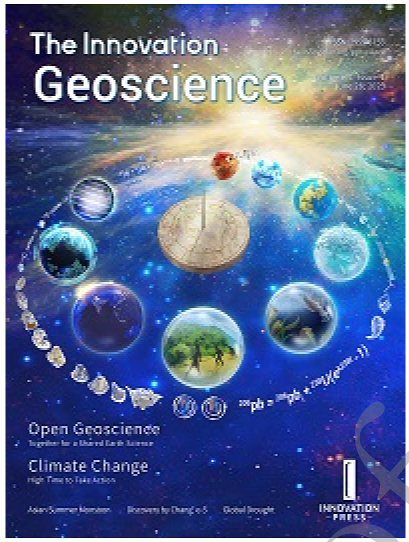


Beyond the multimodel ensemble mean (MME): assessing plausible climate risk at a given warming level

Journal:	*The Innovation Geoscience*
Manuscript ID	TIG-2026-0094.R2
Manuscript Type:	Commentary
Keywords:	Climate risk assessment, Extreme climate outcomes, Multimodel ensemble mean, Internal climate variability, Constrained climate projection

Beyond the multimodel ensemble mean (MME): Assessing plausible climate risk at a given warming level

Hainan Gong[1,2], Wentao Xia[1,5], Lin Wang[1,2*], Kangjie Ma[3], Wen Chen[4]

[1]Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[2]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC - FEMD), Nanjing University of Information Science & Technology, Nanjing 210044, China.

[3]College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China.

[4]Yunnan International Joint Laboratory of Monsoon and Extreme Climate Disasters /Department of Atmospheric Sciences, Yunnan University, Kunming 650500, China

[5]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

*Correspondence: wanglin@mail.iap.ac.cn (L.W.)

Subject Area: Atmospheric Sciences

The 2°C threshold enshrined in the Paris Agreement is widely viewed as a guardrail against the most dangerous consequences of climate change. Yet the risks associated with any given level of warming are still commonly framed through the multimodel ensemble mean (MME), which mainly summarizes the externally forced response shared across climate models. That benchmark is useful, but it is not sufficient.

The climate that societies ultimately confront will be shaped not only by external forcing, but also by differences in modelled process responses and by internal climate variability. Consequently, some of the most consequential future impacts may arise not from the ensemble average, but from plausible high-impact outcomes located in the tails of the distribution.

THE ILLUSION OF MODERATE RISK

A recent study by Bevacqua and colleagues brings this point into sharp relief.[1] Drawing on CMIP6 simulations, they show that even in a 2°C warmer world, some plausible climate outcomes can produce sectoral hazards that are more severe than the MME response at 3°C or even 4°C warming. Their analysis challenges a familiar assumption in climate communication: that the most extreme climate outcomes are chiefly a concern under very high warming. In reality, moderate warming does not necessarily translate into moderate risk. The significance of the study lies not merely in emphasizing model spread, but in demonstrating how that spread can be interpreted in a sector-relevant manner. Rather than focusing only on local extremes or broad global averages, the authors define sector-specific "global climatic impact-drivers" that connect each hazard to the regions where it matters most. In their framework, extreme five-day precipitation over densely populated areas serves as a proxy for flood risk, drought frequency across major breadbasket regions represents food-security risk, and fire weather over forests reflects threats to ecosystems and wildfire exposure (Figure 1). Models are then ranked according to these sector-relevant impact drivers, making it possible to identify plausible best- and worst-case outcomes at the same warming level.

These results have important implications. For drought across global breadbaskets, the spread at 2°C is so large that several model outcomes exceed the MME response at 4°C. For extreme precipitation over populated regions and fire weather over forests, the worst-case outcomes at 2°C surpass the MME response at 3°C. Put differently, the upper tail of risk in a 2°C world can overlap with, or even exceed, the average risk expected in a substantially warmer world. For climate-sensitive sectors, this is not a minor statistical detail. It is a direct challenge to the way risk is assessed and adaptation is planned.

WHY THE MME IS NOT THE FUTURE CLIMATE

The MME remains a powerful benchmark for identifying robust forced responses and for reducing the influence of idiosyncratic biases in individual models. However, it should not be interpreted as the full risk space of future climate. It is a statistical summary of model responses, not the climate outcome that societies will necessarily experience. Future climate hazards relevant to risk emerges from the combined influence of external forcing, differences in model responses, and internal variability,[2,3] all of which can shape regional hazards. This distinction is especially important for extreme events. Many societally disruptive extremes are not simply linear extensions of the forced mean state. Instead, they often depend on specific circulation regimes and oceanic background conditions. These ingredients are sensitive to both model formulation and internal variability, which can amplify, mute, or reorganize regional hazards across interannual to decadal timescales.[3] A projection framework centered only on MME change may therefore underestimate the likelihood, intensity, or spatial coherence of future high-impact extremes.

At the same time, the study offers an important clarification regarding the sources of uncertainty. For the sectors emphasized by Bevacqua and colleagues, much of the spread at 2°C appears to stem from model differences rather than from internal variability alone. That point matters because it underscores the central importance of model performance and process representation for improving future risk assessments. Yet internal variability cannot be dismissed. It still contributes to the envelope of plausible outcomes and becomes particularly relevant when considering realized regional climate trajectories over the coming decades. The relative importance of these uncertainty sources also depends strongly on timescale. Internal variability is especially relevant for near- to medium-term projections (e.g., 2030–2050), because it can modulate the timing, regional expression, and realized severity of climate hazards. By contrast, over longer timescales, model structural uncertainty, including differences in climate sensitivity and process representation, generally becomes a larger contributor to the spread of projected sectoral hazards. This distinction reinforces the need to consider both sources of uncertainty, while recognizing that their relative roles vary with lead time, region, and impact metric. A balanced reading of the study is therefore that both model performance and internal variability matter, even though their relative importance is likely to vary across sectors, timescales, and impact metrics.

CONSTRAINING THE FUTURE THAT MATTER

Recognizing these multiple sources of uncertainty has an important implication: climate-risk assessment should not stop at describing average change, but should also examine the plausibility and relevance of outcomes away from the MME. The contribution of Bevacqua and colleagues is precisely to show that such outcomes can be highly consequential even under moderate warming. This perspective points to several priorities for future research. First, extreme outcomes identified within climate ensembles should be subjected to careful process-based evaluation. Ensemble tails should not be equated automatically with physically plausible futures. Some high-impact outcomes may reflect credible process responses, whereas others may arise from known model deficiencies or unrealistic regional teleconnections. Assessing plausibility therefore requires process-based evaluation, including whether the associated circulation changes, moisture transport, land–atmosphere feedbacks, or fire-weather responses are dynamically and physically consistent with observations and established mechanisms.[4]

Second, improving model performance remains fundamental. Differences in climate sensitivity and in the treatment of clouds, convection, land–atmosphere coupling, and hydrological processes can strongly affect projections of drought, heavy rainfall, and wildfire-conducive weather. Better model fidelity, together with physically grounded constraints, should therefore help refine the range of plausible future hazards.[4] Third, internal variability deserves more explicit attention rather than being treated merely as residual noise. Even when it is not the leading source of spread for a given sector, it can still modulate the timing, regional expression, and realized severity of future extremes.[2,3] This is particularly relevant for near-term regional planning, where decadal circulation and ocean–atmosphere variability may amplify, delay, or reorganize the forced response. For longer-term assessments, however, model structural uncertainty often becomes more important, because differences in climate sensitivity, circulation responses, convection, land–atmosphere coupling, and hydrological processes shape the spread of sectoral hazards such as drought, heavy precipitation, and fire-weather changes. Thus, adaptation planning should combine near-term regional storylines of internal variability with longer-term process-based evaluation of model credibility and high-impact ensemble tails. Although predictability remains limited and depends on the phenomenon considered, better characterization of internal variability would provide a more realistic picture of near- to medium-term risk. The goal is not to replace the MME, but to place it in a broader risk framework that combines the average forced response with physically grounded scrutiny of high-impact tails, improved model evaluation, and clearer characterization of internal variability. Such an approach would be particularly valuable in climate-risk hotspots, including densely populated urban areas, major crop-producing regions, ecologically fragile systems, and critical infrastructure corridors, where underestimating plausible extremes can carry outsized societal costs.[5]

IMPLICATIONS FOR RISK AND POLICY

This perspective also carries a clear message for climate policy. Warming thresholds such as 1.5°C and 2°C remain indispensable benchmarks, but they should not be interpreted as guarantees of proportionally moderate impacts. A 2°C world may still contain plausible futures in which drought, flood, and fire risks are comparable to, or even exceed, the average impacts associated with much warmer worlds. This message should not be interpreted as weakening the case for mitigation. Although some high-impact outcomes at 2°C may exceed the ensemble-mean response at higher warming levels, further warming would still shift the overall distribution of hazards toward more severe outcomes and increase the likelihood of crossing critical impact thresholds. Attention to high-impact outcomes at moderate warming therefore reinforces the need for both rapid mitigation and robust adaptation[5]. Adaptation planning should therefore be informed not only by likely futures summarized by the MME, but also by credible high-impact tails within the same warming level. The broader lesson is straightforward. Hazard-based indicators of sectoral risk are not defined solely by where the MME lies. It is also shaped by what remains physically plausible within the ensemble spread, and by whether that spread can be better understood through process evaluation, improved model performance, and clearer characterization of internal variability. Bevacqua and colleagues offer an important warning against equating moderate warming with moderate risk. The next step is to translate that warning into climate assessments that move beyond average projections and are better able to inform decisions about the plausible extremes that matter most.

REFERENCES

1. Bevacqua E., Fischer E., Sillmann J., et al. (2026). Moderate global warming does not rule out extreme global climate outcomes. *Nature* **651**: 946–953. DOI: 10.1038/s41586-026-10237-9

2. Deser C., Lehner F., Rodgers K. B., et al. (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**: 277-

286. DOI: 10.1038/s41558-020-0731-2

3. Gong H., Wang L., Screen J. et al. (2025). Teleconnection from Arctic warming suppresses long-term warming in central Eurasia. *Sci. Adv.* **11**: eadq9461. DOI: 10.1126/sciadv.adq9461

4. Williamson M. S., Thackeray C. W., Cox P. M.et al. (2021). Emergent constraints on climate sensitivities. *Rev. Mod. Phys.* **93**: 025004. DOI: 10.1103/RevModPhys.93.025004

5. Intergovernmental Panel on Climate Change (IPCC). (2023). Key risks across sectors and regions. *Climate Change 2022 – Impacts, adaptation and vulnerability: Working group II contribution to the sixth assessment report of the intergovernmental panel on climate change* pp:2411–2538. Cambridge University Press. DOI: https://doi.org/10.1017/9781009325844.025

FUNDING AND ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (42522508, 42275023), and the Chinese Academy of Sciences Project for Young Scientists in Basic Research (YSBR-137). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Hainan Gong: Writing-review & editing, Writing–original draft, Methodology, Funding, Investigation. Wentao Xia: Visualization, Software, Investigation. Lin Wang: Validation, Conceptualization, Writing-review & editing. Kangjie Ma: Visualization, Software, Investigation. Wen Chen: Validation, Conceptualization. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

Figure caption

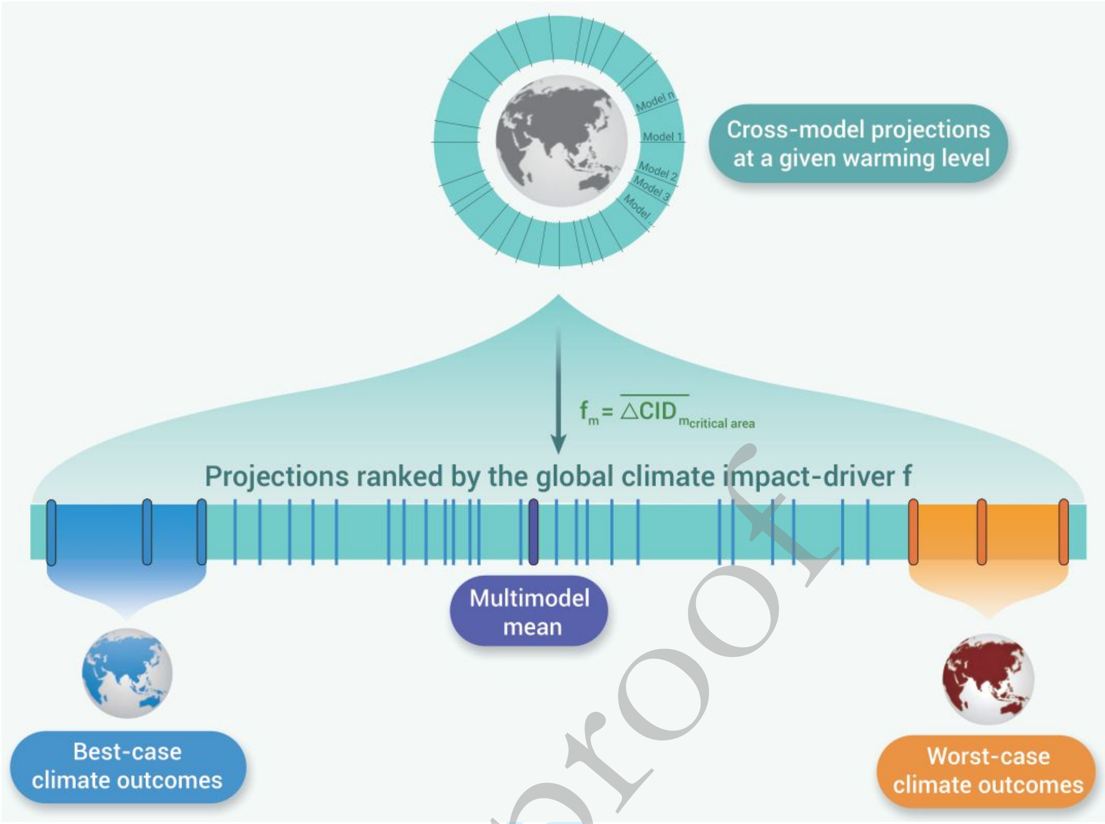
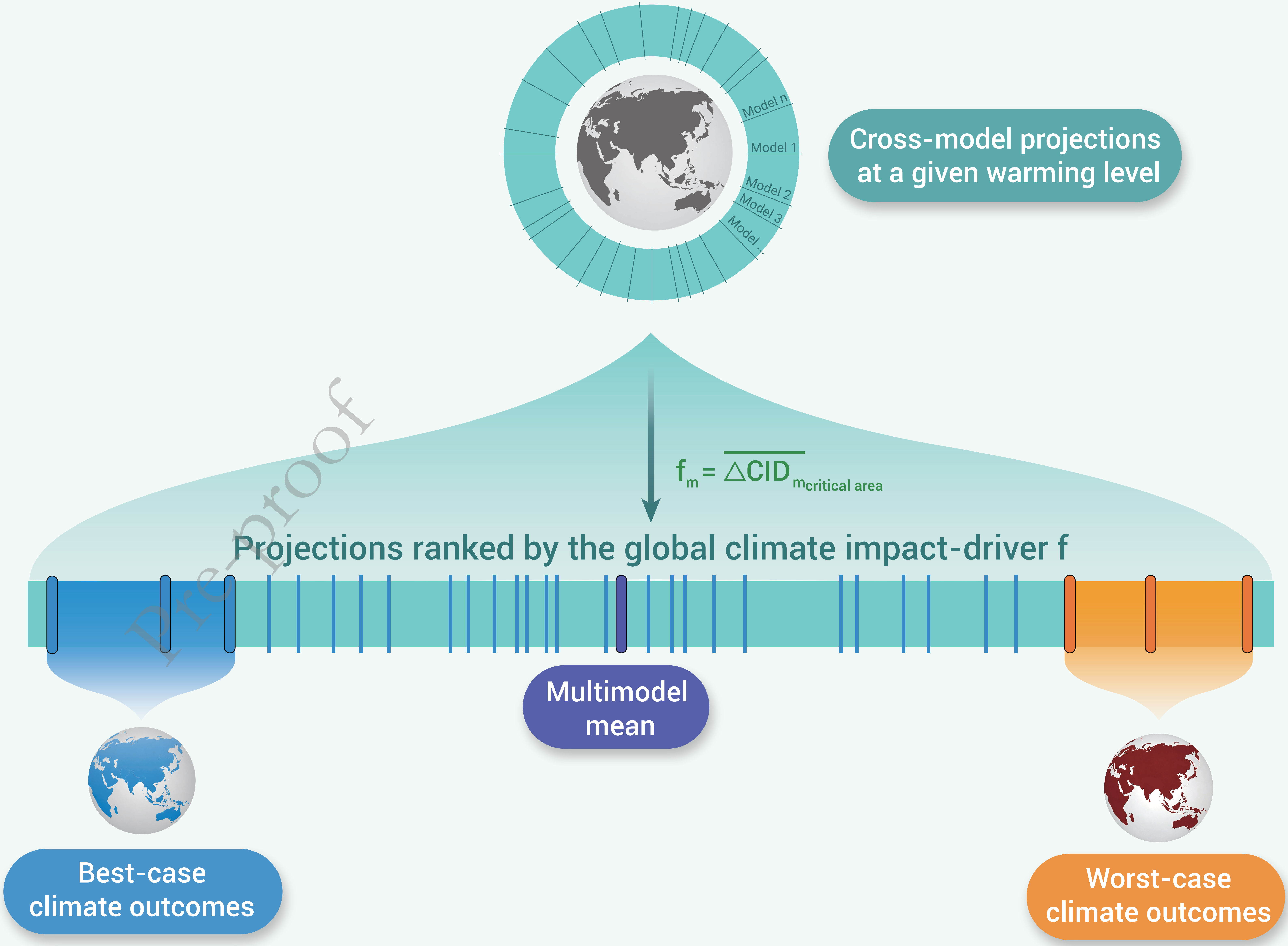


Figure. 1 Framework for identifying plausible extreme climate outcomes from multi-model projections At a given warming level (2°C in *Bevacqua et al.*)[1], model projections are ranked according to a sector-specific global climatic impact-driver, calculated as the spatially averaged change in a relevant hazard across the critical region for that sector. This procedure enables the identification of plausible high- and low-impact futures within the model ensemble. For food security, for example, the metric may be based on projected drought changes across major breadbasket regions. Adapted from Fig. 1 in *Bevacqua et al.*[1] under the terms of the CC BY 4.0 license.


Model n
Model 1
Model 2
Model 3
Model ...
Cross-model projections at a given warming level
$f_m = \overline{\triangle CID}_{m_{\text{critical area}}}$
Projections ranked by the global climate impact-driver f
Multimodel mean
Best-case climate outcomes
Worst-case climate outcomes

Dear Editor,

Thank you for your careful evaluation of our manuscript entitled "Beyond the multimodel ensemble mean (MME): assessing plausible climate risk at a given warming level" (Manuscript ID: TIG-2026-0094).

As no further comments were raised regarding the scientific content of the manuscript, no substantive changes to the text or analyses were required. Following the journal's instructions, we have carefully revised the manuscript and associated files to ensure that they comply with the required formatting and final-file guidelines. We have also completed the Article Processing Charge payment as requested.

We sincerely appreciate the time and efforts of the editors and reviewers in handling our manuscript. We hope that the revised submission now meets all the journal's requirements and look forward to the final evaluation.

Sincerely,

Hainan Gong
On behalf of all authors

Beyond the multimodel ensemble mean (MME): Assessing plausible climate risk at a given warming level

Hainan Gong[1,2], Wentao Xia[1,5], Lin Wang[1,2*], Kangjie Ma[3], Wen Chen[4]

[1]Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[2]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC‑FEMD), Nanjing University of Information Science & Technology, Nanjing 210044, China.

[3]College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China.

[4]Yunnan International Joint Laboratory of Monsoon and Extreme Climate Disasters /Department of Atmospheric Sciences, Yunnan University, Kunming 650500, China

[5]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

*Corresponding author: Prof. Lin Wang (wanglin@mail.iap.ac.cn)

The 2°C threshold enshrined in the Paris Agreement is widely viewed as a guardrail against the most dangerous consequences of climate change. Yet the risks associated with any given level of warming are still commonly framed through the multimodel ensemble mean (MME), which mainly summarizes the externally forced response shared across climate models. That benchmark is useful, but it is not sufficient. The climate that societies ultimately confront will be shaped not only by external forcing, but also by differences in modelled process responses and by internal climate variability. Consequently, some of the most consequential future impacts may arise not from the ensemble average, but from plausible high-impact outcomes located in the tails of the distribution.

THE ILLUSION OF MODERATE RISK

A recent study by Bevacqua and colleagues brings this point into sharp relief.[1] Drawing on CMIP6 simulations, they show that even in a 2°C warmer world, some plausible climate outcomes can produce sectoral hazards that are more severe than the MME response at 3°C or even 4°C warming. Their analysis challenges a familiar assumption in climate communication: that the most extreme climate outcomes are chiefly a concern under very high warming. In reality, moderate warming does not necessarily translate into moderate risk. The significance of the study lies not merely in emphasizing model spread, but in demonstrating how that spread can be interpreted in a sector-relevant manner. Rather than focusing only on local extremes or broad global averages, the authors define sector-specific "global climatic impact-drivers" that connect each hazard to the regions where it matters most. In their framework, extreme five-day precipitation over densely populated areas serves as a proxy for flood risk, drought frequency across major breadbasket regions represents food-security risk, and fire weather over forests reflects threats to ecosystems and wildfire exposure (Figure 1). Models are then ranked according to these sector-relevant impact drivers, making it possible to identify plausible best- and worst-case outcomes at the same warming level.

These results have important implications. For drought across global breadbaskets, the spread at 2°C is so large that several model outcomes exceed the MME response at 4°C. For extreme precipitation over populated regions and fire weather over forests, the worst-case outcomes at 2°C surpass the MME response at 3°C. Put differently, the upper tail of risk in a 2°C world can overlap with, or even exceed, the average risk expected in a substantially warmer world. For climate-sensitive sectors, this is not a minor statistical detail. It is a direct challenge to the way risk is assessed and adaptation is planned.

WHY THE MME IS NOT THE FUTURE CLIMATE

The MME remains a powerful benchmark for identifying robust forced responses and for reducing the influence of idiosyncratic biases in individual models. However, it should not be interpreted as the full risk space of future climate. It is a statistical summary of model responses, not the climate outcome that societies will necessarily experience. Future climate hazards relevant to risk emerges from the combined influence of external forcing, differences in model responses, and internal variability,[2,3] all of which can shape regional hazards. This distinction is especially important for extreme events. Many societally disruptive extremes are not simply linear extensions of the forced mean state. Instead, they often depend on specific circulation regimes and oceanic background conditions. These ingredients are sensitive to both model formulation and internal variability, which can amplify, mute, or reorganize regional hazards across interannual to decadal timescales.[3] A projection framework centered only on MME change may therefore underestimate the likelihood, intensity, or spatial coherence of future high-impact extremes.

At the same time, the study offers an important clarification regarding the sources of uncertainty. For the sectors emphasized by Bevacqua and colleagues, much of the spread at 2°C appears to stem from model differences rather than from internal variability alone. That point matters because it underscores the central importance of model performance and process representation for improving future risk assessments. Yet internal variability cannot be dismissed. It still contributes to the envelope of plausible outcomes and becomes particularly relevant when considering realized regional climate trajectories over the coming decades. The relative importance of these uncertainty sources also depends strongly on timescale. Internal variability is especially relevant for near- to medium-term projections (e.g., 2030–2050), because it can modulate the timing, regional expression, and realized severity of climate hazards. By contrast, over longer timescales, model structural uncertainty, including differences in climate sensitivity and process representation, generally becomes a larger contributor to the spread of projected sectoral hazards. This distinction reinforces the need to consider both sources of uncertainty, while recognizing that their relative roles vary with lead time, region, and impact metric. A balanced reading of the study is therefore that both model performance and internal variability matter, even though their relative importance is likely to vary across sectors, timescales, and impact metrics.

CONSTRAINING THE FUTURE THAT MATTER

Recognizing these multiple sources of uncertainty has an important implication: climate-risk assessment should not stop at describing average change, but should also examine the plausibility and relevance of outcomes away from the MME. The contribution of Bevacqua and colleagues is precisely to show that such outcomes can be highly consequential even under moderate warming. This perspective points to several priorities for future research. First, extreme outcomes identified within climate ensembles should be subjected to careful process-based evaluation. Ensemble tails should not be equated automatically with physically plausible futures. Some high-impact outcomes may reflect credible process responses, whereas others may arise from known model deficiencies or unrealistic regional teleconnections. Assessing plausibility therefore requires process-based evaluation, including whether the associated circulation changes, moisture transport, land–atmosphere feedbacks, or fire-weather responses are dynamically and physically consistent with observations and established mechanisms.[4]

Second, improving model performance remains fundamental. Differences in climate sensitivity and in the treatment of clouds, convection, land–atmosphere coupling, and hydrological processes can strongly affect projections of drought, heavy rainfall, and wildfire-conducive weather. Better model fidelity, together with physically grounded constraints, should therefore help refine the range of plausible future hazards.[4] Third, internal variability deserves more explicit attention rather than being treated merely as residual noise. Even when it is not the leading source of spread for a given sector, it can still modulate the timing, regional expression, and realized severity of future extremes.[2,3] This is particularly relevant for near-term regional planning, where decadal circulation and ocean–atmosphere variability may amplify, delay, or reorganize the forced response. For longer-term assessments, however, model structural uncertainty often becomes more important, because differences in climate sensitivity, circulation responses, convection, land–atmosphere coupling, and hydrological processes shape the spread of sectoral hazards such as drought, heavy precipitation, and fire-weather changes. Thus, adaptation planning should combine near-term regional storylines of internal variability with longer-term process-based evaluation of model credibility and high-impact ensemble tails. Although predictability remains limited and depends on the phenomenon considered, better characterization of internal variability would provide a more realistic picture of near- to medium-term risk. The goal is not to replace the MME, but to place it in a broader risk framework that combines the average forced response with physically grounded scrutiny of high-impact tails, improved model evaluation, and clearer characterization of internal variability. Such an approach would be particularly valuable in climate-risk hotspots, including densely populated urban areas, major crop-producing regions, ecologically fragile systems, and critical infrastructure corridors, where underestimating plausible extremes can carry outsized societal costs.[5]

IMPLICATIONS FOR RISK AND POLICY

This perspective also carries a clear message for climate policy. Warming thresholds such as 1.5°C and 2°C remain indispensable benchmarks, but they should not be interpreted as guarantees of proportionally moderate impacts. A 2°C world may still contain plausible futures in which drought, flood, and fire risks are comparable to, or even exceed, the average impacts associated with much warmer worlds. This message should not be interpreted as weakening the case for mitigation. Although some high-impact outcomes at 2°C may exceed the ensemble-mean response at higher warming levels, further warming would still shift the overall distribution of hazards toward more severe outcomes and increase the likelihood of crossing critical impact thresholds. Attention to high-impact outcomes at moderate warming therefore reinforces the need for both rapid mitigation and robust adaptation[5]. Adaptation planning should therefore be informed not only by likely futures summarized by the MME, but also by credible high-impact tails within the same warming level. The broader lesson is straightforward. Hazard-based indicators of sectoral risk are not defined solely by where the MME lies. It is also shaped by what remains physically plausible within the ensemble spread, and by whether that spread can be better understood through process evaluation, improved model performance, and clearer characterization of internal variability. Bevacqua and colleagues offer an important warning against equating moderate warming with moderate risk. The next step is to translate that warning into climate assessments that move beyond average projections and are better able to inform decisions about the plausible extremes that matter most.

REFERENCES

1. Bevacqua E., Fischer E., Sillmann J., et al. (2026). Moderate global warming does not rule out extreme global climate outcomes. *Nature* **651**: 946–953. DOI: 10.1038/s41586-026-10237-9
2. Deser C., Lehner F., Rodgers K. B., et al. (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**: 277-286. DOI: 10.1038/s41558-020-0731-2

3. Gong H., Wang L., Screen J. et al. (2025). Teleconnection from Arctic warming suppresses long-term warming in central Eurasia. *Sci. Adv.* **11**: eadq9461. DOI: 10.1126/sciadv.adq9461

4. Williamson M. S., Thackeray C. W., Cox P. M.et al. (2021). Emergent constraints on climate sensitivities. *Rev. Mod. Phys.* **93**: 025004. DOI: 10.1103/RevModPhys.93.025004

5. Intergovernmental Panel on Climate Change (IPCC). (2023). Key risks across sectors and regions. *Climate Change 2022 – Impacts, adaptation and vulnerability: Working group II contribution to the sixth assessment report of the intergovernmental panel on climate change* pp:2411–2538. Cambridge University Press. DOI: https://doi.org/10.1017/9781009325844.025

FUNDING AND ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (42522508, 42275023), and the Chinese Academy of Sciences Project for Young Scientists in Basic Research (YSBR-137). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Hainan Gong: Writing-review & editing, Writing–original draft, Methodology, Funding, Investigation. Wentao Xia: Visualization, Software, Investigation. Lin Wang: Validation, Conceptualization, Writing-review & editing. Kangjie Ma: Visualization, Software, Investigation. Wen Chen: Validation, Conceptualization. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

Figure caption

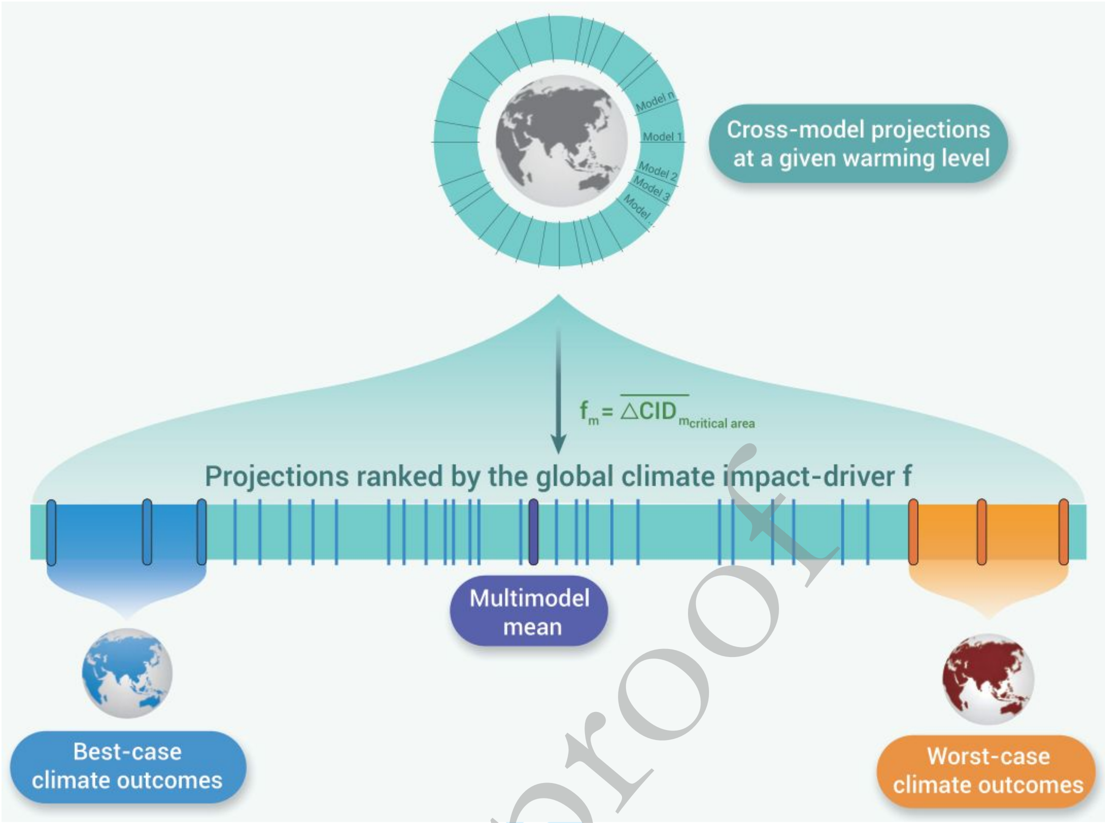


Figure. 1 Framework for identifying plausible extreme climate outcomes from multi-model projections At a given warming level (2°C in *Bevacqua et al.*)[1], model projections are ranked according to a sector-specific global climatic impact-driver, calculated as the spatially averaged change in a relevant hazard across the critical region for that sector. This procedure enables the identification of plausible high- and low-impact futures within the model ensemble. For food security, for example, the metric may be based on projected drought changes across major breadbasket regions. Adapted from Fig. 1 in *Bevacqua et al.*[1] under the terms of the CC BY 4.0 license.

Beyond the multimodel ensemble mean (MME): Assessing plausible climate risk at a given warming level

Hainan Gong[1,2], Wentao Xia[1,5], Lin Wang[1,2*], Kangjie Ma[3], Wen Chen[4]

[1]Key Laboratory of Earth System Numerical Modeling and Application, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[2]Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC‐FEMD), Nanjing University of Information Science & Technology, Nanjing 210044, China.

[3]College of Atmospheric Sciences, Lanzhou University, Lanzhou 730000, China.

[4]Yunnan International Joint Laboratory of Monsoon and Extreme Climate Disasters /Department of Atmospheric Sciences, Yunnan University, Kunming 650500, China

[5]College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

*Corresponding author: Prof. Lin Wang (wanglin@mail.iap.ac.cn)

The 2°C threshold enshrined in the Paris Agreement is widely viewed as a guardrail against the most dangerous consequences of climate change. Yet the risks associated with any given level of warming are still commonly framed through the multimodel ensemble mean (MME), which mainly summarizes the externally forced response shared across climate models. That benchmark is useful, but it is not sufficient. The climate that societies ultimately confront will be shaped not only by external forcing, but also by differences in modelled process responses and by internal climate variability. Consequently, some of the most consequential future impacts may arise not from the ensemble average, but from plausible high-impact outcomes located in the tails of the distribution.

THE ILLUSION OF MODERATE RISK

A recent study by Bevacqua and colleagues brings this point into sharp relief.[1] Drawing on CMIP6 simulations, they show that even in a 2°C warmer world, some plausible climate outcomes can produce sectoral hazards that are more severe than the MME response at 3°C or even 4°C warming. Their analysis challenges a familiar assumption in climate communication: that the most extreme climate outcomes are chiefly a concern under very high warming. In reality, moderate warming does not necessarily translate into moderate risk. The significance of the study lies not merely in emphasizing model spread, but in demonstrating how that spread can be interpreted in a sector-relevant manner. Rather than focusing only on local extremes or broad global averages, the authors define sector-specific "global climatic impact-drivers" that connect each hazard to the regions where it matters most. In their framework, extreme five-day precipitation over densely populated areas serves as a proxy for flood risk, drought frequency across major breadbasket regions represents food-security risk, and fire weather over forests reflects threats to ecosystems and wildfire exposure (Figure 1). Models are then ranked according to these sector-relevant impact drivers, making it possible to identify plausible best- and worst-case outcomes at the same warming level.

These results have important implications. For drought across global breadbaskets, the spread at 2°C is so large that several model outcomes exceed the MME response at 4°C. For extreme precipitation over populated regions and fire weather over forests, the worst-case outcomes at 2°C surpass the MME response at 3°C. Put differently, the upper tail of risk in a 2°C world can overlap with, or even exceed, the average risk expected in a substantially warmer world. For climate-sensitive sectors, this is not a minor statistical detail. It is a direct challenge to the way risk is assessed and adaptation is planned.

WHY THE MME IS NOT THE FUTURE CLIMATE

The MME remains a powerful benchmark for identifying robust forced responses and for reducing the influence of idiosyncratic biases in individual models. However, it should not be interpreted as the full risk space of future climate. It is a statistical summary of model responses, not the climate outcome that societies will necessarily experience. Future climate hazards relevant to risk emerges from the combined influence of external forcing, differences in model responses, and internal variability,[2,3] all of which can shape regional hazards. This distinction is especially important for extreme events. Many societally disruptive extremes are not simply linear extensions of the forced mean state. Instead, they often depend on specific circulation regimes and oceanic background conditions. These ingredients are sensitive to both model formulation and internal variability, which can amplify, mute, or reorganize regional hazards across interannual to decadal timescales.[3] A projection framework centered only on MME change may therefore underestimate the likelihood, intensity, or spatial coherence of future high-impact extremes.

At the same time, the study offers an important clarification regarding the sources of uncertainty. For the sectors emphasized by Bevacqua and colleagues, much of the spread at 2°C appears to stem from model differences rather than from internal variability alone. That point matters because it underscores the central importance of model performance and process representation for improving future risk assessments. Yet internal variability cannot be dismissed. It still contributes to the envelope of plausible outcomes and becomes particularly relevant when considering realized regional climate trajectories over the coming decades. The relative importance of these uncertainty sources also depends strongly on timescale. Internal variability is especially relevant for near- to medium-term projections (e.g., 2030–2050), because it can modulate the timing, regional expression, and realized severity of climate hazards. By contrast, over longer timescales, model structural uncertainty, including differences in climate sensitivity and process representation, generally becomes a larger contributor to the spread of projected sectoral hazards. This distinction reinforces the need to consider both sources of uncertainty, while recognizing that their relative roles vary with lead time, region, and impact metric. A balanced reading of the study is therefore that both model performance and internal variability matter, even though their relative importance is likely to vary across sectors, timescales, and impact metrics.

CONSTRAINING THE FUTURE THAT MATTER

Recognizing these multiple sources of uncertainty has an important implication: climate-risk assessment should not stop at describing average change, but should also examine the plausibility and relevance of outcomes away from the MME. The contribution of Bevacqua and colleagues is precisely to show that such outcomes can be highly consequential even under moderate warming. This perspective points to several priorities for future research. First, extreme outcomes identified within climate ensembles should be subjected to careful process-based evaluation. Ensemble tails should not be equated automatically with physically plausible futures. Some high-impact outcomes may reflect credible process responses, whereas others may arise from known model deficiencies or unrealistic regional teleconnections. Assessing plausibility therefore requires process-based evaluation, including whether the associated circulation changes, moisture transport, land–atmosphere feedbacks, or fire-weather responses are dynamically and physically consistent with observations and established mechanisms.[4]

Second, improving model performance remains fundamental. Differences in climate sensitivity and in the treatment of clouds, convection, land–atmosphere coupling, and hydrological processes can strongly affect projections of drought, heavy rainfall, and wildfire-conducive weather. Better model fidelity, together with physically grounded constraints, should therefore help refine the range of plausible future hazards.[4] Third, internal variability deserves more explicit attention rather than being treated merely as residual noise. Even when it is not the leading source of spread for a given sector, it can still modulate the timing, regional expression, and realized severity of future extremes.[2,3] This is particularly relevant for near-term regional planning, where decadal circulation and ocean–atmosphere variability may amplify, delay, or reorganize the forced response. For longer-term assessments, however, model structural uncertainty often becomes more important, because differences in climate sensitivity, circulation responses, convection, land–atmosphere coupling, and hydrological processes shape the spread of sectoral hazards such as drought, heavy precipitation, and fire-weather changes. Thus, adaptation planning should combine near-term regional storylines of internal variability with longer-term process-based evaluation of model credibility and high-impact ensemble tails. Although predictability remains limited and depends on the phenomenon considered, better characterization of internal variability would provide a more realistic picture of near- to medium-term risk. The goal is not to replace the MME, but to place it in a broader risk framework that combines the average forced response with physically grounded scrutiny of high-impact tails, improved model evaluation, and clearer characterization of internal variability. Such an approach would be particularly valuable in climate-risk hotspots, including densely populated urban areas, major crop-producing regions, ecologically fragile systems, and critical infrastructure corridors, where underestimating plausible extremes can carry outsized societal costs.[5]

IMPLICATIONS FOR RISK AND POLICY

This perspective also carries a clear message for climate policy. Warming thresholds such as 1.5°C and 2°C remain indispensable benchmarks, but they should not be interpreted as guarantees of proportionally moderate impacts. A 2°C world may still contain plausible futures in which drought, flood, and fire risks are comparable to, or even exceed, the average impacts associated with much warmer worlds. This message should not be interpreted as weakening the case for mitigation. Although some high-impact outcomes at 2°C may exceed the ensemble-mean response at higher warming levels, further warming would still shift the overall distribution of hazards toward more severe outcomes and increase the likelihood of crossing critical impact thresholds. Attention to high-impact outcomes at moderate warming therefore reinforces the need for both rapid mitigation and robust adaptation[5]. Adaptation planning should therefore be informed not only by likely futures summarized by the MME, but also by credible high-impact tails within the same warming level. The broader lesson is straightforward. Hazard-based indicators of sectoral risk are not defined solely by where the MME lies. It is also shaped by what remains physically plausible within the ensemble spread, and by whether that spread can be better understood through process evaluation, improved model performance, and clearer characterization of internal variability. Bevacqua and colleagues offer an important warning against equating moderate warming with moderate risk. The next step is to translate that warning into climate assessments that move beyond average projections and are better able to inform decisions about the plausible extremes that matter most.

REFERENCES

1. Bevacqua E., Fischer E., Sillmann J., et al. (2026). Moderate global warming does not rule out extreme global climate outcomes. *Nature* **651**: 946–953. DOI: 10.1038/s41586-026-10237-9
2. Deser C., Lehner F., Rodgers K. B., et al. (2020). Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**: 277-286. DOI: 10.1038/s41558-020-0731-2

3. Gong H., Wang L., Screen J. et al. (2025). Teleconnection from Arctic warming suppresses long-term warming in central Eurasia. *Sci. Adv.* **11**: eadq9461. DOI: 10.1126/sciadv.adq9461

4. Williamson M. S., Thackeray C. W., Cox P. M.et al. (2021). Emergent constraints on climate sensitivities. *Rev. Mod. Phys.* **93**: 025004. DOI: 10.1103/RevModPhys.93.025004

5. Intergovernmental Panel on Climate Change (IPCC). (2023). Key risks across sectors and regions. *Climate Change 2022 – Impacts, adaptation and vulnerability: Working group II contribution to the sixth assessment report of the intergovernmental panel on climate change* pp:2411–2538. Cambridge University Press. DOI: https://doi.org/10.1017/9781009325844.025

168 **FUNDING AND ACKNOWLEDGMENT**

169 This work was supported by the National Natural Science Foundation of China
170 (42522508, 42275023), and the Chinese Academy of Sciences Project for Young
171 Scientists in Basic Research (YSBR-137). The funders had no role in study design, data
172 collection and analysis, decision to publish, or preparation of the manuscript.

173 **AUTHOR CONTRIBUTIONS**

174 Hainan Gong: Writing-review & editing, Writing–original draft, Methodology,
175 Funding, Investigation. Wentao Xia: Visualization, Software, Investigation. Lin Wang:
176 Validation, Conceptualization, Writing-review & editing. Kangjie Ma: Visualization,
177 Software, Investigation. Wen Chen: Validation, Conceptualization. All authors
178 contributed to the manuscript and approved the final version.

179 **DECLARATION OF INTERESTS**

180 The authors declare no conflicts of interest.

181

Figure caption

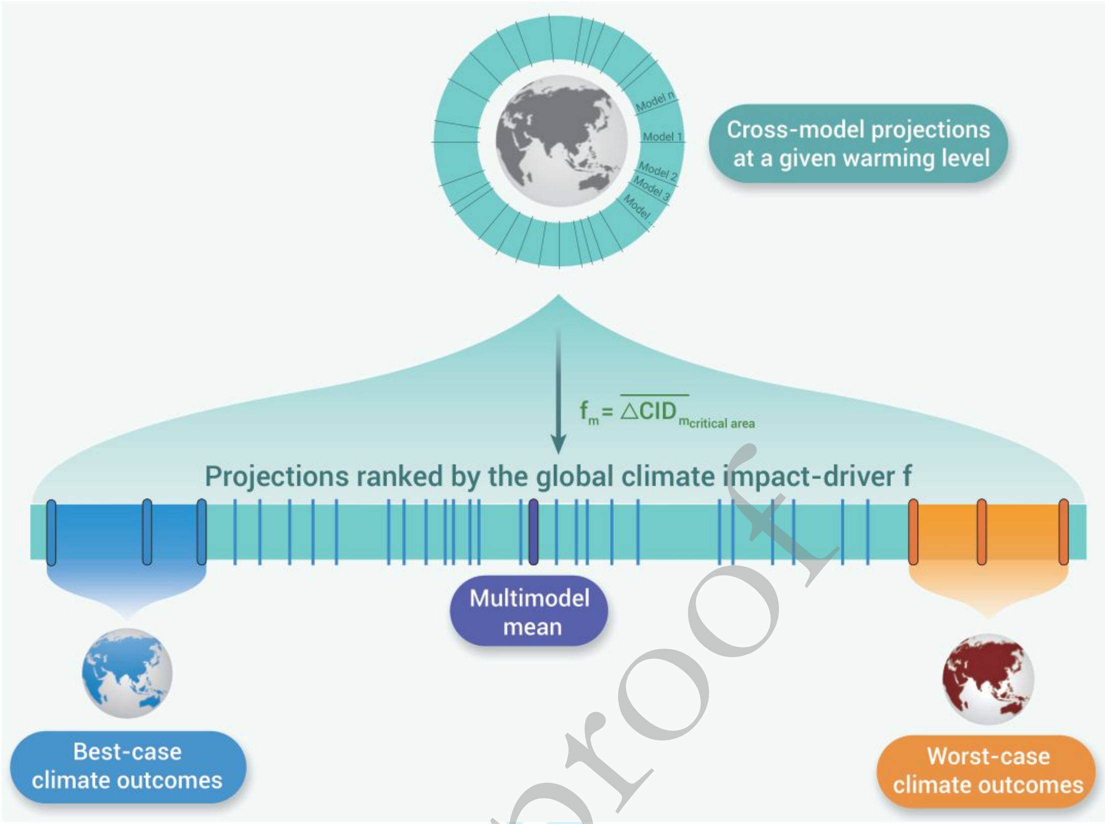


Figure. 1 Framework for identifying plausible extreme climate outcomes from multi-model projections At a given warming level (2°C in *Bevacqua et al.*)[1], model projections are ranked according to a sector-specific global climatic impact-driver, calculated as the spatially averaged change in a relevant hazard across the critical region for that sector. This procedure enables the identification of plausible high- and low-impact futures within the model ensemble. For food security, for example, the metric may be based on projected drought changes across major breadbasket regions. Adapted from Fig. 1 in *Bevacqua et al.*[1] under the terms of the CC BY 4.0 license.