

# Widespread changes in surface temperature persistence under climate change

<https://doi.org/10.1038/s41586-021-03943-z>

Jingyuan Li<sup>1</sup> & David W. J. Thompson<sup>1,2✉</sup>

Received: 12 October 2020

Accepted: 24 August 2021

Published online: 17 November 2021

 Check for updates

Climate change has been and will be accompanied by widespread changes in surface temperature. It is clear that these changes include global-wide increases in mean surface temperature and changes in temperature variance that are more regionally-dependent<sup>1–3</sup>. It is less clear whether they also include changes in the persistence of surface temperature. This is important as the effects of weather events on ecosystems and society depend critically on the length of the event. Here we provide an extensive survey of the response of surface temperature persistence to climate change over the twenty-first century from the output of 150 simulations run on four different Earth system models, and from simulations run on simplified models with varying representations of radiative processes and large-scale dynamics. Together, the results indicate that climate change simulations are marked by widespread changes in surface temperature persistence that are generally most robust over ocean areas and arise due to a seemingly broad range of physical processes. The findings point to both the robustness of widespread changes in persistence under climate change, and the critical need to better understand, simulate and constrain such changes.

The climate system response to increasing greenhouse gases includes changes not only in the mean temperature but also in the shape of the temperature distribution<sup>1–3</sup>. Both are essential for understanding changes in the amplitude and timescale of temperature events under climate change. Increases in the mean temperature alter the likelihood of temperature events by shifting the probability distribution towards higher values; changes in the temperature variance and skewness affect the incidence of temperature events by altering the shape of the distribution. The evidence for changes in the mean surface temperature is overwhelming and incontrovertible in both observations and numerical simulations of climate change<sup>1–3</sup>. The evidence for changes in the variance and skewness of the temperature distribution is less extensive. Nevertheless, there is evidence that both have changed in select regions for a variety of physical reasons<sup>4–15</sup>.

Here we examine the evidence and physics of changes in the persistence of temperature events under climate change. Persistence has a key role in the climate impacts of a given temperature event. But interestingly, relatively few studies have explored how and why persistence will evolve under climate change in a global context. To date, the evidence is limited.

One set of studies argues for increases in observed temperature persistence in data through the early 2000s (refs. <sup>16,17</sup>), but the changes in observed persistence are not clearly reproduced in different datasets, at different locations, and over different time periods (for example, see Figs. 5 and S10 of ref. <sup>17</sup>). Another observational study indicates sporadic changes in the duration of observed extreme temperature events over the latter half of the 20th century<sup>18</sup>. But the most significant changes are limited to the summer months and to land data averaged

over select spatial regions. An analysis of output from CMIP5 (the Coupled Model Intercomparison Project Phase 5) infers changes in temperature persistence from changes in the shape of temperature power spectra<sup>19</sup>, but the results are characterized by considerable regional and inter-model variability. An analysis of atmospheric models forced with prescribed 2 K increases in sea surface temperatures (SSTs) reveals regional increases in land surface temperature persistence<sup>20</sup>, but the experiment design has known biases in the attribution of extreme events<sup>21</sup> and, importantly, prescribing the SST field renders surface temperature persistence fixed over ocean areas. Numerous studies argue that Arctic sea ice loss leads to systematic changes in the mid-latitude circulation, including its persistence (for example, ref. <sup>22</sup>). However, the proposed linkages have not been reproducible in many observational and modelling studies and are therefore controversial (for example, ref. <sup>23</sup>).

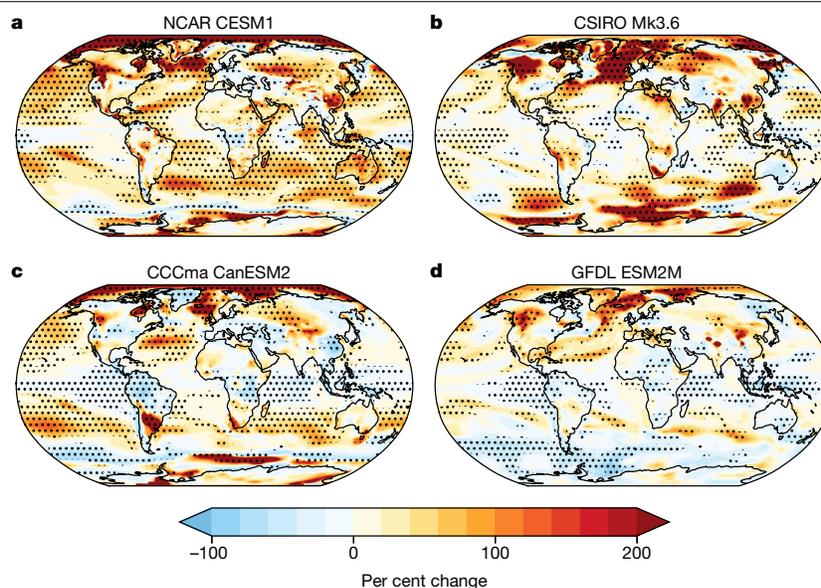
Here we reveal that climate change leads to pervasive changes in temperature persistence in simulations run on a range of climate models, including comprehensive Earth system models (ESMs) and simpler models with varying representations of radiative and dynamical processes. The changes in persistence arise from a seemingly wide range of physical processes. The results highlight the ubiquitous nature of changes in temperature persistence under climate change, and the critical importance of better understanding and constraining such changes.

## Persistence changes in ESMs

We begin by exploring changes in persistence in large ensembles of climate change simulations run on four ESMs: the NCAR CESM1, CSIRO

<sup>1</sup>Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA. <sup>2</sup>School of Environmental Sciences, University of East Anglia, Norwich, UK.

✉e-mail: [davet@atmos.colostate.edu](mailto:davet@atmos.colostate.edu)



**Fig. 1 | Changes in surface temperature persistence in ESMs. a–d.** The changes in persistence between the Historical period (1970–1999) and the Future period (2070–2099) calculated from large ensembles run on four ESMs. Warm (red) colours represent an increase in persistence from the Historical to Future periods, while cool (blue) colours represent a decrease in persistence over the same period. Results show the per cent changes in the variance explained by the lag 10-day autocorrelation; that is, the per cent changes in  $r_{10}^2$ . Note that  $r_{10}^2$  is calculated first for individual ensemble members and then averaged over all

ensemble members. The results are derived from 40 ensemble members run on the NCAR CESM1 (a), 30 ensemble members run on the CSIRO Mk3.6 (b), 50 ensemble members run on the CCCma CanESM2 (c), and 30 members run on the GFDL ESM2M (d). Stippling indicates grid points where at least 75% of the ensemble members agree on the sign of the change (a likelihood of  $>0.1\%$  by chance) and where the ensemble mean results exceed the 95% confidence threshold based on a two-tailed test of the  $t$ -statistic. See Methods for details of the ESM output, analysis, statistical significance and reproducibility.

Mk3.6, CCCma CanESM2 and GFDL ESM2M (Methods). Surface temperature persistence is estimated as the lag 10-day autocorrelation of daily mean temperatures,  $r_{10}$ . The variance explained by persistence is thus estimated as  $r_{10}^2$ , and changes in persistence under climate change are estimated as the per cent changes in  $r_{10}^2$  between two 30-year periods: the ‘Historical’ period 1970–1999 and ‘Future’ period 2070–2099. The autocorrelation provides a simple and robust way of measuring persistence and is directly related to the length of temperature events (Methods; Extended Data Figs. 1 and 2). In addition, results are largely insensitive to the choice of time lag used in the analyses (Methods; Extended Data Figs. 3, 4 and 5). See Methods for details of the ESM output, analysis, statistical significance and reproducibility.

The climatological-mean values of  $r_{10}$  for the Historical and Future periods are shown for reference in Extended Data Fig. 6. In general, the climatological-mean persistence is highest in the tropics and decreases rapidly with latitude in accordance with the more pronounced amplitude of weather ‘noise’ at middle and high latitudes. The per cent changes in  $r_{10}^2$  between the two periods are shown in Fig. 1. Both the CESM1 and Mk3.6 ESMs indicate widespread increases in persistence over time that exceed 50% over much of the globe (Fig. 1a,b). The CanESM2 and ESM2M ESMs likewise indicate widespread increases in persistence across the Northern Hemisphere—particularly over the ocean basins, northwestern North America, and central Siberia—and throughout the Southern Hemisphere subtropics (Fig. 1c,d). However, the CanESM2 and ESM2M results also indicate decreases in persistence in the tropics and—particularly in the ESM2M—over the high latitudes of the Southern Hemisphere. As discussed below, many of the inter-model differences are consistent with inter-model differences in the El-Niño/Southern Oscillation (ENSO) response to climate change and the amplitude of mixing in the Southern Ocean.

The robustness of the changes in persistence is further explored in Fig. 2. All four ESMs indicate robust increases in regional-scale surface temperature persistence throughout the Arctic (Fig. 2a–d) and the Northern

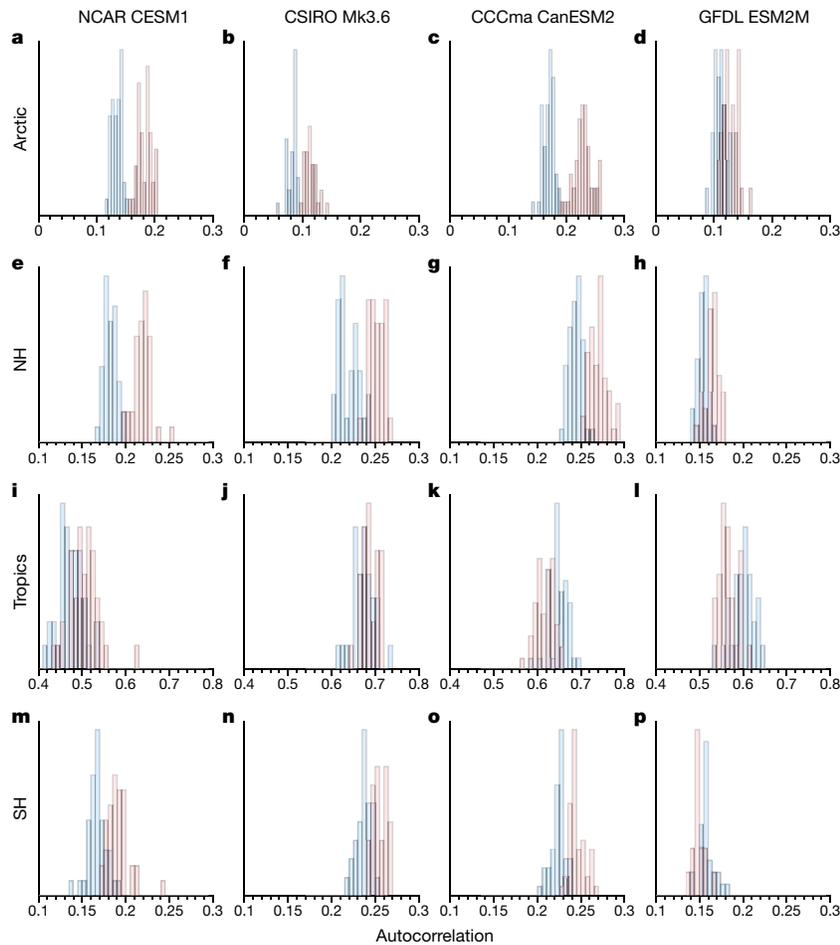
Hemisphere (Fig. 2e–h). The increases over the Arctic are particularly pronounced and are consistent with observed increases in the duration of Arctic warm events<sup>24</sup>. Three out of four ESMs indicate similarly robust increases in regional-scale surface temperature persistence in the Southern Hemisphere (Fig. 2m–o). The most notable exception is the ESM2M, which indicates weak decreases in persistence over the Southern Ocean (Figs. 1 and 2p). Consistent with Fig. 1, the simulations run on the CESM1 indicate increases in persistence in the tropics (Fig. 2i) while the simulations run on the CanESM2 and ESM2M indicate decreases there (Fig. 2k–l).

The results in Figs. 1 and 2 are shown separately for output from four large ensembles, and thus account for the uncertainty due to internal climate variability. Fig. 3 probes the same results averaged over all four large ensembles, and thus accounts for the uncertainty due to both internal and model–model variability. As can also be inferred from Fig. 1, the most robust responses include: (1) widespread increases in persistence across the middle and high latitudes of the Northern Hemisphere, including the Arctic, the North Pacific and North Atlantic basins, northwestern North America and central Siberia; (2) widespread increases in persistence throughout the subtropical oceans in both hemispheres; and (3) localized decreases in persistence along the periphery of Antarctica. The changes in the tropics are generally not robust from one model to the next. As noted above and discussed further below, they are highly dependent on model simulations of changes in ENSO under climate change.

### Changes in simplified numerical models

Similarly robust changes in persistence are found in climate change experiments run on relatively simple numerical models. Here we consider output from three sets of climate change experiments run with no ocean dynamics, no land or sea ice, and varying representations of atmospheric radiative processes and large-scale dynamics.

The first two sets are from the ‘longwave hierarchy’ of numerical models published in ref. <sup>25</sup>. Both sets include a ‘control’ and a ‘4 × CO<sub>2</sub>’



**Fig. 2 | Spatially averaged surface temperature persistence.** **a–p.** The lag-10-day autocorrelation  $r_{10}$  averaged over the indicated regions and shown for individual ensemble members during the Historical period (1970–1999; blue) and Future period (2070–2099; red). Note that  $r_{10}$  is calculated first at individual grid boxes and then averaged over the indicated regions. As such, the results reflect persistence calculated at the grid box level. The Arctic is defined as the region poleward of 65 degrees North; the Northern Hemisphere (NH) and Southern Hemisphere (SH) extratropics are defined as the regions poleward of 20 degrees; the tropics are defined as the region 20S to 20N. The results are derived from 40 ensemble members run on the NCAR CESM1 (**a, e, i, m**), 30 ensemble members run on the CSIRO Mk3.6 (**b, f, j, n**), 50 ensemble members run on the CCCma CanESM2 (**c, g, k, o**), and 30 members run on the GFDL ESM2M (**d, h, l, p**). See Methods for details of the ESM output and analysis.

simulation run on an atmospheric general circulation model coupled to a slab ocean model. The differences between the two sets lies in their treatment of atmospheric radiation: in the first set, the radiative effects of atmospheric composition are parameterized by prescribing the optical depth (that is, a ‘gray radiation’ scheme); in the second set, they are calculated using a comprehensive radiation scheme (RRTMG). See Methods and ref. <sup>25</sup> for details.

Figure 4 shows the autocorrelation functions of near-surface temperatures from both sets of experiments averaged over all grid points in the extratropics (Fig. 4a,b) and tropics (Fig. 4c,d). In both regions, surface temperature persistence increases notably under climate change when the atmospheric optical depth is increased either explicitly (Fig. 4a,c) or through the attendant changes in greenhouse gases (Fig. 4b,d). The increases in persistence are most pronounced in the numerical configuration that includes an interactive water vapor feedback (Fig. 4b,d), potentially due to the importance of the water vapor feedback in low-frequency climate variability<sup>26</sup>. In both sets of simulations, the increases are notable as they arise in the absence of changes in sea ice, ocean dynamics, clouds, the land surface, and the ocean mixed-layer depth.

The third set of climate change experiments is from the radiative/convective equilibrium (RCE) simulations published in ref. <sup>27</sup>. Again, the experiments include a control and a  $4 \times \text{CO}_2$  simulation run on an atmospheric general circulation model coupled to a slab ocean model. But in this case the insolation is spatially homogeneous and the Coriolis parameter is set to zero everywhere (that is, the model is run in a ‘tropics-world’ configuration (Methods)<sup>27</sup>. The model configuration

includes clouds and comprehensive radiation, but has no sea ice, no land surface, no ENSO, fixed surface heat capacity, and no extratropical dynamics. The RCE output is only available in monthly-mean form and thus the lag correlations are discretized by month. Nevertheless, the output affords a very large sample size (Methods). As shown in Fig. 5, the simulations again indicate marked increases in surface temperature persistence under increasing levels of atmospheric  $\text{CO}_2$ .

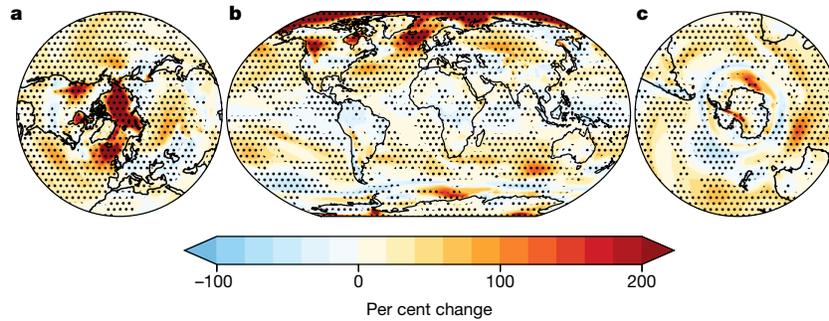
## Interpretation

What physical processes might underlie the widespread changes in temperature persistence indicated above? For pedagogical purposes, consider a linearized version of the surface energy budget:

$$C_{\text{eff}} \frac{dT'_s}{dt} = F' - \lambda T'_s, \quad (1)$$

where primes denote variations about a long-term mean state;  $T'_s$  denotes the surface temperature anomaly;  $C_{\text{eff}}$  denotes the effective heat capacity of the ocean-mixed layer and/or land surface;  $F'$  represents the anomalous surface fluxes associated with internal climate variability; and  $\lambda$  is a damping coefficient. The term  $-\lambda T'_s$  represents the damping of surface temperature anomalies by the surface fluxes of latent, sensible and radiant heat.

There are three ways to change the persistence of  $T'_s$  in equation (1). One is to change the time-varying characteristics of the forcing,  $F'$ ; a second is to change the effective heat capacity  $C_{\text{eff}}$ ; and a third is to change the damping coefficient  $\lambda$ . We consider all three below.



**Fig. 3 | Changes in persistence averaged over all four ESMs.** a–c, The changes in persistence between the Historical period (1970–1999) and the Future period (2070–2099) averaged over 150 ensemble members run on four ESMs. Warm (red) colours represent an increase in persistence from the Historical to Future periods, while cool (blue) colours represent a decrease in persistence over the same period. Results show the per cent changes in the variance explained by the lag 10-day autocorrelation; that is, the per cent changes in  $r_{10}^2$ . Note that  $r_{10}^2$  is calculated first for individual ensemble members and then averaged over all ensemble members from all four ESMs. The results are averaged over 40

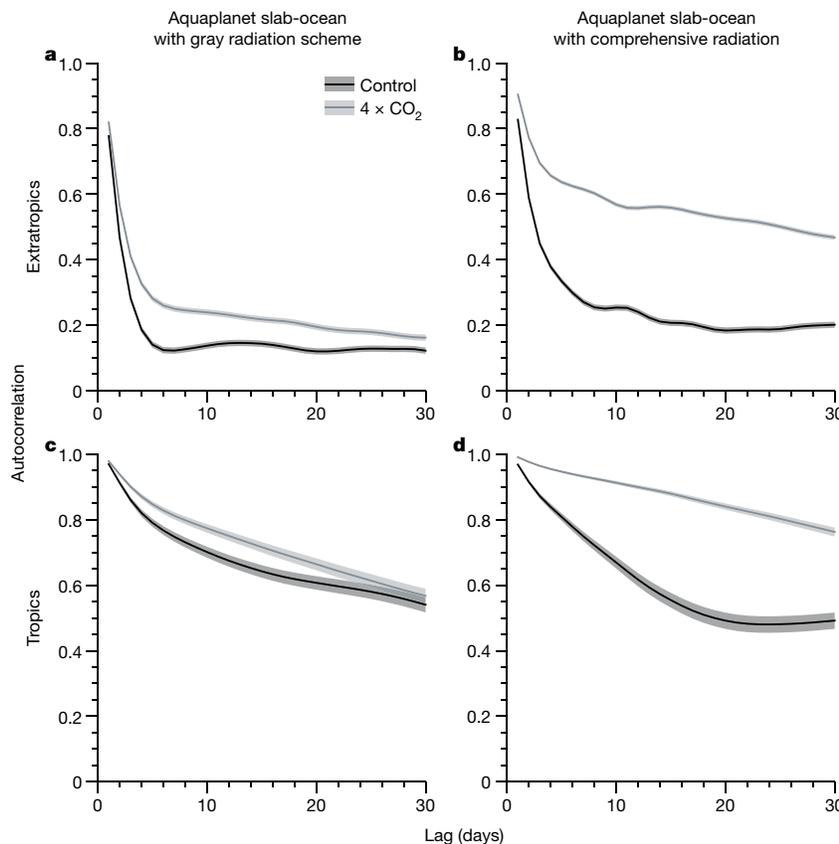
ensemble members run on the NCAR CESM1, 30 ensemble members run on the CSIRO Mk3.6, 50 ensemble members run on the CCCma CanESM2 and 30 members run on the GFDL ESM2M. Stippling indicates grid boxes where the multi-model mean changes are significant at the 95% level based on the  $t$ -statistic, and the signs of the changes are the same in ensemble means from at least three out of the four ESMs. See Methods for details of the ESM output, analysis, statistical significance and reproducibility. a, b, c show the same results from three different perspectives to highlight different regions of the globe.

**Potential changes in the forcing**

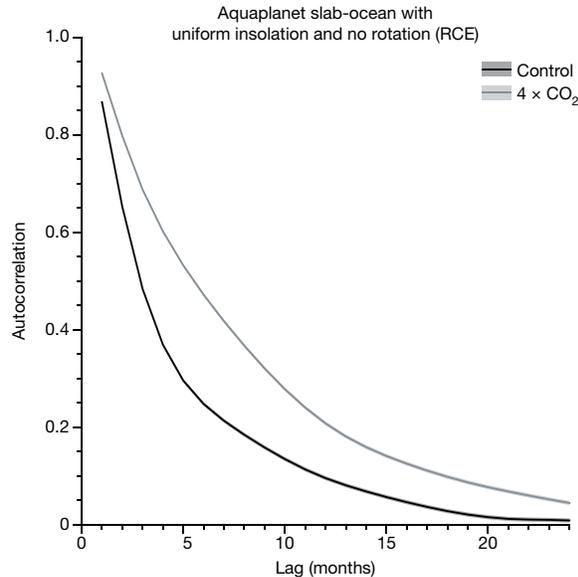
The timescale of the forcing ( $F'$ ) is linked to the persistence of large-scale atmospheric dynamics and their attendant surface fluxes. Trends in the midlatitude circulation have been linked to changes in surface temperature persistence in some climate simulations<sup>20</sup>, the persistence of the extratropical circulation has been suggested to vary with the latitude of the storm track<sup>28</sup>, and many climate models indicate decreases

in the incidence of atmospheric blocking<sup>29</sup>. However, trends in the midlatitude circulation vary from one climate change simulation to the next<sup>30</sup> and there is considerable uncertainty in the response of blocking to climate change; for example, the simulated changes in blocking are generally smaller than model biases in blocking incidence<sup>29</sup>.

The timescale of the forcing is also influenced by the amplitude of the ENSO phenomenon, as ENSO has pronounced amplitude on seasonal timescales and influences climate throughout the tropics<sup>31</sup> and



**Fig. 4 | Changes in persistence in slab-ocean numerical models.** a–d, The autocorrelation functions of surface temperature from two sets of climate change simulations run on an aquaplanet model with a fixed slab-ocean depth. Shown are results from a version of the model with a gray-radiation scheme in which the longwave optical depth is prescribed (a, c), and from a version of the model with a comprehensive radiation scheme (RRTMG) (b, d). Results are first calculated at individual grid points and then averaged poleward of 20 degrees latitude (a, b) and between 20 N and 20 S (c, d). Dark grey curves show results from the control simulations; light grey curves show results from climate change simulations in which optical depth or greenhouse gas concentrations are increased in a manner consistent with a fourfold increase in carbon dioxide from the control. Shading represents the 95% confidence intervals on the autocorrelation values (Methods). The model output is from ref. <sup>25</sup>. See Methods for details of the model output, analysis and statistical significance.



**Fig. 5 | Changes in persistence in a simplified ‘tropics-world’ simulation.** The autocorrelation functions of near-surface temperature from a climate change simulation run on an aquaplanet model with a fixed slab-ocean depth, no rotation and spatially homogenous insolation. Results are first calculated at individual grid points and then averaged over the globe. The dark grey curve shows results from a control simulation; the light grey curve shows results from a climate change simulation in which greenhouse gas concentrations are increased by a factor of four. Shading represents the 95% confidence intervals on the autocorrelation values (Methods; note that the confidence intervals are very small due to the very large sample size used in the analysis). The model output is from ref. <sup>27</sup>. See Methods for details of the model output, analysis and statistical significance.

extratropics<sup>32</sup>. The inter-model differences in persistence changes indicated in Figs. 1 and 2 are consistent with inter-model differences in simulated changes in ENSO variability. The CESM1 and Mk3.6 simulations indicate increases over the twenty-first century in both tropical surface temperature persistence (Figs. 1 and 2) and the variance of the ENSO phenomenon (Extended Data Fig. 7; see also ref. <sup>33</sup> and references therein). The ESM2M and CanESM2 simulations indicate opposite changes in both (Figs. 1 and 2, and Extended Data Fig. 7; see also ref. <sup>34</sup>). Despite the probable importance of ENSO for future changes in temperature persistence, it is worth emphasizing that robust changes in temperature persistence readily arise in simulations with no ENSO variability, as shown in Figs. 4 and 5.

### Potential changes in the heat capacity

A second way to increase the persistence of  $T'_s$  is to change the heat capacity  $C_{\text{eff}}$ . If the timescale of  $F'$  is much shorter than the timescale of  $T'_s$ , then the autocorrelation decays exponentially with lag as follows<sup>35,36</sup>:

$$r(\tau) = \exp\left[\frac{-\lambda}{C_{\text{eff}}}\tau\right]. \quad (2)$$

Thus, surface temperature persistence increases as the surface heat capacity increases and/or the damping coefficient decreases, and vice versa.

Over the Arctic, increases in the areal coverage of open water under climate change should lead to increases in  $C_{\text{eff}}$ , as the effective heat capacity of open water is much higher than that of sea ice. The large

increases in Arctic surface temperature persistence simulated by the ESMs (Figs. 1–3) are broadly consistent with the melting of sea ice there.

Over the oceans, the effective heat capacity of the surface is related to the depth of the ocean mixed layer. Numerous simulations suggest increases in near-surface ocean stratification and thus decreases in mixed-layer depth under climate change<sup>37,38</sup>, and the simulations run on CESM1 indicate decreases in mixed-layer depth over the twenty-first century<sup>39</sup>. All else being equal, decreases in mixed-layer depth will lead to decreases in  $C_{\text{eff}}$  and thus in surface temperature persistence. Interestingly, the ESMs explored here indicate increases in persistence over most extratropical maritime regions (Fig. 1) including, notably, the North Pacific in the CESM1 where mixed-layer depths are decreasing (see Fig. 1g in ref. <sup>39</sup>). Thus, other factors are probably key in these regions.

Changes in ocean stratification and mixed-layer depth are probably dominant for the simulated decreases in Southern Ocean temperature persistence in the ESM2M. Relative to other ESMs, the ESM2M has comparatively robust ocean convection and thus deep mixed layers in the Southern Ocean<sup>40</sup>. This is supported by the comparatively weak Southern Ocean warming, small sea-ice losses, and large decreases in Southern Ocean convection in ESM2M under climate change (Extended Data Fig. 8; refs. <sup>41–44</sup>). In general, the largest increases in ocean stratification under climate change should arise in regions where the mixed layer is deepest under historical conditions (for example, ref. <sup>37</sup>). Thus, the relatively robust Southern Ocean convection under historical conditions in the ESM2M should lead to relatively large increases in stratification there and (all else being equal) decreases in SST persistence.

Over terrestrial regions, the effective heat capacity of the surface may change; for example, via variations in snow cover, soil moisture and vegetation. Soil moisture–temperature feedbacks have been linked to changes in surface temperature variance over select areas (for example, ref. <sup>5</sup>). However, the influence of land surface changes on surface temperature variability varies from one region to the next and depends on model representations of land-surface processes<sup>45</sup>. As shown in Fig. 1, the most robust changes in persistence are found not over land areas but over the oceans.

Importantly, the idealized climate change simulations highlighted in Figs. 4 and 5 are run with fixed mixed-layer depths, no land surface, and no sea ice. They therefore indicate that widespread increases in persistence can readily arise even in the absence of changes in the surface heat capacity.

### Potential changes in damping

A third way to change surface temperature persistence is through changes in the damping term  $-\lambda T'_s$ . The damping term reflects a parameterization of the feedback between a perturbation in surface temperature and the resulting anomalous fluxes of sensible, latent and radiative heat.

The feedbacks due to the surface fluxes of sensible and latent heat are a function of the near-surface wind speed and the vertical gradients in the temperature and specific humidity perturbation within the atmospheric boundary layer (for example, ref. <sup>46</sup>). Maritime regions where the mean wind increases under climate change may thus be marked by increases in the turbulent air–sea feedback and (from equation (2)) decreases in surface temperature persistence. This process may prove important over the high latitude Southern Ocean where the mean winds are expected to increase under climate change<sup>2</sup>. In the absence of changes in the surface wind speed, the turbulent air–sea feedback generally increases with SSTs (see Fig. 4 in ref. <sup>47</sup>) and is thus unlikely to contribute to increases in surface temperature persistence.

The feedback due to the surface fluxes of longwave radiation is a function of the longwave optical depth of the atmosphere. Increases in the longwave optical depth under climate change should lead to a reduction in the thermal damping—and thus an increase in the persistence—of

# Article

surface temperature anomalies. For example, consider the case of a warm temperature anomaly applied at the surface. The damping of the anomaly by the anomalous upward flux of longwave radiation is attenuated by the anomalous downward flux of longwave radiation that arises as the lower atmosphere warms<sup>48</sup>. Under climate change, the surface is radiatively coupled to a lower and warmer level of the atmosphere, and thus the surface longwave radiative fluxes should be less efficient in damping a given surface temperature anomaly. Radiative coupling between the surface and atmosphere should contribute to the persistence of surface temperature anomalies for the same reasons that thermal coupling contributes to the persistence of midlatitude SST anomalies<sup>49</sup> and the tropospheric relaxation timescale to radiative–convective equilibrium<sup>50</sup>.

The importance of changes in radiative damping for changes in surface temperature persistence is suggested by the numerical experiments shown in Figs. 4 and 5. The simulations indicate marked increases in persistence, despite the fact that there is no land surface (Figs. 4 and 5), no sea ice (Figs. 4 and 5), no clouds (Fig. 4), no ocean dynamics (Figs. 4 and 5), no changes in slab-ocean depth (Figs. 4 and 5), and no extratropical dynamics (Fig. 5). We reason that changes in longwave radiative damping have a potentially key but largely overlooked role in changes in temperature persistence under climate change.

## Conclusions

Robust and widespread changes in surface temperature persistence arise in climate change simulations run on a variety of numerical configurations. They derive from a seemingly broad array of physical factors. Such changes in temperature persistence project directly onto the timescales of extreme temperature events, and thus have potentially key implications for various societies and ecosystems. The results here point to the critical need to better understand, simulate and constrain changes in surface temperature persistence under climate change.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03943-z>.

- Hartmann, D. L. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 2 (IPCC, Cambridge Univ. Press, 2013).
- Kirtman, B. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 11 (IPCC, Cambridge Univ. Press, 2013).
- Collins, M. et al. in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. et al.) Ch. 12 (IPCC, Cambridge Univ. Press, 2013).
- Schar, C. et al. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
- Seneviratne, S. I., Luthi, D., Litschi, M. & Schar, C. Land–atmosphere coupling and climate change in Europe. *Nature* **443**, 205–207 (2006).
- Fischer, E. M. & Schar, C. Future changes in daily summer temperature variability: driving processes and role for temperature extremes. *Clim. Dyn.* **33**, 917–935 (2009).
- Donat, M. G. & Alexander, L. V. The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.* <https://doi.org/10.1029/2012GL052459> (2012).
- Fischer, E. M., J. Rajczak, & Schar, C. Changes in European summer temperature variability revisited. *Geophys. Res. Lett.* **39**, 19 (2012).
- Volodin, E. M. & Yurova, A. Y. Summer temperature standard deviation, skewness and strong positive temperature anomalies in the present day climate and under global warming conditions. *Clim. Dyn.* **40**, 1387–1398 (2013).
- Lewis, S. C. & Karoly, D. J. Anthropogenic contributions to Australia's record summer temperatures of 2013. *Geophys. Res. Lett.* **40**, 3705–3709 (2013).
- Screen, J. A. Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Clim. Change* **4**, 577–582 (2014).
- Schneider, T., Bischoff, T. & Plotka, H. Physics of changes in synoptic midlatitude temperature variability. *J. Climate* **28**, 2312–2331 (2015).
- McKinnon, K. A., Rhines, A., Tingley, M. P. & Huybers, P. The changing shape of Northern hemisphere summer temperature distributions. *J. Geophys. Res. Atm.* **121**, 8849–8868 (2016).
- Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J. & Shepherd, T. G. A dynamical perspective on atmospheric temperature variability and its response to climate change. *J. Climate* **32**, 1707–1724 (2019).
- Tamarin-Brodsky, T., Hodges, K. L., Hoskins, B. J. & Shepherd, T. Changes in Northern Hemisphere temperature variability shaped by regional warming patterns. *Nat. Geosci.* **13**, 414–421 (2020).
- Boulton, C. A. & Lenton, T. M. Slowing down of North Pacific climate variability and its implications for abrupt ecosystem change. *Proc. Natl Acad. Sci. USA* **112**, 11496–11501 (2015).
- Lenton, T. M., Dakos, V., Bathiany, S. & Scheffer, M. Observed trends in the magnitude and persistence of monthly temperature variability. *Sci. Rep.* **7**, 5940 (2017).
- Pfleiderer, P. & Coumou, D. Quantification of temperature persistence over the Northern Hemisphere land-area. *Clim. Dyn.* **51**, 627–637 (2018).
- Di Cecco, G. J., Gouhier, T. C. Increased spatial and temporal autocorrelation of temperature under climate change. *Sci. Rep.* **8**, 14850 (2018).
- Pfleiderer, P., Schleussner, C., Kornhuber, K. & Coumou, D. Summer weather becomes more persistent in a 2 C world. *Nat. Clim. Change* **9**, 666–671 (2019).
- Fischer, E. M., Beyerle, U., Schleussner, C. F., King, A. D. & Knutti, R. Biased estimates of changes in climate extremes from prescribed SST simulations. *Geophys. Res. Lett.* **45**, 8500–8509 (2018).
- Cohen, J. et al. Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.* **7**, 627–637 (2014).
- Barnes, E. A. & Screen, J. The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *WIREs Clim. Change* **6**, 277–286 (2015).
- Graham, R. M. et al. Increasing frequency and duration of Arctic winter warming events. *Geophys. Res. Lett.* **44**, 6974–6983 (2013).
- Tan, Z., Lachmy, O. & Shaw, T. A. The sensitivity of the jet stream response to climate change to radiative assumptions. *JAMES* **11**, 934–956 (2019).
- Hall, A. & Manabe, S. The role of water vapor feedback in unperturbed climate variability and global warming. *J. Climate* **12**, 2327–2346 (1999).
- Drotos, G., Becker, T., Mauritsen, T. & Stevens, B. Global variability in radiative-convective equilibrium with a slab ocean under a wide range of CO<sub>2</sub> concentrations. *Tellus* **72**, 1–19 (2020).
- Barnes, E. A. & Hartmann, D. L. Testing a theory for the effect of latitude on the persistence of eddy-driven jets using CMIP3 simulations. *Geophys. Res. Lett.* **37**, L15801 (2010).
- Woollings, T. et al. Blocking and its response to climate change. *Curr. Clim. Change Rep.* **4**, 287–300 (2018).
- Barnes, E. A. & Polvani, L. Response of the midlatitude jets and of their variability to increased greenhouse gases in the CMIP5 models. *J. Climate* **26**, 7117–7135 (2013).
- Yulaeva, E. & Wallace, J. M. The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate* **7**, 1719–1736 (1994).
- Alexander, M. A. et al. The atmospheric bridge: the influence of ENSO teleconnections on air–sea interaction over the global oceans. *J. Climate* **15**, 2205–2231 (2002).
- Cai, W. et al. Increasing frequency of extreme El Niño events due to greenhouse warming. *Nat. Clim. Change* **4**, 111–116 (2014).
- Kohyama, T., Hartmann, D. L. & Battisti, D. S. La Nina-like mean-state response to global warming and potential oceanic roles. *J. Climate* **30**, 4207–4225 (2017).
- Frankignoul, C. & Hasselman, K. Stochastic climate models, Part II Application to sea-surface temperature variability and thermocline variability. *Tellus* **29**, 289–305 (1977).
- Deser, C., Alexander, M. A. & Timlin, M. S. Understanding the persistence of sea surface temperature anomalies in midlatitudes. *J. Climate* **16**, 57–72 (2003).
- Capotondi, A., Alexander, M. A., Bond, N. A., Curchitser, E. N. & Scott, J. D. Enhanced upper ocean stratification with climate change in the CMIP3 models. *J. Geophys. Res.* **117**, C04031 (2012).
- Li, G. et al. Increasing ocean stratification over the past half-century. *Nat. Clim. Change* **10**, 1116–1123 (2020).
- Amaya, D. J. et al. Are long-term changes in mixed layer depth influencing North Pacific marine heatwaves? *Bull. Am. Meteorol. Soc.* **102**, S59–S66 (2021).
- Sallée, J. B. et al. Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: historical bias and forcing response. *J. Geophys. Res. Oceans* **118**, 1830–1844 (2013).
- Manabe, S., Stouffer, R. J., Spelman, M. J. & Bryan, K. Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO<sub>2</sub>. Part I. Annual mean response. *J. Climate* **4**, 785–818 (1991).
- de Lavergne, C. et al. Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. *Nat. Clim. Change* **4**, 278–282 (2014).
- Kostov, Y. et al. Fast and slow responses of Southern Ocean sea surface temperature to SAM in coupled climate models. *Climate Dyn.* **48**, 1595–1609 (2017).
- Bronselaer, B. et al. Change in future climate due to Antarctic meltwater. *Nature* **564**, 53–58 (2018).
- Jia, G. et al. in *IPCC Special Report on Climate Change and Land* (eds Shukla, P. R. et al.) Ch. 2 (IPCC, 2019).
- Frankignoul, C., Czaja, A. & L'Heveder, B. Air–sea feedback in the North Atlantic and surface boundary conditions for ocean models. *J. Climate* **11**, 2310–2324 (1998).
- Hausmann, U., Czaja, A. & Marshall, J. Mechanisms controlling the SST air–sea heat flux feedback and its dependence on spatial scale. *Clim. Dyn.* **48**, 1297–1307 (2016).
- Vargas Zeppetello, L. R., Donohoe, A. & Battisti, D. S. Does surface temperature respond to or determine downwelling longwave radiation? *Geophys. Res. Lett.* **46**, 2781–2789 (2019).
- Barsugli, J. J. and D. S. Battisti. The basic effects of atmosphere–ocean thermal coupling on midlatitude variability. *J. Atmos. Sci.* **55**, 477–493 (1998).
- Cronin, T. W. & Emanuel, K. A. The climate time scale in the approach to radiative-convective equilibrium. *JAMES* **5**, 843–849 (2013).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

### Earth system models

We analyse climate change simulations run on four different ESMs: the National Center for Atmospheric Research CESM1, the Commonwealth Scientific and Industrial Research Organisation Mk3.6, the Canadian Centre for Climate Modeling and Analysis CanESM2, and the Geophysical Fluid Dynamics Laboratory ESM2M. The output includes a total of 150 runs, including: (1) 40 simulations run on the CESM1 (ref. <sup>51</sup>); (2) 30 simulations run on the Mk3.6 (ref. <sup>52</sup>); (3) 50 simulations run on the CanESM2 (ref. <sup>53</sup>); and (4) 30 simulations run on the ESM2M<sup>54</sup>. All four ESMs are fully coupled atmosphere–ocean–land–ice climate models. The large-ensemble output is made available through the Multi-Model Large Ensemble Archive with the guidance of the US CLIVAR Working Group on Large Ensembles<sup>55,56</sup>. Output is accessed via the NCAR Climate Variability Diagnostics Package for Large Ensembles<sup>56</sup> and the Climate Data Gateway.

### Persistence in the ESMs

Persistence is estimated from the autocorrelation function. We calculate the autocorrelation of surface temperature in the ESMs as follows.

For each ESM, the daily mean surface temperature output is obtained from all ensemble members from two periods that represent the historical and future climate: 1970–1999 for the Historical period and 2070–2099 for the Future period.

The time-varying ensemble mean temperature is subtracted from all grid points in all ensemble members over both periods. Removing the ensemble mean is essential so that the surface temperature persistence in each ensemble member is not biased by low-frequency variability in the forcings applied to the model. For example, the rate of global warming is not stationary over the entire 1970–2099 period, and thus the latter part of the twenty-first century may exhibit changes in persistence due simply to changes in the rate of global warming.

The persistence of surface temperature is estimated from the autocorrelation of the temperature time series. The autocorrelation is calculated as:

$$r(\tau) = \frac{\overline{T'(t)T'(t+\tau)}}{T'(t)^2}$$

where  $T$  denotes surface temperature, primes denote departures from the ensemble mean, overbars denote the time mean, and  $\tau$  denotes the lag autocorrelation in days.

The ensemble and zonal mean autocorrelations are calculated using the Fisher  $z$ -transformation. That is, we transform autocorrelations to their respective  $z$ -values, compute means and differences of the resulting  $z$ -values, and then back-transform the mean or differenced  $z$ -values to obtain the resulting correlations.

Lastly, changes in persistence are quantified as the per cent changes in variance explained by the lag 10-day autocorrelation between the Future and Historical periods; that is,  $\frac{r_{10,Future}^2}{r_{10,Historical}^2} - 1$  where  $r_{10}^2$  denotes the variance explained by the lag 10-day autocorrelation. The per cent changes are calculated from the ensemble-mean autocorrelations mentioned above.

### Use of the autocorrelation to assess persistence

There are numerous ways to quantify the persistence of weather (for example, refs. <sup>57,58</sup>). We use the lag autocorrelation for three primary reasons. First, it is simple and requires few design choices as it involves correlating a time series with itself at another lag. There is no need to make subjective choices on the threshold or timescale of a temperature event, and the results can be easily reproduced. Second, the autocorrelation is calculated from all elements in a time series. Methods based on a predetermined threshold inevitably rely on a subjectively truncated data set. And third, the autocorrelation is directly related to the length of events on the wings of the frequency distribution; that is, it is directly related to the length of extreme warm and cold events.

Extended Data Figures 1 and 2 demonstrate the close relationship between the autocorrelation of temperature data and the average length of ‘warm events’, where the length of a warm event is defined as the number of consecutive days on which temperatures exceed one standard deviation. The conclusions are not influenced by the amplitude of the threshold, albeit the sample size becomes smaller as the threshold increases.

Extended Data Figure 1 shows the 2d density plots of the lag 1-day autocorrelation and the average length of warm events calculated at all grid boxes in the CESM1 historical output for four sample ensemble members. The average length of warm events increases monotonically with autocorrelation: time series with a larger autocorrelation are associated with longer warm events, and vice versa. As shown in Extended Data Fig. 2a, results calculated first for individual ensemble members and then averaged over all 40 ensemble members bear close resemblance to those derived from individual ensemble members.

The results shown in Extended Data Fig. 1 are derived from more than 10,000 days at more than 55,000 grid points, and the results shown in Extended Data Fig. 2a are further averaged over 40 ensemble members. The relationships are thus derived from a very large sample size and make clear the robust relationship between the autocorrelation and the length of warm events. The results are also consistent with the fundamental characteristics of a red-noise process. The grey dots in Extended Data Fig. 2b are generated by averaging the CESM1 ensemble-mean output from Extended Data Fig. 2a over bins that span 0.001 on the abscissa. The averaging serves to emphasize the centroid of the more than 55,000 data points used to generate Extended Data Fig. 2a. The black line shows the corresponding results derived not for CESM1 output but for red-noise time series generated as

$$x(t) = r(\Delta t)x(t - \Delta t) + (1 - r(\Delta t))^2 \varepsilon(t),$$

where  $\varepsilon(t)$  is randomly drawn from a standardized normal distribution and  $r(\Delta t)$  is the autocorrelation of  $x(t)$  at lag  $\Delta t$ . The specified autocorrelation  $r(\Delta t)$  is shown on the abscissa of Extended Data Fig. 2b. As evidenced in the results, the relationship between the autocorrelation and persistence of temperature events in the CESM1 output is consistent with the characteristics of a red-noise process.

### Statistical significance of ESM results

The stippling in Fig. 1 indicates results that are significant in the following two ways.

First, the sign of the results is the same in at least 75% of the ensemble members. If there is no change in temperature persistence (the null hypothesis), then there is an equal chance of either an increase or a decrease in the autocorrelation at a given grid point in a given ensemble member. The likelihood that results from at least 75% of all ensemble members will exhibit the same sign (either positive or negative) can be modelled using a binomial distribution with a probability of success of 0.5. In the case of 40 ensemble members:

$$P(x \geq 30) = \sum_{x=30}^{40} \binom{40}{x} (0.5)^x (0.5)^{40-x} \sim 0.1\%$$

Second, the differences in  $r_{10}^2$  averaged over all ensemble members exceed the 95% confidence threshold. The significance of the differences in means is assessed from a two-tailed test of the  $t$ -statistic

$$t = \frac{\bar{x}_F - \bar{x}_H}{\sqrt{\frac{\sigma_F^2}{N_F} + \frac{\sigma_H^2}{N_H}}},$$

where the subscripts F and H denote the Future and Historical periods, respectively; overbars denote the mean of  $r_{10}^2$  within each period;  $\sigma$  denotes the standard deviation of  $r_{10}^2$  within each period; and  $N$  denotes the number of independent values in each period, assumed to be

## Article

number of ensemble members (for example,  $N_F = N_H = 40$  if there are 40 ensemble members). The results are controlled for the false discovery rate<sup>59</sup>.

The stippling in Fig. 3 indicates results that are significant in the following two ways.

First, the multi-model mean changes are significant at the 95% level based on a two-tailed test of the  $t$ -statistic.

Second, the signs of the changes are the same in ensemble means from at least three out of the four ESMs.

Note that the results derived from the ESMs are based on a very large sample size. The results in Figs. 1 and 3 are derived from more than 10,000 days and averaged over 30–50 ensemble members (Fig. 1) and 150 ensemble members (Fig. 3). The results in Fig. 2 are further averaged over many grid points.

### Robustness to different lags

Figures 1–3 show results for the lag 10-day autocorrelation. Extended Data Fig. 3 highlights results from the CESM1 output for other lags, and Extended Data Fig. 4 highlights results from the CESM1 output for the total differences rather than percentage differences. Note that the total differences yield the same spatial patterns as the per cent differences, but by construction have larger amplitudes where the climatological-mean persistence  $r_{10, \text{Historical}}^2$  is largest, such as the tropics.

For the most part, the patterns of the differences are highly stable to the choice of the time lag used in the analysis. In the case of the results shown in Extended Data Figure 4, the spatial correlation between maps derived for results at lag  $i$  and lag  $i+1$  increases from  $r = 0.95$  when  $i=1$  to  $r = 0.995$  when  $i > 5$  (Extended Data Fig. 5a). Likewise, the spatial correlations between the lag 10-day map and maps calculated for results at all other lags exceeds  $r = 0.9$  for all lags between  $i=1$  and  $i=30$  (Extended Data Fig. 5b). In other words, the results shown in Figs. 1–3 are largely insensitive to changes in the time lag used to estimate the autocorrelation.

The most notable differences between results derived for different lags in Extended Data Figs. 3 and 4 are found in association with results from the lag 5-day map. For example, the lag 5-day map indicates negligible changes in persistence over the Sahara, whereas the lag 10-, 15- and 20-day maps indicate increasingly large changes. In the specific case of the Sahara, the discrepancies arise because the autocorrelation curves do not indicate clear separation until a lag of approximately 7 days (not shown). Note that even in the specific case of the Sahara, the lag autocorrelations are stable at lags longer than 10 days (Extended Data Figs. 3 and 4).

### Simpler models

The climate change simulations run with the gray radiation scheme and comprehensive radiation scheme shown in Fig. 4 were performed as part of a longwave hierarchy of numerical model configurations<sup>25</sup>. Details of the simulations are provided in ref. <sup>25</sup>. Details of the gray radiation scheme are provided in refs. <sup>25</sup> and <sup>60</sup>. Details of the correlated  $k$  method used in the RRTMG scheme are provided in ref. <sup>61</sup>.

The gray radiation and RRTMG simulations were run on the GFDL AM2.1 atmospheric general circulation model<sup>62</sup> at a uniform horizontal resolution of 2.8 degrees, with 48 vertical levels, and at a time step of 900 s. The model was run in an aquaplanet configuration and coupled to a slab-ocean model with a depth of 30 m (ref. <sup>63</sup>). There is no land surface, there is no ocean heat transport, there is no sea ice and there are no clouds. The simulations are forced with roughly equinoctial values of insolation (as per equation 3 in ref. <sup>25</sup>) and there is no diurnal or seasonal cycle. The simulations were spun-up for 10 years (gray radiation) and 20 years (RRTMG). The analyses shown here are based on 10 years of four times daily output.

The control and climate change experiments are forced as follows. In the case of the RRTMG configuration, the control simulation is forced with CO<sub>2</sub> concentrations of 355 ppmv, and the climate change simulation is forced with a fourfold increase in CO<sub>2</sub> concentrations relative to the control. In the case of the gray radiation configuration,

the simulations are forced with prescribed optical depths that are configured so that the long-term mean sea surface temperature fields in the gray radiation control and climate change simulations closely match those in the corresponding RRTMG simulations. In all simulations, the ocean mixed-layer depth is fixed at 30 m.

The climate change simulations run in RCE and shown in Fig. 5 are from ref. <sup>27</sup>. The simulations were run on the atmospheric component of the MPI-ESM1.2 (ref. <sup>64</sup>) at horizontal resolution T31 and with 47 vertical levels. Insolation is spatially homogeneous with a global mean value of 340 W m<sup>-2</sup>. The Coriolis parameter is zero everywhere. The model was run in an aquaplanet configuration and—unlike models used in many RCE experiments—is coupled to a slab-ocean model with a depth of 25 m. There is no land or ocean heat transport. The model simulates cloud processes and employs a comprehensive radiation scheme (RRTMG). There are no extratropical dynamical processes as the Coriolis parameter is zero and there is no meridional structure in the insolation. The analyses shown here are based on 10<sup>4</sup> months of output. The control and climate change experiments are forced with spatially homogeneous CO<sub>2</sub> concentrations corresponding to preindustrial and four times preindustrial levels, respectively. Details of the simulations are provided in ref. <sup>27</sup>.

### Statistical significance of the simpler model results

The error bars on the autocorrelation curves in Figs. 4 and 5 are calculated as follows.

First, the correlations are transformed to their corresponding  $z$ -values using the Fischer  $z$ -transformation.

The 95% error bars on the  $z$ -values are then calculated, assuming the following number of samples: one sample for every 30 days in the extratropics in Fig. 4; one sample for every 180 days in the tropics in Fig. 4; one sample for every 24 months in Fig. 5; and one sample for every 1,000,000 km<sup>2</sup> (~500 spatial samples). For example, in the case of the 10<sup>4</sup> months of grid-point output used to generate Fig. 5, the procedure amounts to approximately  $2 \times 10^5$  samples. The inclusion of spatial resolution in the number of samples used in the  $z$ -value confidence intervals is justified, as the mean correlations averaged across all grid boxes (as used in Fig. 5) are effectively identical to the correlations found by concatenating time series at all grid boxes into a single time series.

Lastly, the  $z$ -values were back-transformed to provide 95% confidence intervals on the respective correlation coefficients.

Note that the confidence intervals are generally very small due to the very large sample sizes used in the analyses.

### Data availability

The large-ensemble output is publicly available via the Multi-Model Large Ensemble Archive (MMLEA) at the National Center for Atmospheric Research (<https://doi.org/10.1038/s41558-020-0731-2>). The output from the gray radiation and RRTMG simulations were provided by Zhihong Tan at the NOAA Geophysical Fluid Dynamics Laboratory; the output from the RCE simulations were provided by Gabor Drotos at the Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca, Spain. All data used to construct the figures are archived in Figshare (<https://doi.org/10.6084/m9.figshare.15078807.v1>). All other data that support the findings of the study are available from the corresponding author upon reasonable request.

### Code availability

Code that was used in this study is available from the corresponding author upon reasonable request.

51. Kay, J. E. et al. The Community Earth System Model (CESM) Large Ensemble Project: a community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.* **96**, 1333–1349 (2015).

52. Jeffrey, S. et al. Australia's CMIP5 submission using the CSIRO-Mk3.6 model. *Aust. Meteorol. Ocean* **63**, 1–13 (2013).
53. Kirchmeier-Young, M., Zwiers, F. W. & Gillett, N. P. Attribution of extreme events in arctic sea ice extent. *J. Climate* **30**, 553–571 (2017).
54. Rodgers, K. B., Lin, J. & Frolicher, T. L. Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences* **12**, 3301–3320 (2015).
55. Deser, C. et al. Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).
56. Phillips, A. S., Deser, C., Fasullo, J., Schneider, D. P. & Simpson, I. R. Assessing climate variability and change in model large ensembles: a user's guide to the "climate variability diagnostics package for large ensembles" version 1.0 (2020).
57. Fraedrich, K. and Blender, R. Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.* **90**, 108501 (2003).
58. Franzke, C. L. E. et al. The structure of climate variability across scales. *Rev. Geophys.* **58**, e2019RG000657 (2020).
59. Wilks, D. S. "The stippling shows statistically significant grid points": how research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.* **97**, 2263–2273 (2016).
60. Frierson, D. M. W., Held, I. M. & Zurita-Gotor, P. A gray-radiation aquaplanet moist gcm. Part I: static stability and eddy scale. *J. Atmos. Sci.* **63**, 2548–2566 (2006).
61. Iacono, M. J., Mlawer, E. J., Clough, S. A. & Morcrette, J.-J. Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR community climate model, CCM3. *J. Geophys. Res.* **105**, 14873–14890 (2000).
62. Delworth, T. L., Broccoli, A. J., Stouffer, R. J., Balaji, V. & Beesley, J. A. GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate* **19**, 643–667 (2006).
63. Kang, S. M., Held, I. M., Frierson, D. M. W. & Zhao, M. The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *J. Climate* **21**, 3521–3532 (2008).

64. Mauritsen, T. et al. Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO<sub>2</sub>. *JAMES* **11**, 998–1038 (2019).

**Acknowledgements** We thank A. Phillips at NCAR for assistance with the Climate Variability Diagnostics Package analyses of the MMLEA; M. Winton and K. Rodgers for discussion of the GFDL ESM2M output; Z. Tan for providing output from the gray radiation and RRTMG simulations, and comments on the text; G. Drotos for providing the output from the RCE simulations and comments on the text; and F. Lehner, A. Czaja and B. Medeiros for helpful discussions. The large-ensemble output was obtained from the MMLEA produced by the US CLIVAR Working Group on Large Ensembles. We also acknowledge high-performance computing support from Cheyenne (<https://doi.org/10.5065/D6RX99HX>) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation (NSF). All maps were made using the Proplot Matplotlib wrapper and map projections from the Cartopy Python package. J.L. and D.W.J.T. are supported by the NSF Climate and Large-Scale Dynamics program.

**Author contributions** D.W.J.T. led the writing. J.L. performed the analyses and produced the figures.

**Competing interests** The authors declare no competing interests.

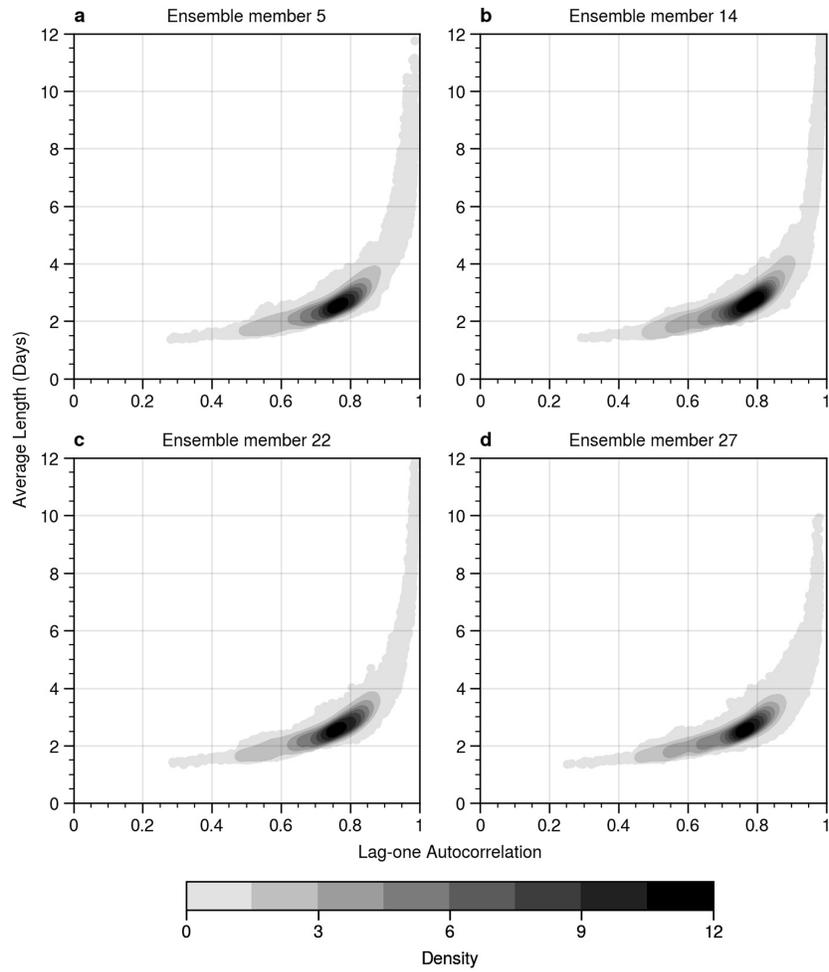
#### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03943-z>.

**Correspondence and requests for materials** should be addressed to David W. J. Thompson.

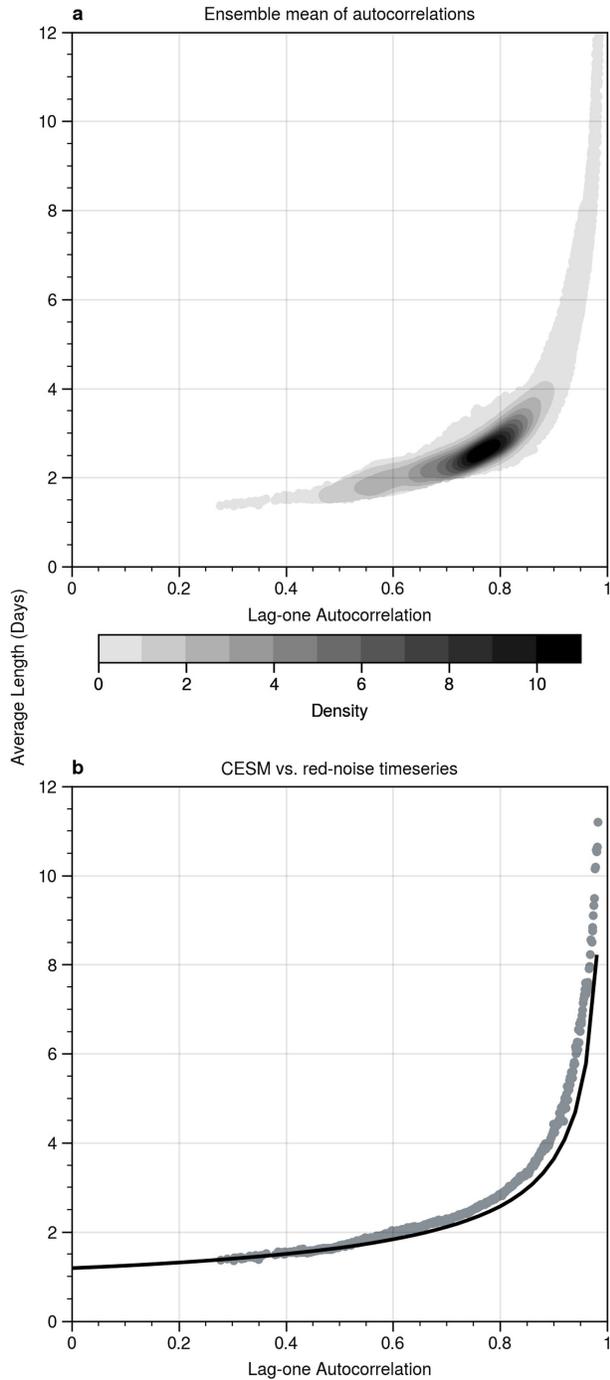
**Peer review information** *Nature* thanks Daniel Koll and Christian Franzke for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

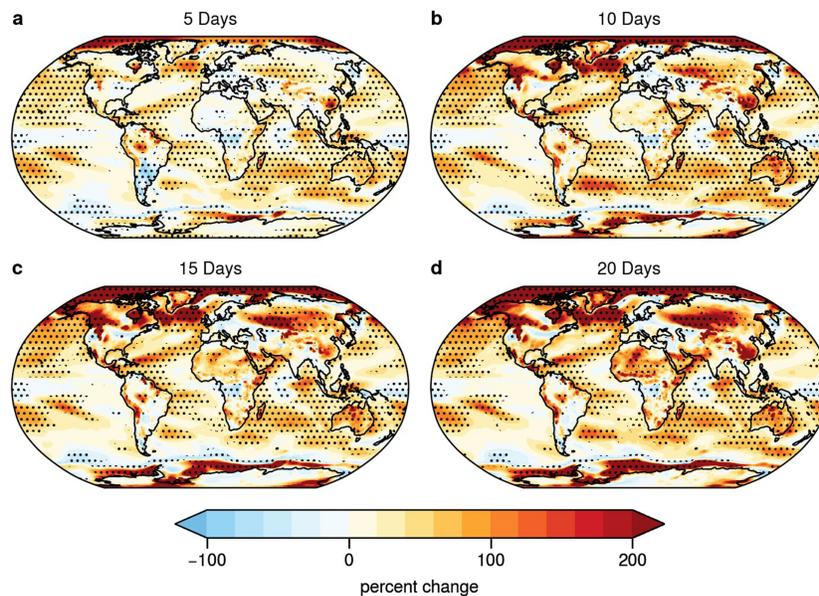


**Extended Data Fig. 1 | The relationship between the autocorrelation and the average length of a warm event.** The 2d density plot of the lag-one autocorrelation and the average length of warm events calculated as a function of grid box in the CESM1 historical output. Warm events are defined as periods

when temperatures exceed one standard deviation. Panels (a-d) show results for four sample ensemble members in the CESM1. Each panel includes results from 55296 grid boxes. Data density is found using a Gaussian kernel density estimate.

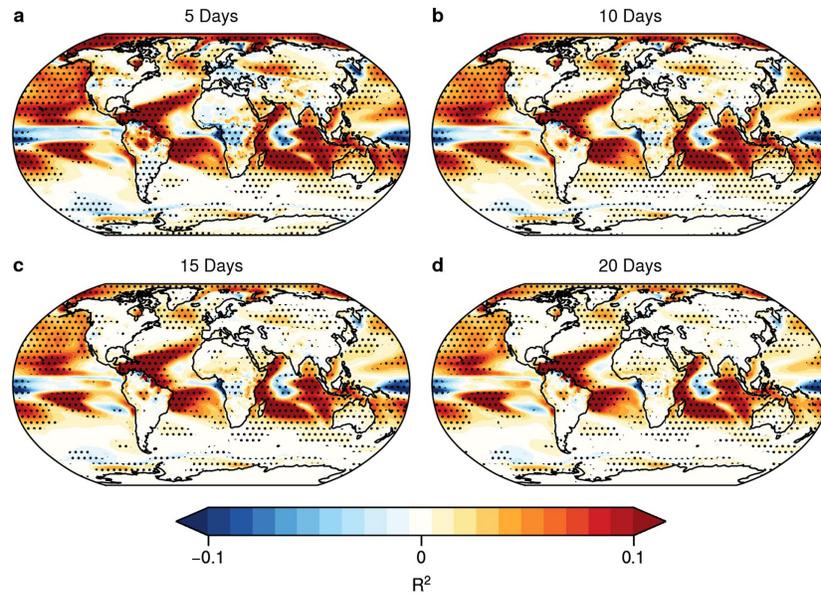


**Extended Data Fig. 2 | The ensemble-mean relationship between the autocorrelation and the average length of a warm event.** (a) The 2d density plot of the lag-one autocorrelation and the average length of warm events calculated as a function of grid box in the CESM1 historical output. Results are calculated for each ensemble member and then averaged over all ensemble members. Warm events are defined as periods when temperatures exceed one standard deviation. Each panel includes results from 55296 grid boxes. (b; shading) As in the top panel, but results are averaged over bins that span 0.001 on the abscissa. (b; black line) Results derived from randomly generated red-noise time series with autocorrelation specified on the abscissa. Data density is found using a Gaussian kernel density estimate.



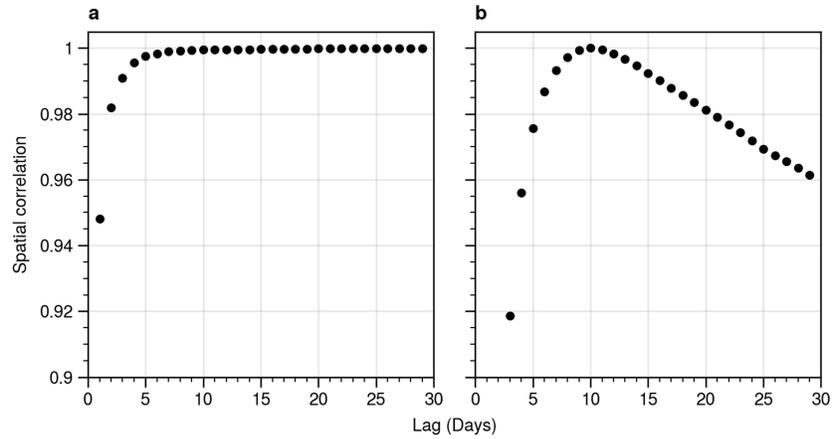
**Extended Data Fig. 3 | Changes in variance explained by persistence as a function of lag.** The changes in persistence between the “historical” period 1970-1999 and the “future” period 2070-2099 calculated from 40 large-ensembles run on the NCAR CESM1. Warm (red) colours represent an increase in persistence from the Historical to Future periods, while cool (blue) colours represent a decrease in persistence over the same period. Results show the percent changes in the variance explained by the (a) lag 5, (b) lag 10, (c) lag 15, and (d) lag 20-day autocorrelations. That is, they show:  $\frac{r_{i,Future}^2}{r_{i,Historical}^2} - 1$  where

$r_i^2$  denotes the variance explained by the lag  $i$ -day autocorrelation. Note that the autocorrelations are calculated first for individual ensemble members and then averaged over all ensembles using the Fisher-z transformation. Stippling indicates grid points where at least 75% of the ensemble members agree on the sign of the change (a likelihood of  $< 0.1\%$  by chance) and where the ensemble mean results exceed the 95% confidence threshold based on a two-tailed test of the t-statistic. Note that panel (b) is identical to Figure 1a. See Methods for details of the ESM output, analysis, statistical significance, and reproducibility.



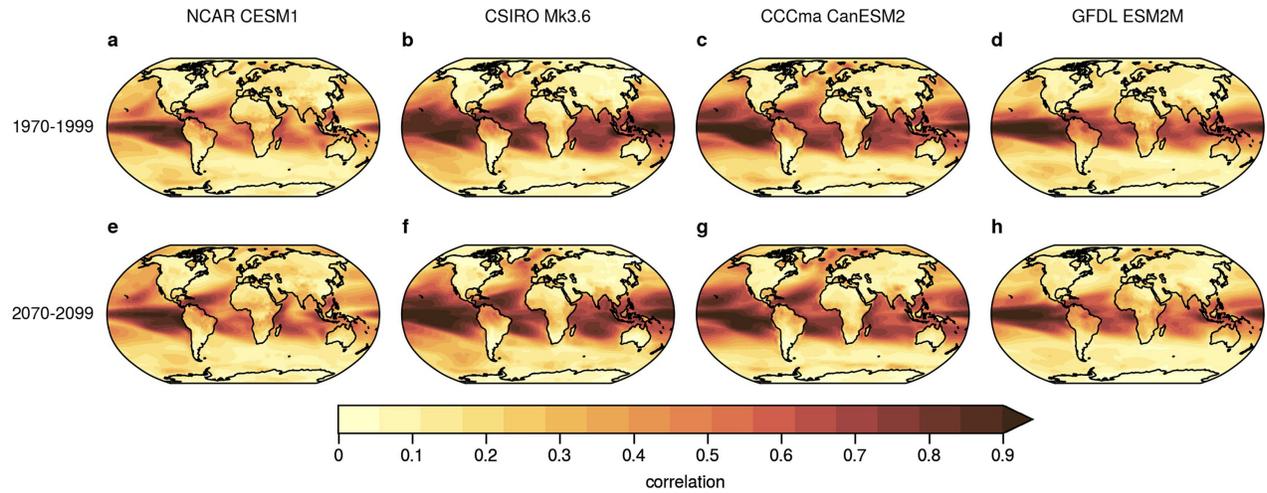
**Extended Data Fig. 4 | Changes in persistence as a function of lag.** The changes in persistence between the “historical” period 1970-1999 and the “future” period 2070-2099 calculated from 40 large-ensembles run on the NCAR CESM1. Warm (cool) colors represent an increase (decrease) in persistence from the historical to future period. Results show the *actual* changes in the variance explained by the (a) lag 5, (b) lag 10, (c) lag 15, and (d) lag 20-day autocorrelations, not the *percent* changes as shown in Extended Data Figure 3. That is, they show:  $r^2_{i,Future} - r^2_{i,Historical}$  where  $r^2_i$  denotes the variance

explained by the lag  $i$ -day autocorrelation. The autocorrelations are calculated first for individual ensemble members and then averaged over all ensembles using the Fisher-z transformation. Stippling indicates grid points where at least 75% of the ensemble members agree on the sign of the change (a likelihood of  $\sim 0.1\%$  by chance) and where the ensemble mean results exceed the 95% confidence threshold based on a two-tailed test of the t-statistic. See Methods for details of the ESM output, analysis, statistical significance, and reproducibility.



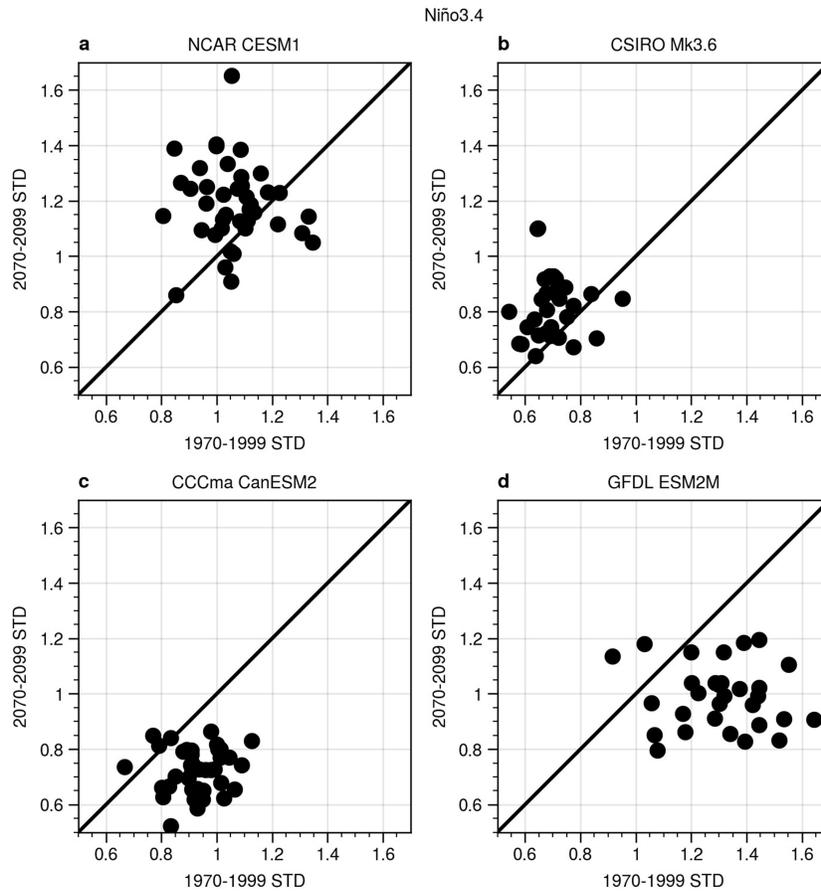
**Extended Data Fig. 5 | Testing the robustness of changes in persistence to lag.** (a) The results at lag  $i$  on the abscissa indicate the spatial correlation between 1) the spatial map formed as  $r^2_{i,Future} - r^2_{i,Historical}$ , where  $r^2_{i,Future}$  and  $r^2_{i,Historical}$  indicate the variance explained by the lag  $i$ -day autocorrelation in the Future and Historical periods, respectively, and the autocorrelations are

calculated first for individual ensemble members and then averaged over all ensembles (e.g., the lag  $i=10$  map is shown in Extended Data Figure 4b); and 2) the corresponding map calculated for lag  $i+1$ . (b) As in panel (a), but for the spatial correlations between 1) the map formed for lag  $i$  and 2) the map formed for lag  $i=10$ . Results are based on all members from the CESM1 output.



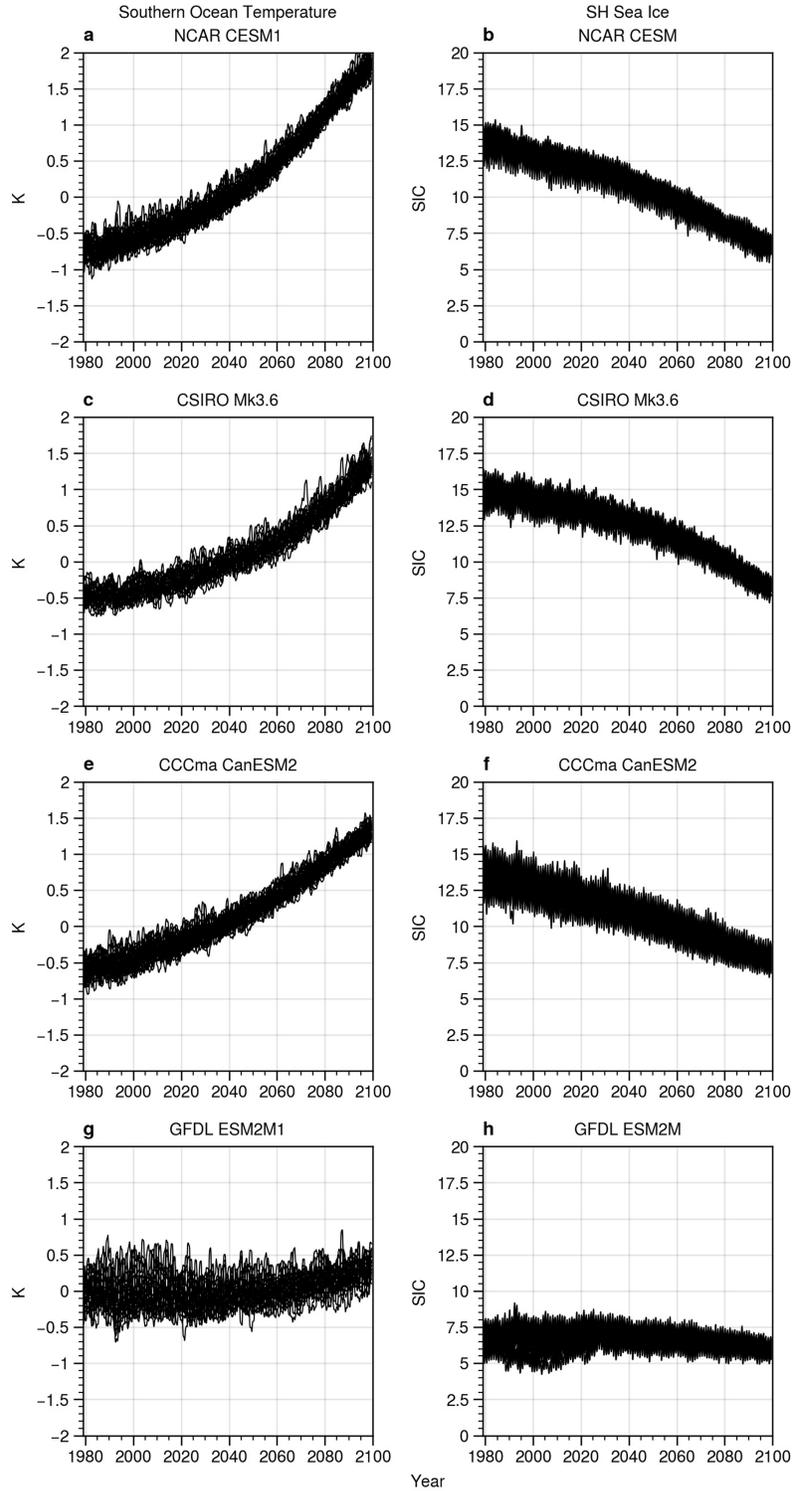
**Extended Data Fig. 6 | Climatological-mean autocorrelations of surface temperature in the historical and future periods.** The lag 10-day autocorrelations of surface temperature in large ensembles run on the four indicated ESMs for (top) the 1970-1999 historical period; (bottom) the

2070-2099 future period. The results are derived from (a, e) 40 ensemble members run on the NCAR CESM1, (b, f) 30 ensemble members run on the CSIRO Mk3.6, (c, g) 50 ensemble members run on the CCCma CanESM2, and (d, h) 30 members run on the GFDL ESM2M.



**Extended Data Fig. 7 | Assessing changes in ENSO in large ensembles run on four ESMS.** Scatter plots of the standard deviation of the monthly mean Niño 3.4 index during the historical period 1970-1999 and the future period 2070-2099 derived from (a) 40 ensemble members run on the NCAR CESM1, (b) 30

ensemble members run on the CSIRO Mk3.6, (c) 50 ensemble members run on the CCCma CanESM2, and (d) 30 members run on the GFDL ESM2M. The black diagonal lines represent the 1:1 line. Dots indicate results from individual ensemble members. The output was obtained from the NCAR CVDLP-LE.



**Extended Data Fig. 8 | Southern Ocean temperatures and SH sea ice extent in large ensembles run on four ESMs.** Monthly mean values of (left) Southern Ocean temperatures; (right) Southern Hemisphere sea ice extent in large ensembles from the indicated ESMs. Results are shown for individual ensemble members and smoothed for display purposes using a 13 month running mean.

Results are derived from (a, b) 40 ensemble members run on the NCAR CESM1, (c, d) 30 ensemble members run on the CSIRO Mk3.6, (e, f) 50 ensemble members run on the CCCma CanESM2, and (g, h) 30 members run on the GFDL ESM2M. The output was obtained from the NCAR CVDP-LE.