against mitigation costs, rather than to conduct a formal cost-benefit analysis of transitioning from one emission path to another. Formal cost-benefit analyses typically find that the net benefits of mitigation only emerge after 2050[5], which may lead some to conclude that physical damages from climate change are simply not large enough to outweigh mitigation costs until the latter half of the century. Our simple comparison of their magnitudes makes clear that damages are actually already considerably larger than mitigation costs, and the delayed emergence of net mitigation benefits results primarily from the fact that damages across different emission paths are indistinguishable until mid-century (Figure 1)."

L342-345

“Our simple comparison of the magnitude of damages and mitigation costs makes clear that this is primarily because damages are indistinguishable across emissions scenarios – i.e. committed - until mid-century (Figure 1), and that they are actually already much larger than mitigation costs.”