Higher trends but larger uncertainty and geographic variability in 21st century temperature and heat waves

Auroop R. Ganguly[a,1], Karsten Steinhaeuser[a,b], David J. Erickson III[c], Marcia Branstetter[c], Esther S. Parish[a], Nagendra Singh[a], John B. Drake[c], and Lawrence Buja[d]

[a]Geographic Information Science and Technology Group, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [b]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556; [c]Computational Earth Sciences Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831; and [d]National Center for Atmospheric Research, Boulder, CO 80305

Edited by Stephen H. Schneider, Stanford University, Stanford, CA, and approved July 31, 2009 (received for review April 23, 2009)

Generating credible climate change and extremes projections remains a high-priority challenge, especially since recent observed emissions are above the worst-case scenario. Bias and uncertainty analyses of ensemble simulations from a global earth systems model show increased warming and more intense heat waves combined with greater uncertainty and large regional variability in the 21st century. Global warming trends are statistically validated across ensembles and investigated at regional scales. Observed heat wave intensities in the current decade are larger than worst-case projections. Model projections are relatively insensitive to initial conditions, while uncertainty bounds obtained by comparison with recent observations are wider than ensemble ranges. Increased trends in temperature and heat waves, concurrent with larger uncertainty and variability, suggest greater urgency and complexity of adaptation or mitigation decisions.

climate change | extremes | regional analysis

Recent observations of global-average emissions (1, 2) show higher trajectories than the worst-case A1FI scenario reported in IPCC AR4 (3). Average A1FI temperatures (1, 4) trend higher than the best-case B1 as well as the relatively worse-case A2 scenario (5). Model simulations, validated with observations, have pointed to more intense, longer lasting, and more frequent heat waves in the 21st century (6). However, a rigorous statistical validation of the increased global warming and heat waves, followed by an investigation of the trends at regional scales, is required for decision-makers and end-users. Larger trends in warming and extremes suggest a greater urgency to develop adaptation and mitigation strategies (7, 8). On the other hand, a comprehensive assessment of the uncertainties and geographical variability provide an understanding of the tradeoff space for risk-informed decisions (9), which refers to different tactical or strategic options that may be available to a decision-maker for climate change adaptation and mitigation. Uncertainty of climate model projections has been quantified (10-14) either by comparing model hindcasts with observations or by comparing an ensemble of simulations. However, hindcasts validate models after the fact and hence risk underestimating predictive ability (15), while ensembles may only capture specific aspects of the variability. Hence the reliable and timely analysis of evolving climate model projections, extremes, and uncertainty remains a challenge (16-21).

Results

Statistically Higher Warming Trends. First, we show that the global-average temperatures from the middle to end of the 21st century are likely to be higher than previously believed (3). This is suggested by the fact that recent observed emissions trend toward or above A1FI assumptions (1, 2). The fact that observed emissions are at or above the level of A1FI, or any given scenario, in the current decade may not be a compelling reason to support conclusions about temperature in the late 21st century, as the trends could change considerably. However, when recent observations match or exceed the higher end of the emissions scenarios, then the latter cannot be ruled out as an implausible scenario. Moreover, we are not aware of any studies that clearly show that the higher temperature trends based on A1FI are statistically significant compared to other scenarios like A2 or B1. Here, A1FI simulations from CCSM 3.0 (22) are being evaluated. The assumptions inherent in the design of the A1FI and the A2 scenarios cause the A1FI emissions trajectories to be higher than A2 in the latter half of the 21st century; but while A2 continues to increase thereafter, A1FI begins to stabilize. The two scenarios converge toward the end of the century because of competing factors. Specifically, the A1FI envisions a more fossil-fuel intensive situation but also a more convergent world as compared to A2 (see reference 5 for details).

We performed a t-test ($\alpha = 0.05$) to determine if the mean A1FI outputs are higher than the mean A2 and B1 outputs at significant levels. Fig. 1 shows the global-average temperature projections, along with confidence bounds (three standard deviations on either side) at each decade. Our results (details in SI) show that both A1FI and A2 temperature projections are statistically distinguishable from the B1 scenario from 2040–2100 at 95% confidence; A1FI projections are statistically distinguishable from A2 from 2060–2090, but become indistinguishable again in 2100. During 2000–2007, when comparisons with observations are made, and until 2030, the B1, A2, and A1FI scenarios are statistically indistinguishable at 95% confidence. These statistical significance tests rely on important assumptions and uncertainty estimates (SI). The bottom-right panel of Fig. 1 shows monthly global-average temperatures (the "seasonality" in this case is caused by the distribution of land surfaces in the northern versus the southern hemispheres); visually, there is a clear match with reanalysis and observations, as well as an increasing trend in 2050 and 2100.

Significant Geographic Variability. Global averages are important (21), but a complete picture of projected trends and uncertainty emerges only when the results are analyzed geographically. Furthermore, stakeholders and end-users require credible as-

Author contributions: A.R.G. designed research; A.R.G. and K.S. performed research; D.J.E., M.B., J.B.D., and L.B. contributed new reagents/analytic tools; K.S., E.S.P., and N.S. analyzed data; and A.R.G., K.S., and D.J.E. wrote the paper.

The authors declare no conflicts of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: gangulyar@ornl.gov.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904495106/DCSupplemental.

ENVIRONMENTAL SCIENCES

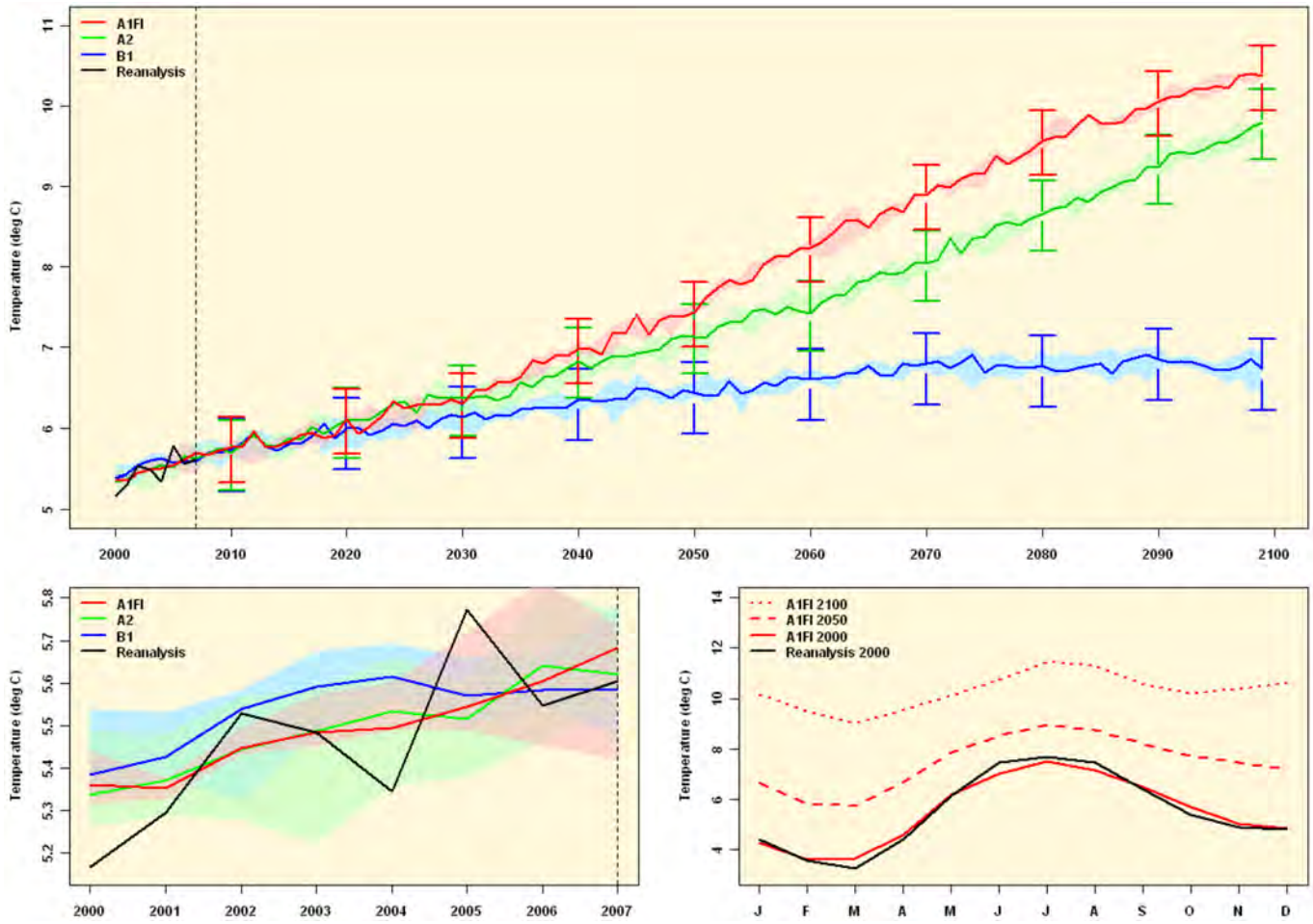

Fig. 1. Global average projections of temperatures and uncertainty. The top panel shows globally-average temperature (°C) projections from CCSM 3.0, based on A1FI, A2, and B1, along with error bars. The bias and standard deviations are calculated for each projection by comparing NCEP Reanalysis data with model outputs in 2000–2007, which forms the basis in the generation of the error bars for 2010 to 2100; note that the error bars are based solely on this bias and variance, but do not take into account the effect of projection lead times. The shaded areas indicate uncertainties caused by five initial-condition ensembles. The bottom left panel zooms in on 2000–2007. The bottom right panel shows monthly global-average temperatures.

sessments of climate change and extremes at local to regional scales for adaptation decisions and policy negotiations (7, 8). Here we compare A1FI-driven CCSM 3.0 (22) model projections and NCEP Reanalysis (23) observations for 2000–2007, when both are available, and develop grid-based estimates for model bias and standard deviation. We use the results to generate bias-corrected "most likely" projections and corresponding confidence bounds based on three standard deviations at each grid cell. Grid-based decadal averages are calculated for three time periods: The current "2000" decade, the mid-century "2050" decade, and the end-century "2100." The "2050" and "2100" values are subtracted from the "2000" values to show the change. A1FI temperatures start to significantly depart from B1 and A2 scenarios around 2050 and reconverge with A2 around 2100 (Fig. 1). The variables plotted in Figs. 2 and 3 are based on decadal average temperatures and Fig. 4 on a measure of the intensity of heat waves (6), all in degrees Celsius. While the maps have global coverage, the precision of the numbers are the same as CCSM 3.0 model outputs (1.4° × 1.4° grid), so the results can be used for regional analysis; visualizations were interpolated with commercial GIS software.

Decadal average temperatures from the model and observations, as well as the bias (details in SI), are shown in top panel of Fig. 2. While model and observations appear visually similar in 2000–2007, the geographical variability is significant. Temperatures over land appear to be mostly overpredicted (especially at high altitudes), with exceptions in higher latitudes, where the biases are low, and in desert areas, where there is some underprediction. Oceans are well-predicted due to thermal inertial effects, except for parts of the Atlantic around the current ice-edges that are underpredicted. The bottom two panels show temperature differences in 2050 and 2100 compared to 2000. The most likely temperature increases in 2050 are high almost all over the globe; however, the disparity between the upper and lower bounds is significant. The 2050 upper bound map looks similar to the 2100 most likely map, while the 2050 lower bound map is covered globally with small negative values. The geographic variability is large with relatively distinct regional patterns. The upper and lower bound maps in 2050, therefore, present two contrasting pictures of the globe. The 2100 maps show significant overall warming, even at the lower bound, while the upper bounds for both 2050 and 2100 show an increasingly grim state of the world. This is a unique method of examining trends in global climate prediction.

The geographical variability of temperature differences from the A1FI-based projections in 2000, 2050, and 2100 are compared with those from B1 and A2 in Fig. 3. Projections from the three scenarios are indistinguishable even at small significance levels in 2000, although A1FI appears slightly lower than both B1 and A2 over large portions of the globe. In 2050 and 2100, A1FI

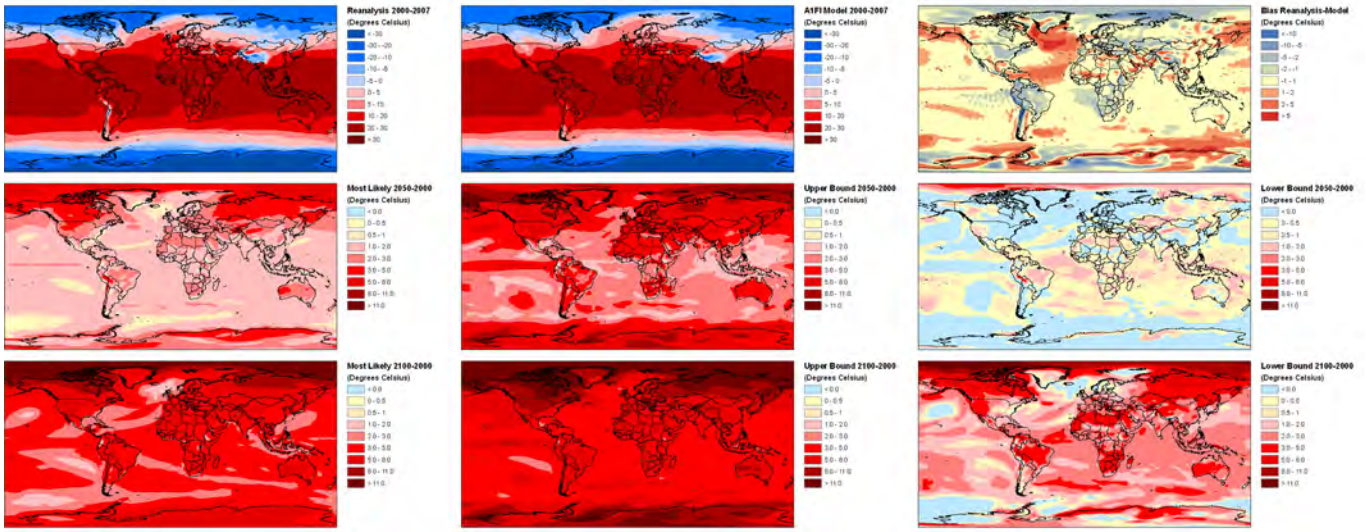

Fig. 2. Grid-based temperature projections with confidence bounds for A1FI. The top panel shows reanalysis and model-simulated annual average temperature (°C) along with the bias for 2000–2007. The bottom panels show 2050 and 2100 temperature projections from A1FI-forced CCSM 3.0 after bias correction (most likely maps, left) as well as upper (center) and lower (right) bounds. The numbers can be used to support local to regional scale analyses of climate change and extreme hydrometeorological stresses or impacts.

projections are much higher than B1 for almost all grid cells. The difference between the A1FI and A2 in 2050 is not statistically significant on average, but we find significant variability in scattered regions across the map. However, by 2100, A1FI is visually much higher than A2, especially in the northern hemisphere. To emphasize the uncertainty stemming from the geographic variability of the model, we also perform a comparison between individual ensemble members of the A1FI scenario (see SI).

Increasing Intensity-Duration-Frequency of Heat Waves. The intensity of heat waves across the world at CCSM 3.0 resolutions are investigated in Fig. 4. A heat wave here is defined as the mean annual 3-day warmest nighttime minima event, following previous researchers (6). The top two panels show that, while the model-generated and observed severity of heat waves in 2000–2007 exhibit similar patterns, there is a distinct bias. The observed heat wave intensity is consistently higher than A1FI projections in the current decade when global averages are considered. However, the large geographic variability of the observed and modeled heat wave intensities, as well as the significant uncertainties and geographic variability of the model biases, imply that this may not necessarily be true for all regions. Heat waves over land masses are nearly all overpredicted, with exception of the higher latitudes and certain desert regions, while the oceans are better predicted except in the Atlantic, where they are underpredicted. The two bottom panels show the projections for 2050 and 2100. Increased severity of heat waves is observed in nearly all of the land masses around the globe (except at high latitudes), as well as most of the low- to mid-latitude oceanic regions. The increases in heat wave intensity do not necessarily follow the warming patterns. Thus, the Western part of the United States show larger temperature increase, but the heat waves appear to be concentrated in the Midwest and Southeast. This clear difference in the geographical distribution of heat waves and simple warming is a critical point of this communication. We also investigate the duration and frequency of heat waves (see SI).

Discussion

We analyze NCEP Reanalysis observations and climate model simulations, including developed A1FI-forced ensembles generated from the CCSM 3.0 model, to develop bias-corrected

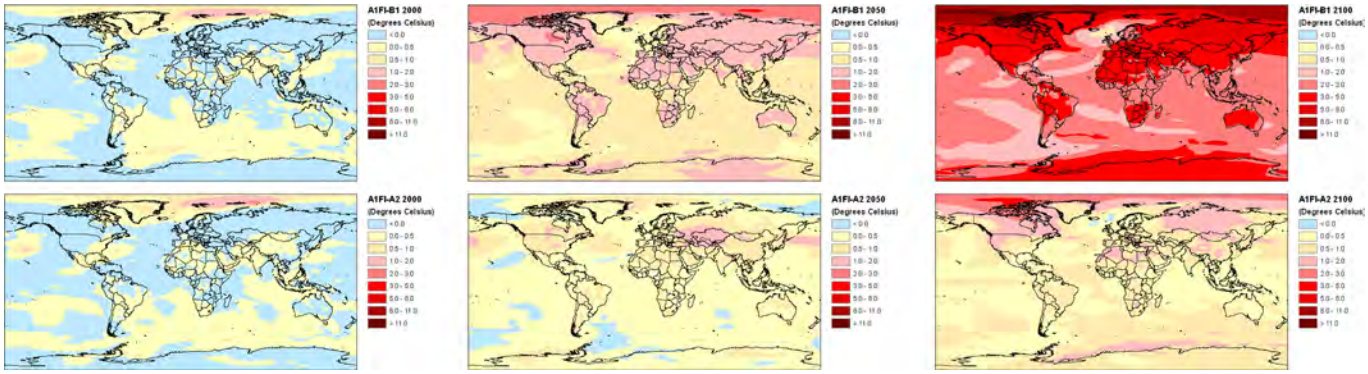

Fig. 3. Comparing grid-based A1FI projections with A2 and B1. Temperatures generated for the A1FI scenario in 2000, 2050, and 2100 are subtracted from the other scenarios: The figures on the left show 2000, during which time the scenarios are not separated. The figures on the right for 2050 and 2100 show the difference of A1FI with B1 (top) and A2 (bottom).

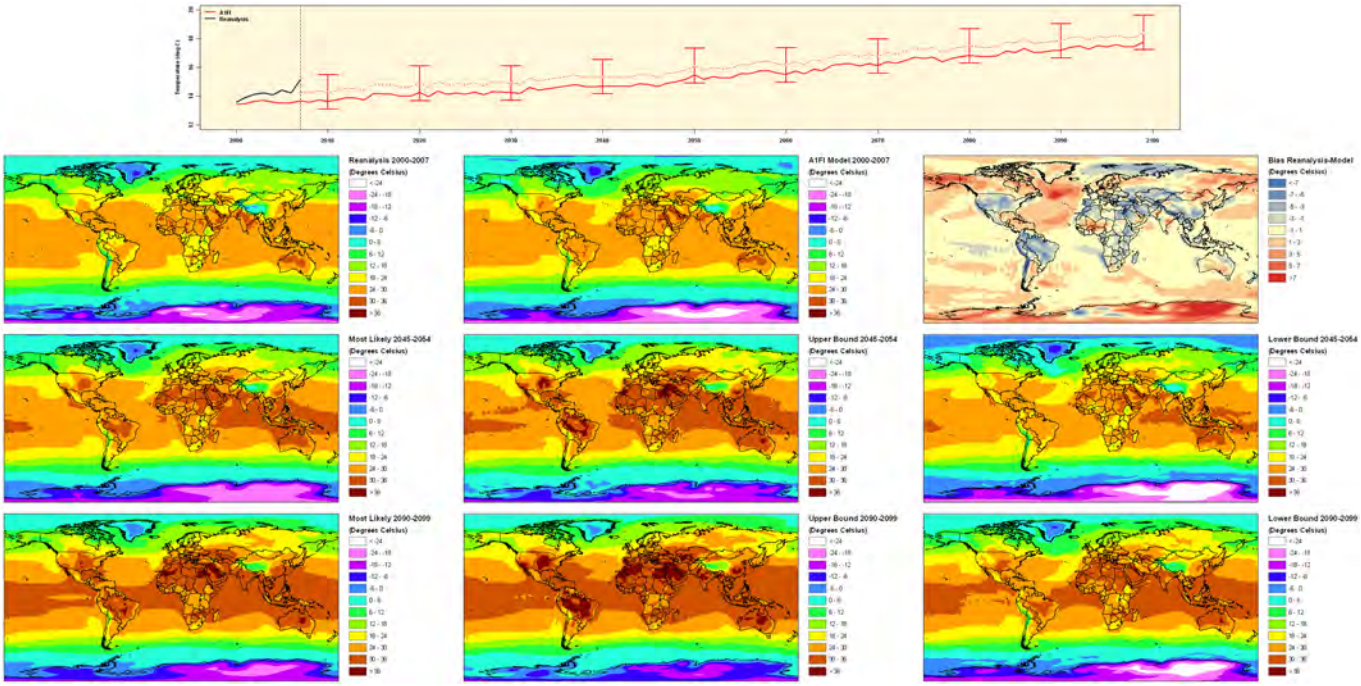

Fig. 4. Intensity of heat waves from A1FI. A heat wave is defined as the mean annual consecutive 3-day warmest nighttime minima event. The top two panels show intensity, graphically and mapped, from reanalysis data and model outputs for 2000–2007 along with the bias. The bottom panels show 2050 and 2100 heat wave projections from A1FI-forced CCSM 3.0 after bias correction (most likely maps, left) as well as upper (center) and lower (right) bounds. The numbers can be used to support local to regional scale analyses of climate change and extreme hydrometeorological stresses or impacts.

projections as well as uncertainty bounds for decadal-averages of temperature and heat waves. The uncertainty bounds are based on the differences between the model-simulations and observations, which are shown to follow a relatively stationary Gaussian distribution. The confidence bounds based on three standard deviations are consistently greater than the maximum ensemble ranges. In addition, these bounds can be larger than the differences between scenarios. Thus, the projected global-average temperatures from the different SRES scenarios cannot be statistically distinguished from each other at 95% confidence levels until about 2030, and the two more extreme scenarios (A2 and A1FI) cannot be distinguished from each other until the middle of the 21st century. This remains true even though the temperature changes are clearly distinguishable when compared across multiple decades for any one given scenario. However, the trends in A1FI-forced global-averaged temperatures are significantly higher from 2050 onwards, until they converge back with A2 toward the end of the century following the emissions trajectories. We note that the IPCC SRES scenarios are a suite of baseline nonpolicy scenarios that are not intended to span the full range of possible future emissions. A worst-case scenario could have higher emissions than A1FI, and a scenario including climate policy could have lower emissions than B1. A strict interpretation would identify A1FI as the "highest" scenario reported in IPCC AR4 and B1 as the "lowest," which is a relative labeling. Our use of the terms "worst case" and "best case," which borrowed from labeling (e.g., 24), may need to be interpreted accordingly. An examination of the regional variability based on daily data at 1.4 ° Gaussian grids reveals that the uncertainty bounds are large enough to make the warming appear insignificant on the lower bounds until 2050, but very significant at regional scales. Larger upper bounds imply that decision-makers need to be prepared for the worst possible consequences even though the most likely and lower bounds provide a way to optimize the allocation of potentially limited resources to manage the adverse effects. An investigation of decadal-average heat wave intensities at regional scales similarly reveals a large bias and uncertainty bounds. The globally averaged intensity of heat waves at decadal scales shows that the observed intensities are higher than the worst-case model projections in the current decade, which implies further exacerbation of heat waves compared to what has been already suggested by previous researchers. Future research needs to further validate the insights developed here through multimodel ensembles. The insights about trends in temperatures and heat waves, as a function of emissions trajectories, are expected to remain unaltered. However, the use of multiple models will likely increase the uncertainties and variability at both global and regional scales.

Materials and Methods

From the CCSM 3.0 model, we obtained five-member ensembles for IPCC SRES A2 and B1 and three runs for A1FI. We consider the ensemble median for visualization where applicable. The model data were provided at T85 resolution (approximately 1.4 ° × 1.4 ° grid) and NCEP/NCAR Reanalysis data at T62 resolution (approximately 2.5 ° × 2.5 ° grid). We use a bivariate spline (25) to interpolate the model data onto the reanalysis grid. Bias was computed for the 8-year period from 2000–2007, due to the need for both model and reanalysis data. The remainder of our analyses use three decades at the beginning, middle, and end of the 21st century: 2000–2009, 2045–2054, and 2090–2099; in our figures these are labeled as 2000, 2050, and 2100, respectively. All figures show decadal averages over each of these periods, in plots as global average and in maps computed individually at each grid location. For temperature extremes, we adopt a definition of heat waves that focuses on intensity of the event (6).

All statistics are performed using the software environment R (www.r-project.org) and the package akima (R package version 0.5–1; http://cran.r-project.org/web/packages). Maps were produced using commercial GIS software ArcGIS 9.3 (www.esri.com/software/arcgis).

ACKNOWLEDGMENTS. We thank Dr. Shih-Chieh Kao of Oak Ridge National Laboratory (ORNL) for his comments. This research was supported by the Laboratory-Directed Research and Development Program of the Oak Ridge National Laboratory, managed by UT Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

1. Raupach MR, et al. (2007) Global and regional drivers of accelerating CO_2 emissions. *Proc Natl Acad Sci USA* 104:10288–10293.
2. Marland G (2008) Uncertainties in accounting for CO_2 from fossil fuels. *J Ind Ecol* 12:136–139.
3. IPCC (2007) *Fourth Assessment Report: Climate Change 2007* (Cambridge University Press, Cambridge, UK).
4. Hadley Centre (2003) *Climate Change: Observations and Predictions* (Met Office, Exeter, UK).
5. Nakicenovic N, Swart R, Eds (2000) *Special Report on Emissions Scenarios* (Cambridge University Press, Cambridge, UK).
6. Meehl G, Tebaldi C (2004) More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* 305:994–997.
7. Stainforth DA, Allen MR, Tredger ER, Smith LA (2007) Confidence, uncertainty and decision-support relevance in climate predictions. *Philos Trans R Soc London Ser A* 365:2145–2161.
8. Stainforth DA, Downing TE, Washington R, Lopez A (2007) Issues in the interpretation of climate model ensembles to inform decisions. *Philos Trans R Soc London Ser A* 365:2163–2177.
9. Dessai S, O'Brien K, Hulme M (2007) Editorial: On uncertainty and climate change. *Global Environ Change* 17:1–3.
10. Lopez A, et al. (2006) Two approaches to quantifying uncertainty in global temperature changes. *J Clim* 19:4785–4796.
11. Barnett DN, Brown SJ, Murphy JM, Sexton DMH, Webb MJ (2006) Quantifying uncertainty in changes in extreme event frequency in response do doubled CO_2 using a large ensemble of GCM simulations. *Clim Dyn* 26:489–511.
12. Tebaldi C, Smith RL, Nychka D, Mearns LO (2005) Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensebles. *J Clim* 18:1524–1540.
13. Stott PA, Forest CE (2007) Ensemble climate predictions using climate models and observational constraints. *Philos Trans R Soc London Ser A* 365:2029–2052.
14. Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP and the breeding method. *Mon Weather Rev* 125:3297–3319.
15. Stainforth D, et al. (2005) Uncertainty in predictions of climate response to rising levels of greenhouse gases. *Nature* 433:403–406.
16. Stott PA, Kettleborough JA (2002) Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* 416:723–726.
17. Collins M, et al. (2006) Towards quantifying uncertainty in transient climate change. *Clim Dyn* 27:127–147.
18. Pielke RA (2008) Overheated claims. *Financial Post*, June 17.
19. Rahmstorf S (2007) Recent climate observations compared to projections. *Science* 316:709.
20. Berliner ML, Kim Y (2008) Bayesian design and analysis for superensemble-based climate forecasting. *J Clim* 21:1891–1910.
21. Watson J (2008) Certainty and uncertainty in climate change predictions: What use are climate models? *Environ Resour Econ* 39:37–44.
22. Drake JB, Jones PW, Carr GR (2005) Overview of the software design of the community climate system model. *Int J High Perform C* 19:177–186.
23. Kistler R, et al. (2001) The NCEP-NCAR 50 year reanalysis. *Bull Am Meteorol Soc* 82:247–268.
24. Tollefson J (2008) Climate war games. *Nature* 454:673.
25. Hiroshi A(1978) A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Trans Math Software* 4:148–159.

Supporting Information

Ganguly et al. 10.1073/pnas.0904495106

SI Text

Statistical Methods for Significance and Uncertainty Estimates. The assignment of confidence bounds to the model-based climate projections makes several underlying assumptions about the nature of the data, specifically, for the differences between reanalysis (''observations'') and climate model simulations:

The application of a *t*-test and the use of sigma levels to assign confidence bounds implicitly assume a normal (Gaussian) distribution. This assumption needs to be tested, especially over the time period 2000–2007, which in turn is used to compute the bias and standard deviation.

The computation of bias assumes that the mean of the normal distribution is stationary over time, while the computation of uncertainty assumes the same is true for variance (i.e., heteroskedasticity is ruled out). The stationary assumptions in the differences cannot be tested for projections, since reanalysis (''observed'') data are not available for the future. However, as a proxy, we can test whether the stationary assumptions hold for previous time steps when climate model hindcasts can indeed be compared with reanalysis data.

Bias and variance are assumed to remain constant with projection lead times. This assumption is difficult to test, since reanalysis data do not exist for the future and because climate models evolve over time. The latter implies that accurate estimates of the bias and variance of the most recent models as a function of lead times typically cannot be obtained from prior versions of climate models. However, normal distributions in decadal differences, a lack of trend in the 2000–2007 bias and variance, as well as nonstationary bias and variance in hindcasts, suggest that these assumptions may not be unreasonable. However, we conjecture that the variance will increase with lead times.

First, we test the assumption that the daily differences between the reanalysis data and the CCSM 3.0 model projections are in fact normally distributed, especially for 2000–2007. The correspondence of the histogram with the fitted normal distribution is visually examined. In addition, we produce the quantile-quantile (Q-Q) plots, which test for normality. As shown in the top left panel of Fig. S1, the histogram of the differences does indeed follow a Gaussian (with mean −0.018) visually, which is reflected in the Q-Q plot as well. The assumptions of normality are further tested by comparing hindcasts from climate model simulations with reanalysis data for each decade from 1950 onward. As shown in Fig. S1, the decadal differences follow the Gaussian quite well, both visually from the histograms as well as from the Q-Q plots.

The mean or standard deviation of the normal distributions in Fig. S1 does not appear to change significantly over six consecutive decades. We test the stationary assumptions further in Fig. S2. Trends in the mean and variance of the differences, along with the confidence bounds along the trend lines, are examined through regression analysis. The top, middle, and bottom panels (there are two plots per panel) show the trends in the mean and variance on the daily difference data after the application of an annual moving average, a decadal moving average, and a 25-year moving average. At 95% confidence levels, the trend lines do not appear to deviate much from the zero-slope regression lines in a statistically significant manner, even though slight trends may exist for the 25-year moving averages. The annual moving averages of the differences do not exhibit any linear trends whatsoever in the mean or the variance, but there are interannual features, which only rarely rise above the 95% confidence bands. The decadal moving averages appear to exhibit a slight upward trend in the mean and downward trend in the variance, which appears slightly enhanced in the 25-year moving averages. However, the deviations from a zero-slope line are relatively small. While decadal and multidecadal features are present in the differences, these are typically small compared to the 95% confidence bands. Thus, we can derive a few broad conclusions from Fig. S2. First, there are no major discernable trends in the mean or variance of the differences that fall beyond the 95% confidence bounds of the regression line. Second, the regression line is not significantly different from a corresponding zero-slope line, which implies stationary mean and variance. However, a slight trend (higher for mean and lower for variance) in the successive moving averages is noted. Third, weak interannual, decadal, and multidecadal features are noted in the differences. While not strong enough to overwhelm the bounds of the linear regression, these features may have physical interpretations. Thus, their relations with the interannual, decadal, and multidecadal sea surface temperature anomalies (SSTA) may need to be investigated, especially since climate models are known to be weak in modeling SSTA and their impacts on regional climate.

Fig. 1 shows global average projections of temperatures with uncertainty bounds. Here we present the complete *t*-test results between the three scenarios and indicate which pairings show separation at 95% confidence ($\alpha = 0.05$). As shown in Table S1, both the A1FI and A2 scenarios are significantly higher than B1 from 2040 onward, while A1FI does not separate from A2 until 2060 and then converges again by 2100.

Uncertainty in climate projections have typically been computed based on multimodel (12) and initial-condition ensembles (15), as well as by comparing model hindcasts with observations (13). Multimodel ensembles attempt to capture the uncertainty caused by our incomplete understanding of the physics and assume that a statistical averaging of the results from multiple models will be better than any one alone. However, since any physical understanding developed by the community is typically embedded within all models and the unknown physics are not captured in any model, the uncertainties may be underestimated from the multimodel approach. As suggested in Fig. 1 (bottom left panel, 2000–2007), the variability in the observations are rarely completely captured by the smoothed model outputs. Initial-condition ensembles capture only one, and perhaps very minor, aspect of the overall uncertainty. Fig. 1 visually depicts that bounds generated by the ensemble ranges for A2 and B1 are relatively small (compared to the bounds developed by comparing with the observations) and that these bounds do not fluctuate significantly over time, which indicates relative insensitivity to initial conditions for these specific projections at global-average scales. Furthermore, uncertainty bounds generated from comparisons with observations and shown as three standard deviation levels around the bias-corrected mean are much larger in all cases compared to the bounds generated from the ensembles based on the median value for the central tendency at any given point and the minimum and maximum as the bounds. We confirmed that the standard deviations are different and statistically distinguishable based on the F-test (see Table S2), especially since the observation-based bounds are always larger than the initial-condition ensemble-based bounds.

Uncertainties based on comparing model hindcasts with observations can provide important insights on the predictability and systematic errors of climate models. However, given that climate models continually improve based on incoming infor-

mation, the hindcasts may be closer to observations than future projections; hence the uncertainty assessments in projections may be underestimated. On the other hand, results from prior model versions originally run in projection mode may be compared with current observations. However, since models improve continually, such an approach may overestimate the current uncertainty. Here we assume that the IPCC SRES scenario definitions (at least conceptually) and the CCSM 3.0 model have not changed significantly since 2000. Thus, here we have been able to compare the 2000–2007 model results almost in a forecast (rather than a hindcast) mode with actual observations obtained from NCEP Reanalysis during that period for uncertainty bounds.

Owing to what are known as cascading uncertainties (the percolation of uncertainties from emissions to carbon and on to climate model projections, followed by regional climate and hydrological models all of the way to assessments of regional impacts), even a small gain in the characterization of uncertainty at the earlier stages of the cascade can lead to significant improvements in risk-informed decision-making related to adaptation and mitigation. Stakeholders and end-users require credible assessments of climate change and extremes at local to regional scales for adaptation decisions (e.g., management of natural water and nutritional resources, development of natural hazards preparedness and humanitarian assistance infrastructures, as well as dealing with climate related migrations) and policy negotiations (e.g., emissions regulations and agreements). The ultimate goal may be to reduce uncertainties by developing precise characterizations of the predictive ability of the current generation of climate models, pointing to areas where enhancements to physical understanding or modeling processes may be necessary, and suggesting pathways for climate model improvements in conjunction with statistical developments for uncertainty treatment.

Geographic Variability. Risk-informed decisions require cost-benefit analyses, for which regional-scale climate projections and impacts become important first steps. The geographic variability of climate change and extremes are important in this context. A recently concluded high-profile climate change war game brought these matters to the forefront (24).

In the main body of the paper, we already illustrated the geographic variability between different IPCC SRES scenarios. Here, we expand upon this discussion by examining variability within the same scenario by comparing the three individual A1FI ensemble members; this is an analysis for ensembles at a global scale. Fig. S3 shows the differences between ensemble outputs 2000 (top), 2050 (middle), and 2100 (bottom). The three columns represent the three possible pairwise comparisons between the ensemble members. First and foremost, we observe that significant spatial variability is indeed present, ranging from $-2.4\,^{\circ}$ to 1.6 °C across all pairings. In addition, note that magnitude and direction, as well as the geographic locality, of observed differences may vary over time. For instance, the third pairing has a large negative difference over much of Europe and northern Asia in 2000 (top right), but by 2100, this pattern has nearly inverted (bottom right), as the majority of the area now shows a moderate positive difference. These time-variant differences of up to 2.4 °C between ensemble members further increase the uncertainty associated with model projections, which should be taken into consideration in particular for regional- or local-scale analyses of future climate change and/or impacts.

Intensity, Duration, and Frequency of Heat Waves. We consider the severity of heat waves based on the A1FI scenario, and Fig. 4 illustrates their current intensity as well as projections including upper and lower bounds for 2050 and 2100. However, in characterizing heat waves, we are often interested in the trifecta of intensity, duration, and frequency for such events. Here we compare heat waves from CCSM 3.0 model outputs to observations and examine the latter two components, i.e., duration and frequency, omitted from the body text for space reasons.

Fig. S4 shows heat waves for observations from 2000–2007 (black) and for A1FI from 2000–2100 (red). The solid line denotes the actual model output, while the dotted line indicates the bias-corrected values with three-sigma error bounds. We note that even the worst-case A1FI scenario under-predicts intensity for the current decade by 0.65 °C on average, with an increase of 4.36 °C by the end of the 21st century. Hence, heat waves may become more intense than previously thought, and impact studies should give this possible scenario, once believed "too extreme," more serious consideration.

For duration and frequency, we adopt a different definition of a heat wave, namely one based on the probability of occurrence. More specifically, we select as threshold the 95th percentile of nighttime minima over the period 2000–2007, and any night exceeding this threshold is considered a heat wave. In Fig. S5, the left column corresponds to average duration (number of days) and the right column to the frequency (number of events per year). The top row represents the reanalysis data for 2000–2007, the next row the A1FI-forced model outputs for the same period, and the bottom two rows the model outputs for 2050 and 2100, respectively. First, we find that there is a fair amount of agreement between the observed and model-based values, both in terms of magnitude and spatial locality. Similar to intensity, we also observe significant increases in both duration and frequency of heat events in 2050, and even more so by 2100. It is important to point out that the decrease in frequency for 2100 (bottom right) along the equator does not represent a true reduction in heat events, but rather is a visual artifact caused by a single heat event lasting the entire year; that is, the nighttime low temperatures in 2100 will exceed the 95th percentile of present levels every single night. Combined with our analysis of intensity in the main body, these insights paint a grim picture for the end of the century if emissions continue to evolve along the trajectory defined by the A1FI scenario.

Computational Issues. Issues of computational complexity arise in a number of different areas related to this work. First, there is the immense complexity of the climate models themselves, which can strain even the most advanced supercomputing resources. In the case of CCSM 3.0, model runs proceed at ≈4.5 years of simulation per day of wall clock time running on 192 processors*—this is the major reason why a Monte Carlo approach using hundreds or thousands of model runs is simply not feasible. And while these numbers are already staggering, there is a continuous push to further reduce both spatial and temporal resolutions, but also to include additional processes that are currently being omitted to keep the computation tractable. In the case of CCSM 3.0, experiments are already being run with output at 6-h intervals, and improvements in the spatial resolution are only a matter of time.

Likewise, there are computational issues associated with the analysis of the data, both model outputs as well as observations. For example, merely calculating the heat waves from 100 years of daily A1FI outputs (20 GB of data) took 60 CPU h on a high-performance compute cluster. If we were to expand this type of analysis to the approximately 100 CCSM 3.0 output variables, it would require the equivalent of nearly 1 year of CPU time—not to mention the fact that most variables are output at each of 26 vertical layers. Moreover, we may be interested in more complex analytic techniques, e.g., the application of extreme value theory instead of an exceedance-over-threshold approach, which would increase the computational complexity of the problem manifold. For these reasons, consideration must

be given to computational requirements and constraints when performing any type of analysis on climate data.

Assumptions and Limitations. In this section, we state all known assumptions in our analysis and limitations of the available data and/or methods.

We assume a Gaussian distribution of differences between model outputs and observations and test the validity of this assumption (shown in *SI Text*). The temperature differences appear to follow the Gaussian quite closely, which is not unexpected. However, this assumption may not hold for other variables like precipitation.

Because bias and variance are stationary in hindcasts (shown in this *SI Text* under *Statistical Methods*), we assume the same will be true for projections as well. This assumption appears reasonable in this context and is not to be confused with the assumption about projection lead times below.

We implicitly assume the uncertainty remains stationary even as a function of projection lead time. While we would expect an increase, this is difficult to quantify as models evolve over time. The assumption is a strong one. However, this does not take away from our analysis of most likely values or our assertions about large uncertainty and geographic variability. If nonstationarity with projection lead times is considered, uncertainty and variability are expected to increase, lending further support to our assertion.

We assume the statistical methods used to compute confidence intervals are valid despite a small number of data points (8) in the sample, namely yearly differences for 2000–2007. This is a strong assumption in general, but given our tests for Gaussian and stationarity, probably not too unreasonable. This may be especially true when global averages are considered. One alternative would be to compute the intervals by comparing model hindcasts and observations. However, this alternative approach has shortcomings as explained in the main paper. First, the models implicitly take into account past observations, and hence a comparison with hindcast may tend to underestimate the uncertainty in projections. Second, hindcasts do not consider the uncertainty caused by the emission scenarios. On the other hand, comparing the most recent (in this case 2000–2007) model outputs to corresponding observations may be considered as close to online validation as may be achievable. Thus, the definitions of IPCC SRES scenarios have remained the same during this period, and the improvements in CCSM3 model specifically to account for observations from 2000–2007 have been minor if any.

Grid-based estimates are only as precise as the model ($1.4° \times 1.4°$ grid), but are interpolated for visualization so that overall spatial trends are more easily discernible. Geospatial visualization of climate data are an important research area and may need to cater to multiple stakeholders like the modeling science community and the consequence analysis community, as well as to policy and decision-makers.

NCEP Reanalysis data are taken as a proxy for observations, even though we are cognizant that these data are not actual ground measurements, but the product of a model applied to observed data from a variety of sources. This assumption may be reasonable for temperature but needs to be tested for variables like precipitation.

Our analysis relies on global averages in certain situations and grid-based analysis in others. However, we use the outputs from general circulation models (GCM) only. Future work needs to examine continental to regional scale analysis, as well as outputs of either regional climate models (RCM) or higher-resolution GCM, when available.

We only use simulations from a single GCM here (namely, CCSM 3.0) but a multimodel ensemble approach is certainly possible and could be used in future research. We conjecture that similar mean trends of global temperature would hold when other models are included, but uncertainty and geographic variability would be further increased.

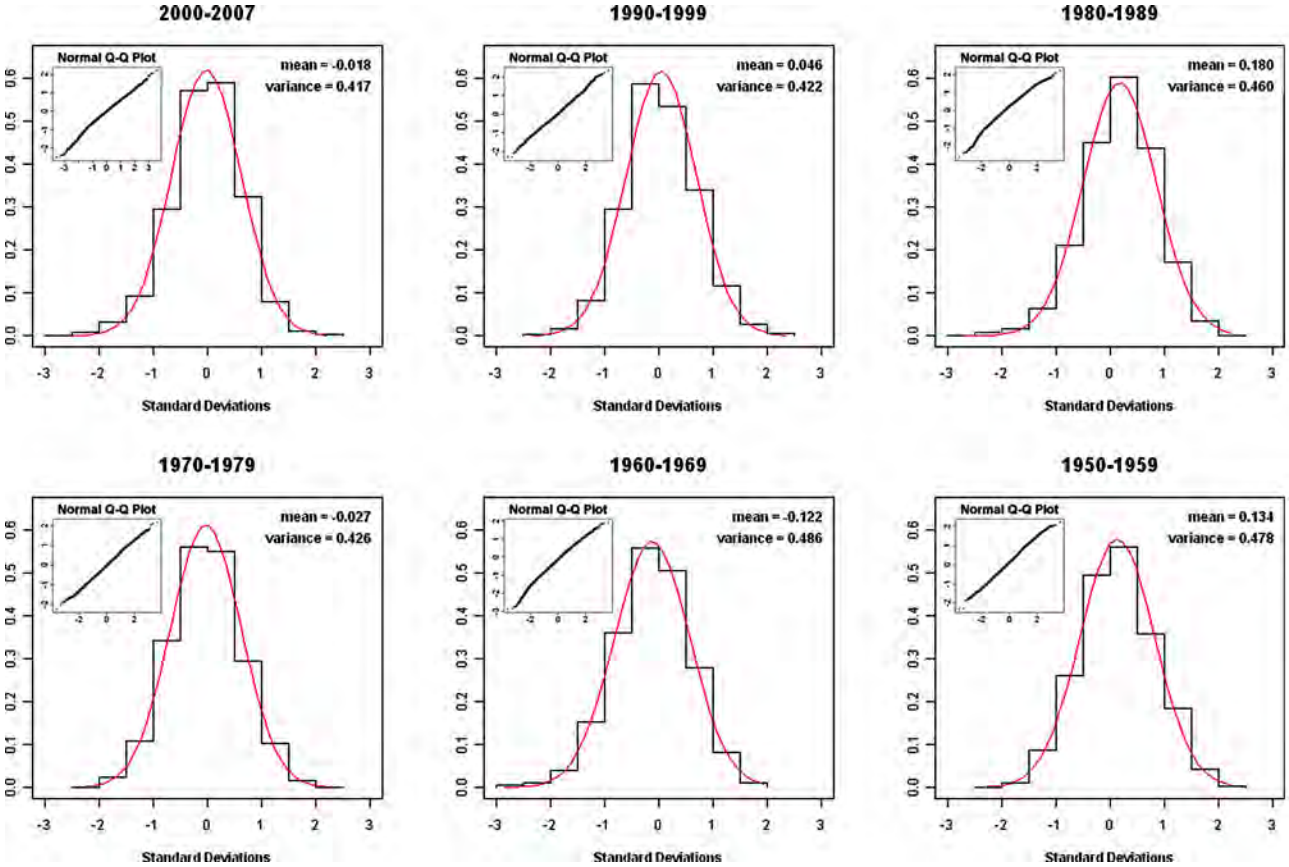

Fig. S1. Histograms of the model-reanalysis bias validate the normality assumption; best-fit curves are shown in red and Q-Q plots as insets.

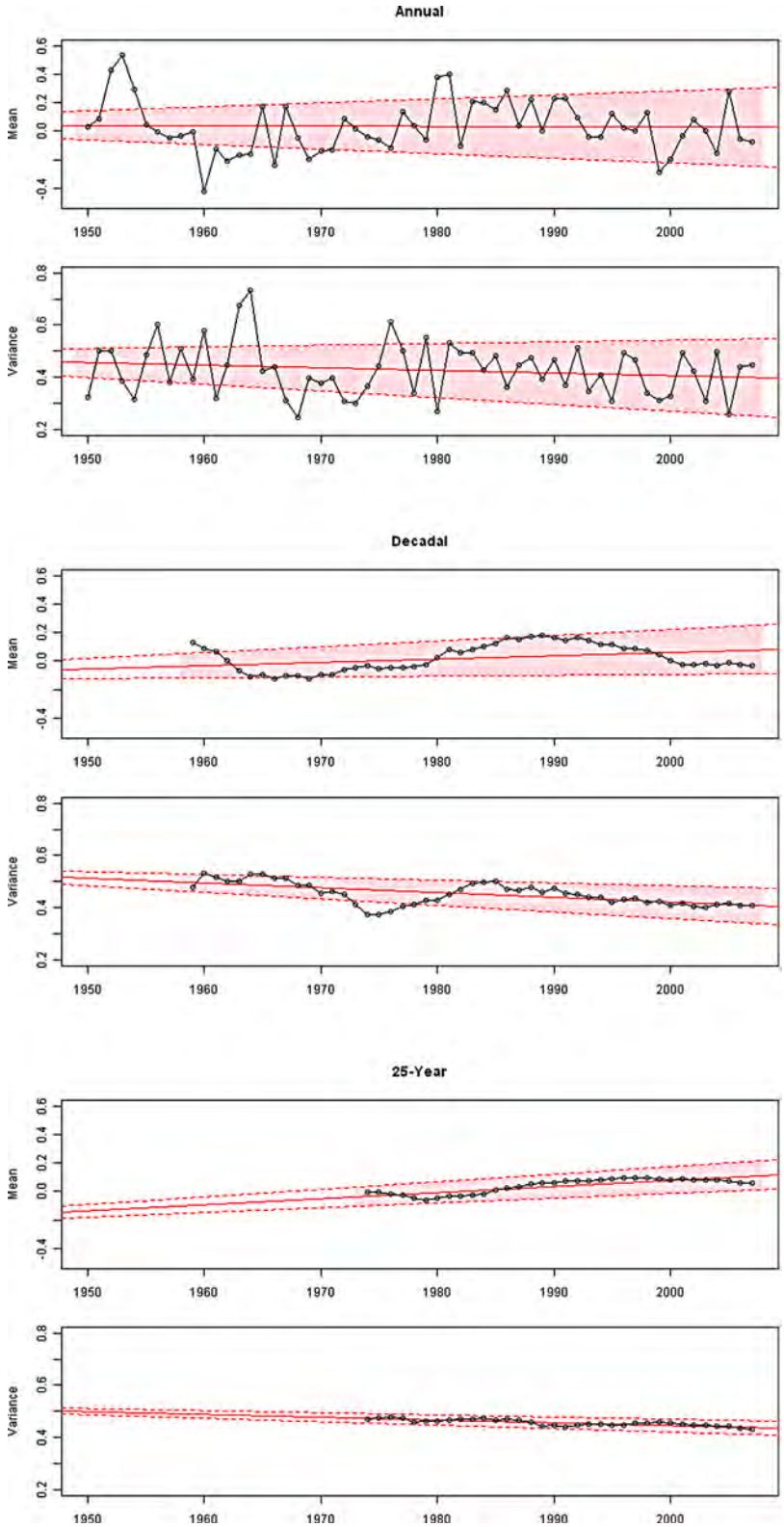

Fig. S2. Analysis of mean and variance of the differences in reanalysis and model simulations computed using 1950–2007 data.

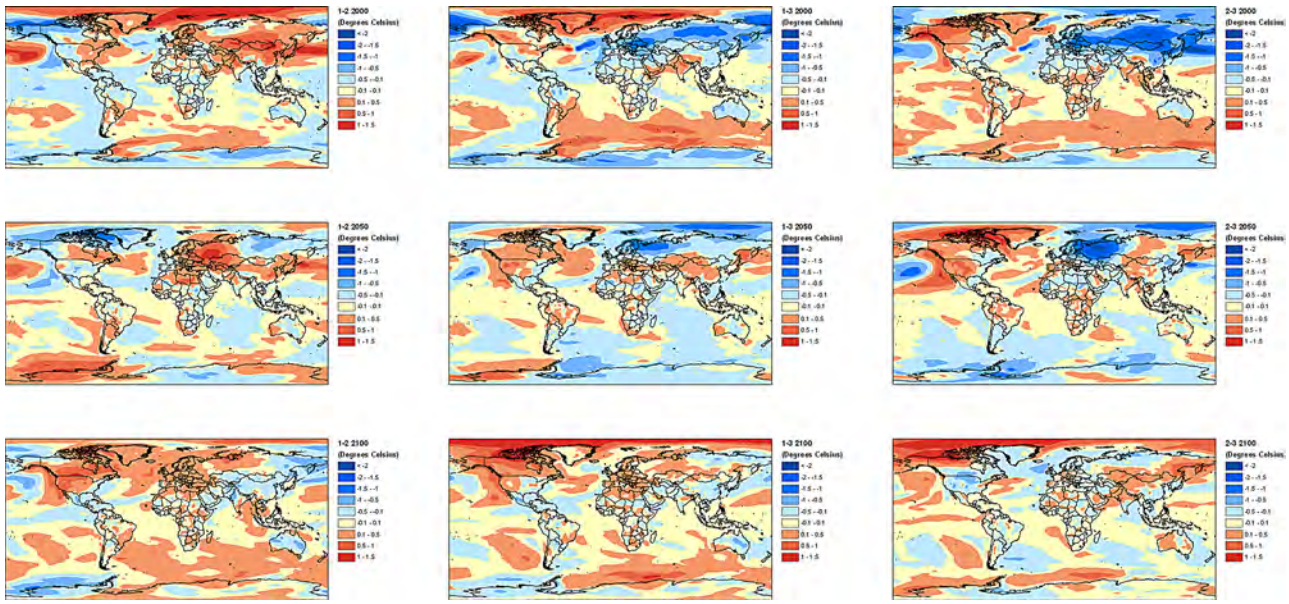

Fig. S3. Geographic variability between three ensembles of the A1FI scenario for 2000, 2050, and 2100.

Fig. S4. Globally-average heat wave (°C) projections from CCSM 3.0, based on A1FI, along with error bars. The bias and standard deviations are calculated for each projection by comparing NCEP Reanalysis data with model outputs in 2000–2007, which forms the basis in the generation of the error bars for 2010 to 2100.

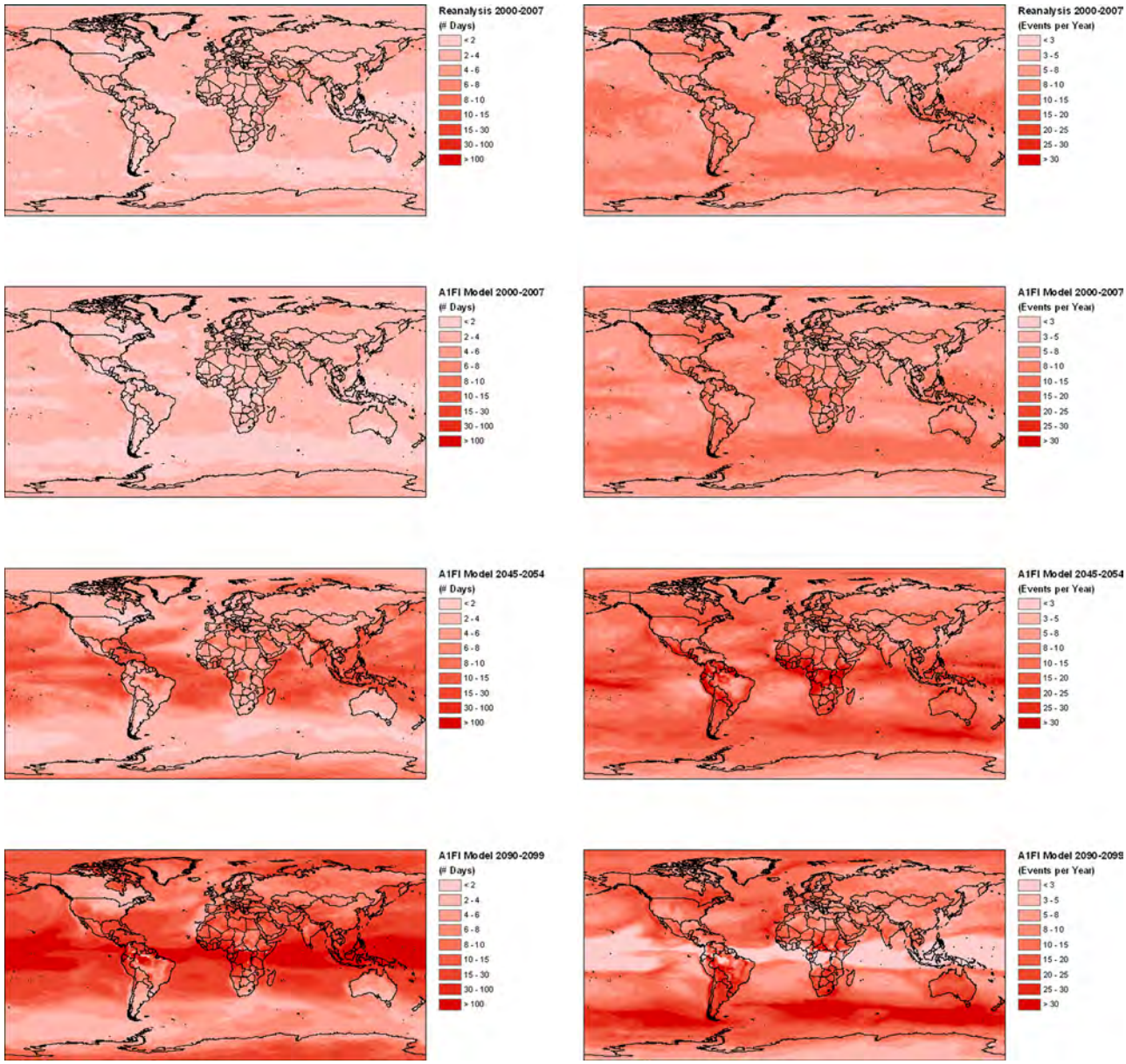

Fig. S5. Duration (*Left*) and frequency (*Right*) of heat waves from NCEP Reanalysis (top) and A1FI for 2000, 2050, and 2100.

Table S1. Pairwise *t* test results between the A1FI, A2, and B1 scenarios at 95% confidence ($\alpha = 0.05$)

		2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
A1FI-A2	*t*	0.385	0.295	0.059	0.599	1.034	1.781	2.018	1.851	2.041	1.371
	95%	No	No	No	No	No	Yes	Yes	Yes	Yes	No
A1FI-B1	*t*	0.485	0.900	0.840	2.978	4.646	6.397	8.575	10.970	13.303	8.640
	95%	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
A2-B1	*t*	0.182	0.393	0.945	1.952	2.873	3.334	5.111	7.807	9.821	12.667
	95%	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table S2. *f* test results for the A2 and B1 scenarios comparing uncertainty bounds to the ensemble bounds at 95% confidence ($\alpha = 0.05$)

		2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
A2	*f*	30.939	7.356	2.939	7.893	3.863	16.494	11.717	212.74	6.423	9.686
	95%	No	No	No	No	No	No	No	No	No	No
B1	*f*	46.012	5.932	5.975	20.988	28.977	23.393	37.963	9.939	8.870	5.106
	95%	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes