Statistically bias-corrected and downscaled climate models underestimate the severity of U.S. maize yield shocks

David Lafferty (davidcl2@illinois.edu)
University of Illinois at Urbana-Champaign https://orcid.org/0000-0003-2098-6885

Ryan Sriver
University of Illinois https://orcid.org/0000-0001-9587-3741

Iman Haqiqi
Purdue University https://orcid.org/0000-0001-7331-4615

Thomas Hertel
Purdue University

Klaus Keller
The Pennsylvania State University

Robert Nicholas
The Pennsylvania State University https://orcid.org/0000-0003-2615-2574

Article

Keywords: climate change, agriculture, climate models

Posted Date: February 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-157774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Communications Earth & Environment on September 20th, 2021. See the published version at https://doi.org/10.1038/s43247-021-00266-9.
Statistically bias-corrected and downscaled climate models underestimate the severity of U.S. maize yield shocks

Abstract

Efforts to understand and quantify how a changing climate can impact agriculture often rely on bias-corrected and downscaled climate information, making it important to quantify potential biases of this approach. Previous studies typically focus their uncertainty analyses on climatic variables and are silent on how these uncertainties propagate into human systems through their subsequent incorporation into econometric models. Here, we use a multi-model ensemble of statistically downscaled and bias-corrected climate models, as well as the corresponding CMIP5 parent models, to analyze uncertainty surrounding annual maize yield variability in the United States. We find that the CMIP5 models considerably overestimate historical yield variability while the bias-corrected and downscaled versions underestimate the largest historically observed yield shocks. We also find large differences in projected yields and other decision-relevant metrics throughout this century, leaving stakeholders with modeling choices that require navigating trade-offs in resolution, historical accuracy, and projection confidence.
Main Text

Introduction

Understanding and managing climate risk hinges on a quantitative description of the human and Earth system dynamics as well as their associated uncertainties\(^1\). The fidelity and utility of such analyses depend on using appropriate underlying climate information. While global climate models (GCMs) can provide useful insights at the global scale, they can face severe problems at regional to local scales. There are two main reasons for this. For one, the native resolution of GCMs is too coarse to be used in fine-scale analysis\(^2\). In addition, GCMs exhibit systematic biases relative to observations that require projections to be appropriately corrected and interpreted\(^3,4\).

As a result, bias-corrected and downscaled climate products are widely employed for a broad range of end-use applications, including diagnosing the impacts of climate change on the economy\(^5\), energy demand\(^6\), and human populations\(^7\); accounting for climate change in regional infrastructure planning\(^8,9\); and devising climate change adaptation strategies from the national level\(^10\) to city and urban areas\(^11\).

Although generally regarded as 'value-added' with respect to raw GCM outputs\(^12\), bias-corrected and downscaled climate products still contain considerable uncertainties. Examples include the validity of the stationarity assumption underpinning many bias-correction and downscaling methods\(^13,14\), the physical plausibility of results if methods are applied without expert knowledge\(^15\), the resulting representation of atmospheric or hydrologic variability\(^16,17\), and the applicability of methods in a changing climate\(^18,19\). It is important to understand how these uncertainties might propagate through econometric models and influence estimates of socioeconomic outcomes. Many previous studies in the climate and hydrology literatures focus on developing new bias-correction and downscaling techniques\(^20–22\) or evaluating the efficacy of existing ones\(^23–25\). However, few perform an 'end-to-end' analysis that incorporates sector-specific models of climate impacts\(^26\).

Here, we analyze how uncertainties from one statistical bias-correction and downscaling technique propagate into simulated agricultural outcomes by employing a parsimonious regression model of maize yields adapted from ref.\(^27\). In keeping with a substantial literature documenting the highly nonlinear effects of temperature on maize
yields\textsuperscript{27-30}, we account for the influence of temperature via growing degree days (GDDs), which are beneficial for yields, and extreme degree days (EDDs), which are indicative of heat stress and can severely reduce yields. Further relying on previous works\textsuperscript{31}, we also include a quadratic term in season-total precipitation. This model can reproduce yield records from the United States Department of Agriculture (USDA) with good accuracy (see Materials and Methods). We run the yield model across the continental United States (CONUS) using a multi-model ensemble of bias-corrected and downscaled climate models from the NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) dataset\textsuperscript{32}, as well as with the corresponding parent GCMs from the Coupled Model Intercomparison Project Phase 5, or CMIP5\textsuperscript{33}. By comparing yield hindcasts over the historical period (1960-2005) to yields simulated by an observational dataset, we quantify uncertainty due to bias-correction and downscaling as it pertains to crop yield outcomes. Using a high-emissions scenario, we then project yields using both ensembles throughout the remainder of this century and discuss notable differences and their implications. Our analysis highlights the importance of uncertainty propagation from global models through bias-correction and downscaling to its effect on simulated yield variability and other decision-relevant metrics.

Results and Discussions

Hindcast Evaluation of the National Yield Distribution. Simulations based on the original CMIP models considerably overestimate the variability of the observed CONUS yield distribution. Conversely, using the bias-corrected and downscaled ensemble (NEX-GDDP) underestimates the severity of the largest historically observed yield shocks. Figure 1.a shows the modeled national-level probability distribution of yield anomalies for each multi-model ensemble, constructed by aggregating county-level yields using yearly production weights. To facilitate a clearer diagnosis of the associated uncertainties, we use yields simulated by the observational climate product as ‘observed yields’, rather than actual USDA records. Further, we choose the same observational climate dataset that was used to bias-correct and downscale the original models. Figure 1.a shows that while the NEX-GDDP ensemble exhibits excellent agreement with the observed histogram over the central portion of the distribution, the CMIP ensemble gives a better representation of
the left or lower yield tail—particularly for the two most extreme shocks over the historical period.

From the probability distributions in Figure 1.a, we also compute sampling distributions of selected summary statistics and compare them with the observed values, shown in Figures 1.b-d. To test how well each ensemble captures the observed variability, with a particular emphasis on the tails of the distribution, we choose as summary statistics the standard deviation (Figure 1.b), median absolute deviation (Figure 1.c), and magnitude of the largest negative yield shock (Figure 1.d). To translate into a quantitative measure, the tail-area probability ($p_{tail}$) is calculated as the right-tail probability for each distribution, measured from the observed value. Figure 1.b shows that in only 17% of cases does the NEX-GDDP ensemble generate a sample with a standard deviation at least as large as the observed value. This rises to 68% for the median absolute deviation (Figure 1.c), a more robust measure of variability that is insensitive to outliers, which here are the largest yield shocks. By contrast, every sample generated by the CMIP ensemble exhibits variability, measured either by the standard deviation or median absolute deviation, larger than that observed.

Decision-making strategies that are robust against low-probability, high-impact events require information about the entire range of outcomes. Here, the largest negative yield shock (Figure 1.d) represents the most extreme event observed throughout the historical period. Consistent with the CMIP-based hindcasts overestimating the variability of the historical distribution, they also overestimate the severity of the largest shock, with 97% of samples exhibiting a greater magnitude than that observed. Conversely, the tendency of the NEX-GDDP ensemble to underestimate variability is exacerbated here—in only 5.7% of samples is the simulated shock as large as observed. Thus, while it is clear that bias-correction and downscaling lead to an overall improvement in the representation of national yield variability, this improvement is concentrated in the central portion of the distribution. For the tails, the net effect is an over-correction of the original large variability bias, resulting in an under-sampling of the most extreme weather-induced shocks.

**Hindcast Evaluation at the County Level.** To what extent are the national-level results replicated across individual counties in CONUS? We find that for many counties, particularly in and around the ‘Corn Belt’ region encompassing historically high-yield counties (Figure S1), conclusions from the national level hold. Figure 2 shows a map of
yield tail-area probabilities calculated from NEX-GDDP for each of the summary statistics in Figure 1. Small tail-area probabilities for the standard deviation (Figure 2.a) demonstrate that driving our yield model with the bias-corrected and downscaled ensemble underestimates the bulk variability of observed yields for many important maize-growing counties. As indicated by the 1% stippling, NEX-GDDP grossly underestimates the variability for several counties in Illinois and Iowa which are among the highest-production counties in CONUS (Figure S1). We observe that the median absolute deviation is underestimated for fewer counties (Figure 2.b) but the magnitude of the largest negative shock is underestimated for more counties, again indicating that the central part of the observed distribution is better captured than the tails.

The equivalent maps for the CMIP ensemble (Figure S2) reveal that the parent models systematically overestimate historical yield variability, displaying large tail-area probabilities for all summary statistics across a majority of maize-growing counties. Thus, we again conclude that bias-correction and downscaling lead to an under-sampling of the tails of the yield distribution. Further, we find that this holds across a large number of high-yield counties in CONUS, which is why this bias persists at the national level.

**Hindcast Evaluation of Climate Variables.** What is the source of this low-variability bias in the NEX-GDDP yield hindcasts? Recall that for yield observations, we use the results of driving the statistical yield model with the same observational product that was used in the bias-correction and downscaling. The yield model itself can hence be eliminated as a source of uncertainty. Instead, this bias is driven by the representation of each of the underlying climate variables in the model. Specifically, we find that the NEX-GDDP ensemble underestimates the variability of extreme degree days. It performs better with the more moderate temperature variable and the precipitation variable. Figure 3 shows the map of tail-area probabilities calculated from the NEX-GDDP ensemble for each climate variable across the eastern half of CONUS, with the standard deviation and median absolute deviation as summary statistics. Both temperature-derived variables show a predominantly latitudinal gradient in tail-area probabilities with variability being underestimated farther north and overestimated to the south. The NEX-GDDP ensemble shows a strong prevalence toward underestimating the variability of both variables throughout the Corn Belt region, which is the primary driver of the resulting low-variability bias in the corresponding maize yield hindcasts. However, as indicated by the 1%
stippling, the representation of extreme degree days, defined as an integrated measure of

temperature above 29°C, is considerably worse.

For season-total precipitation, there is less evidence of a clear bias that would
affect maize yields—the most prominent feature is a tendency to overestimate the
standard deviation across many central and southern counties. One potentially fruitful
avenue for future work is to investigate the effects of bias-correction and downscaling on
the representation of soil moisture, which is more important for plant growth\textsuperscript{35} and may or
may not be strongly correlated with cumulative precipitation depending on factors such as
runoff, drainage, and irrigation\textsuperscript{36}.

All tail-area probabilities generally cluster toward more moderate values when the
median absolute deviation is considered, as demonstrated by the reduction in 1%
stippling. Since the median absolute deviation is insensitive to outliers, this behavior is
consistent with the hypothesis that the NEX-GDDP ensemble tends to perform better at
capturing centrally located observations, now for each climate variable in addition to for
the yield distribution itself. However, for many of the most northern counties, NEX-GDDP
underestimates even the median absolute deviation for the extreme temperature variable.

As might be expected, the original CMIP models overestimate historical variability
for each climate variable for almost every county in the eastern half of CONUS (Figure
S3). The effects of bias-correction and downscaling hence show a clear spatial pattern.
For extreme degree days, the large variability bias is over-corrected for the most northern
counties with the coldest climates and the smallest historical variability (Figure S4). In
contrast, the large-variability bias is under-corrected for many southern counties with
warmer climates and larger historical variability, while the best performance is observed
for central counties. These patterns are also broadly exhibited for the more moderate
growing degree days. Thus, as a clear example of uncertainty propagation, the low
variability bias in NEX-GDDP simulated yields results from a tendency of this statistical
bias-correction and downscaling technique, for many important maize-growing counties in
CONUS, to over-correct the large-variability bias inherent to CMIP. We further highlight
that an analysis based on temperature itself would not uncover this NEX-GDDP bias,
which arises only for the degree day variables used in the yield model.

\textbf{Yield Projections.} Decision-makers care about the reliability of future climate projections,
for which adequate past performance is generally regarded as a necessary but insufficient
condition\textsuperscript{37}. One hypothesis, based on the tendency of the NEX-GDDP ensemble to underestimate historical yield variability, is that projections are similarly overconfident.

There are considerable differences between the yield projections of the NEX-GDDP and CMIP ensembles. Figure 4.a shows the projected national-level distributions for each multi-model ensemble for two 20-year periods during the 21st century: 2040-2059 (mid-century) and 2080-2099 (late century). We use the relatively high-emissions Representative Concentration Pathway (RCP) 8.5 scenario\textsuperscript{38} to emphasize the differences between the ensembles. One robust response is that all four projections show an increase in yield variability relative to historical observations (shown in gray). However, as with the historical hindcasts, the NEX-GDDP ensemble exhibits reduced variability relative to the original CMIP models. Although these projection differences alone do not demonstrate conclusively that the bias-corrected and downscaled models are overconfident (this could be done, for example, by using a perfect model experiment\textsuperscript{39}), they do illustrate the importance of end-users’ modeling choices, particularly in light of the results discussed in previous sections.

To illustrate how projection differences might affect decision-making, we consider how the different distributions affect estimates of return periods as a function of yield shock magnitude. Return periods give the average time between (often rare and damaging) events and are used often as part of risk analysis\textsuperscript{40}. In our case, we estimate the average number of years between large national-level yield drops driven by weather and climate. Figure 4.b shows that the calculated return periods from the CMIP and NEX-GDDP ensembles diverge for increasingly large yield shocks, particularly around the middle of the century. As an example, consider the magnitude of the largest national-level shock throughout the 46-year historical period, around 0.3. NEX-GDDP projects that an event this size will occur approximately every 16 years by the middle of the century; CMIP projects roughly every 7 years. By the end of the century, NEX-GDDP projects a shock this size approximately once every 7 years and for CMIP, roughly every 5 years. This presents a challenge for decision-makers in deciding which product to rely on, given that low-probability, tail events can have disproportionate negative impact. For example, it might pay to hold commodity stocks under NEX-GDDP but not CMIP. Alternatively, investing in adaptation measures might pass a cost-benefit analysis under CMIP, but not NEX-GDDP.

Caveats and Conclusions
Climate models are increasingly used to support regional impact analysis. Understanding, quantifying, and ultimately reducing the uncertainties and biases in climate impact projections is therefore crucial to improving decision-making. Bias-corrected and downscaled GCM output represents the state-of-the-art of climate modeling and many such datasets are used widely in the scientific literature as well as by decision-makers in public and private sectors. It is important to understand how uncertainty associated with methods of bias-correction and downscaling might propagate from the underlying climate information through models of human systems and influence key metrics of interest to stakeholders.

We analyze the effects of bias-correction and downscaling on modeled maize yields throughout the continental United States. Using a transparent, fast, and skillful statistical yield model and minimizing the influence of other factors, we find that the NEX-GDDP ensemble underestimates the largest historically observed negative yield shocks. We can attribute this effect to the process of bias-correction and downscaling of the original GCM outputs. By further evaluating the representation of each climate variable in the yield model, we also provide qualitative evidence that this underestimation of variability can be traced to temperature extremes in the form of extreme degree days. This uncertainty also has important implications regarding the use of projections—yield variability and associated risk metrics are considerably reduced in downscaled and bias-corrected projections relative to the parent models. This presents end-users with non-trivial decisions regarding which models to use. Choosing a climate projection dataset involves navigating potentially harsh trade-offs among resolution, historical performance of bulk metrics and extremes, and possible overconfidence.

Our study has important caveats that point to future research needs. For one, we consider projections based on a relatively simple statistical downscaling and bias-correction method. Our conclusions are hence of limited relevance to potentially more skillful bias-correction and downscaling techniques. Relative to the NEX-GDDP method, dynamic downscaling approaches are better suited for regions of complex terrain and more sophisticated statistical approaches may give a better representation of atmospheric variability. Second, our conclusions may not be valid for all econometric models that employ bias-corrected and downscaled climate information. It may be possible that structurally different yield models from the one employed here—for example, using average seasonal temperature rather than degree days—are less susceptible to biases in
the NEX-GDDP ensemble. Beyond the agricultural sector, econometric models will, in
general, be sufficiently different as to merit a separate investigation into whether biases
persist that are similar to those uncovered here. Given the widespread and
disproportionate impacts of extreme heat on human systems\textsuperscript{43}, these investigations would
be worthwhile. Third, we have treated each ensemble as a ‘model democracy’ where each
model output is weighted equally. There is a growing literature on more sophisticated
weighting schemes that account for model dependence and performance such that key
ensemble metrics are optimized\textsuperscript{44}. Such schemes could be applied here with the aim of
ensuring that yield hindcasts and projections are not overconfident, for example. Finally,
we assume no future adaptation in yield projections and do not account for increased
atmospheric carbon dioxide concentrations, which mitigate yield declines\textsuperscript{45}. Recall that
Figure 4 shows climate- and weather-driven yield deviations around the technological time
trend, meaning that even in years with large shocks, the absolute yield could still be higher
than today’s levels. Figure 4 should not be taken as a prediction but rather as a
demonstration of the differences that arise from bias-correction and downscaling when
subsequently applied to a historically realistic yield model.

Given these caveats, this work raises important questions about the use of bias-
corrected and downscaled climate information. Without careful consideration of the
representation of low-probability, high-impact events, the use of bias-corrected and
downscaled climate projections may lead to underestimates of the severity of impacts and
consequently, poor decisions. This is especially true for coupled human-environment
systems where extremes may be sensitive to feedbacks. In the agricultural sector, for
example, land-use change or irrigation practices can affect the local hydroclimatology\textsuperscript{46,47}. Finally, this work highlights the importance of taking a holistic approach and using metrics
specific to the sector of interest when evaluating the efficacy of bias-corrected and
downscaled climate products.

Methods

Climate Data. The NEX-GDDP dataset consists of statistically bias-corrected and
downscaled climate scenarios derived from 21 different CMIP5 models. One simulation
member from each CMIP5 model was used to derive historical hindcasts (1950-2005) and
future projections (2006-2100) of daily maximum and minimum temperature and mean
precipitation, with projections run under RCPs 4.5 and 8.5. From the varying spatial
resolutions of the parent CMIP5 models (see Table S1), all downscaled outputs have a constant lateral resolution of 0.25°. The observational dataset used was the Global Meteorological Forcing Dataset\textsuperscript{48}. The bias-correction and downscaling were performed via the Bias-Correction Spatial Disaggregation algorithm\textsuperscript{32}. Further details can be found in the SI Text. Academic studies that employ the NEX-GDDP dataset include refs.\textsuperscript{6,49–52}.

**Yield Model Construction and Evaluation.** The concept of degree days is a widely used measure for the amount of cumulative heat a crop is exposed to over the length of the growing season, which in this work is held constant at March through August. Degree days are derived for each grid cell by sinusoidally interpolating the diurnal temperature cycle using maximum and minimum daily temperatures (a detailed derivation can be found in ref.\textsuperscript{31}). GDDs are then calculated as the area under this temperature ($T$) curve subject to the bounds $10°C < T \leq 29°C$, summed over the growing season. EDDs are defined similarly, but with $T > 29°C$. These bounds are specific to maize\textsuperscript{27}. An area average is used to convert from the climate model or observational grid to the county level.

Modeling the yield sensitivity to moisture poses greater challenges. Complex daily interactions between soil moisture extremes and heat stress can lead to drastically different yield outcomes\textsuperscript{36}. As discussed, season-total precipitation is not always strongly correlated with seasonal mean soil moisture. Precipitation fields are also inherently noisier than temperature fields in climate reanalysis products, further complicating the extraction of the response function. Previous works rely on cumulative precipitation\textsuperscript{53,54}. To keep the yield model simple and fast, we include here only a quadratic season-total precipitation term.

Our resulting yield model is then:

$$\log Y_{i,t} = f_i(t) + c_i + \alpha_i \ GDD_{i,t}^i + \beta_i \ EDD_{i,t}^i + \gamma_i \ P_{i,t}^i + \kappa_i \ P_{i,t}^i + \epsilon_{i,t} \ .$$  \hspace{1cm} (1)

Here, $i$ denotes the spatial index (for this analysis, each county in CONUS), $t$ denotes the temporal index (each year), $Y$ is the yield in bushels/acre, $GDD$ denotes growing degree days, $EDD$ denotes extreme degree days, and $P$ denotes total in-season precipitation. A quadratic, county-specific time trend, $f_i(t)$, is included to account for technological progress and a constant term, $c_i$, is included to account for time-invariant group fixed effects such as soil quality. A prime indicates that the sample mean is removed. The
residual error is given by $\epsilon$ and assumed to be Gaussian, independent and identically distributed.

We use yearly county-level yield records to fit the coefficients in Eq. (1) using an ordinary least squares regression subject to the constraints $\alpha > 0, \beta < 0, \gamma > 0, \kappa < 0$. For each county, the group fixed effect $c_i$ is simply the mean of the de-trended yield time series, so the regression need only calculate the coefficients in front of the four climate variables.

Other modeling choices are possible here—our regression implementation is straightforward, and in constraining the coefficients to be the correct sign we rely on prior knowledge from previous works. Bayesian methods would allow a more sophisticated approach to represent prior knowledge. Additionally, the fit residuals are correlated in space (Figures S5, S6). More sophisticated treatments of the residual structure are possible\textsuperscript{55}, but we choose a simple model-fitting approach for parsimony.

Given these caveats, we fit this model using 1960-2005 as the training period and 2006-2016 for an out-of-sample validation, since 2005/2006 marks the hindcast/projection boundary in the climate model runs. Additionally, the fitting is performed only for counties that exhibit at least 50% data coverage in the USDA record over the training period. Model skill is assessed with the coefficient of determination, $R^2$, calculated without the inclusion of the time trend (i.e., it provides only a measure of how well the model captures the year-to-year weather-induced variations). We analyze spatial maps of $R^2$ at the county level as well as at the national level after aggregating by county production share, where production is defined as yield multiplied by harvested area. Additionally, we test whether Eq. (1) represents an improvement over only the quadratic time trend by using leave-one-out cross-validation (for a more detailed description, see the SI Text). In all analyses involving the modeled yields, we include only the counties where including the climate variables gives a lower mean squared error. This is to ensure that our conclusions are drawn from counties where including climate data in the model is meaningful. Equivalents of Figures 1, 2, and 4 that include all counties are shown in Figures S7, S8, and S9, respectively.

Our yield model is able to reproduce USDA-recorded yields with good accuracy (Figure S10). In the training period, the median $R^2$ for all 2,371 counties is 0.35; for the 1,816 counties where including the climate variables improved the fit, the median $R^2$ is 0.41 (these counties made up no less than 80% of total national-level production since 1960—see Figure S11). The equivalent out-of-sample values are 0.14 and 0.20,
respectively (Figure S12). The best fits are consistently found in and around the Corn Belt region, which is why the national-level time series of aggregated yields also exhibits good agreement with USDA records. The national-level $R^2$ for the training period is 0.49 and for the out-of-sample period is 0.71.

**Evaluation Metric.** In evaluating the climate model hindcasts, it is important to note that the time series of simulated yields are generally out of phase with observed yields on a year-by-year basis. As such, we choose an evaluation metric that is insensitive to the temporal dimension of modeled yields and instead focuses on long-term distributions. For each multi-model ensemble, we estimate the yield probability density function for each county in CONUS by constructing a kernel density estimate of the 966 (21 models x 46 years) point values. Then, $10^4$ samples of length equal to the observational data are drawn and for each summary statistic, a sampling distribution is compiled. These sampling distributions are displayed graphically for the national level in Figure 1. At the county level, we report the tail-area probability, calculated as

$$p_{\text{tail}} = Pr(S(\text{y}^{\text{ens}}) \geq S(y^{\text{obs}})).$$

Here, $S(y^{\text{obs}})$ is a given summary statistic of the observed distribution and the probability is understood to be taken over multiple realizations of the simulated distribution for each ensemble, $y^{\text{ens}}$. Thus, the tail-area probability is readily interpreted as the fraction of samples in which the simulated summary statistic was at least as large as the observed value. As a consequence of the structure of the yield model, summary statistics that measure the central tendency such as the mean and median are well represented by both ensembles, so we focus on measures relevant to variability and the tails.
Figure 1. (top) National-level distribution of yield anomalies, modeled by each multi-model ensemble: NEX-GDDP is shown in purple and CMIP in orange. The observational yield distribution is shown as the gray histogram. (bottom) Sampling distribution of each summary statistic: standard deviation (b), median absolute deviation (c), and magnitude of largest negative yield shock (d). NEX-GDDP is shown in purple and CMIP in orange; the single observed value is denoted by the black bar. Corresponding tail-area probabilities are also stated.
Figure 2. Tail-area probabilities of modeled maize yields for the NEX-GDDP ensemble measured against observationally driven yields. Results are shown for standard deviation (a), median absolute deviation (b), and the magnitude of the largest negative yield shock (c). Stippling indicates tail-area probabilities less than 0.01 or greater than 0.99.
Figure 3. Tail-area probabilities for each climate variable used in the yield model, calculated from the NEX-GDDP ensemble. The left column shows the results for standard deviation; the right column for median absolute deviation (MAD). Climate variables are organized by row: growing degree days (top), extreme degree days (middle), and precipitation (bottom). Stippling indicates tail-area probabilities less than 0.01 or greater than 0.99.
Figure 4. (a) RCP8.5 projections of U.S. yield variability for mid-century (2040-2059; full lines) and late century (2080-2099; dashed lines). The NEX-GDDP projections are shown in purple and the CMIP projections in orange; historical observations are also included for reference as the gray histogram. (b) Corresponding return periods for increasingly large magnitudes of negative yield shocks. In the absence of future yearly county-level production shares, we aggregate to the national level by using the mean 1960-2016 weights.
References

1. Change, Intergovernmental Panel on Climate. *Climate Change 2014 – Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report: Volume I: Global and Sectoral Aspects*. vol. 1 (Cambridge University Press, 2014).

2. Huber, V. *et al.* Climate impact research: beyond patchwork. *Earth Syst Dynam* 5, 399–408 (2014).

3. Sriver, R. L., Forest, C. E. & Keller, K. Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophys Res Lett* 42, 5468–5476 (2015).

4. Hawkins, E. & Sutton, R. The Potential to Narrow Uncertainty in Regional Climate Predictions. *B Am Meteorol Soc* 90, 1095–1107 (2009).

5. Hsiang, S. *et al.* Estimating economic damage from climate change in the United States. *Science* 356, 1362–1369 (2017).

6. Ruijven, B. J. van, Cian, E. D. & Wing, I. S. Amplification of future energy demand growth due to climate change. *Nat Commun* 10, 2762 (2019).

7. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J.-C. & Scheffer, M. Future of the human climate niche. *Proc National Acad Sci* 117, 11350 (2020).

8. Cook, L. M., Anderson, C. J. & Samaras, C. Framework for Incorporating Downscaled Climate Output into Existing Engineering Methods: Application to Precipitation Frequency Curves. *J Infrastruct Syst* 23, 04017027 (2017).

9. Anderson, J. *et al.* Progress on incorporating climate change into management of California’s water resources. *Climatic Change* 87, 91–108 (2008).

10. Martinich, J. & Crimmins, A. Climate damages and adaptation potential across diverse sectors of the United States. *Nat Clim Change* 9, 397–404 (2019).

11. Arnbjerg-Nielsen, K., Leonardsen, L. & Madsen, H. Evaluating adaptation options for urban flooding based on new high-end emission scenario regional climate model simulations. *Climate Res* 64, 73–84 (2015).

12. Feser, F., Rockel, B., Storch, H. von, Winterfeldt, J. & Zahn, M. Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples. *B Am Meteorol Soc* 92, 1181–1192 (2011).
13. Dixon, K. W. et al. Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change* **135**, 395–408 (2016).

14. Teutschbein, C. & Seibert, J. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol Earth Syst Sc* **17**, 5061--5077 (2013).

15. Maraun, D. et al. Towards process-informed bias correction of climate change simulations. *Nat Clim Change* **7**, 764–773 (2017).

16. Maraun, D. et al. Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* **48**, (2010).

17. Lanzante, J. R., Nath, M. J., Whitlock, C. E., Dixon, K. W. & Adams-Smith, D. Evaluation and improvement of tail behaviour in the cumulative distribution function transform downscaling method. *Int J Climatol* **39**, 2449–2460 (2019).

18. Lopez-Cantu, T., Prein, A. F. & Samaras, C. Uncertainties in Future U.S. Extreme Precipitation From Downscaled Climate Projections. *Geophys Res Lett* **47**, (2020).

19. Maraun, D. Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue. *J Climate* **26**, 2137–2143 (2013).

20. Walton, D. B., Sun, F., Hall, A. & Capps, S. A Hybrid Dynamical–Statistical Downscaling Technique. Part I: Development and Validation of the Technique. *J Climate* **28**, 4597–4617 (2015).

21. Hertig, E. & Jacobeit, J. A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area. *J Geophys Res Atmospheres* **118**, 520–533 (2013).

22. Sippel, S. et al. A novel bias correction methodology for climate impact simulations. *Earth Syst Dyn* **7**, 71–88 (2016).

23. Cannon, A. J., Sobie, S. R. & Murdock, T. Q. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J Climate* **28**, 6938–6959 (2015).

24. Eum, H.-I., Cannon, A. J. & Murdock, T. Q. Intercomparison of multiple statistical downscaling methods: multi-criteria model selection for South Korea. *Stoch Env Res Risk A* **31**, 683–703 (2017).
25. Teutschbein, C. & Seibert, J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology* **456–457**, 12–29 (2012).

26. Glotter, M. *et al.* Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proc National Acad Sci* **111**, 8776–8781 (2014).

27. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc National Acad Sci* **106**, 15594 (2009).

28. Schaubberger, B. *et al.* Consistent negative response of US crops to high temperatures in observations and crop models. *Nat Commun* **8**, 13931 (2017).

29. Lobell, D. B., Bänziger, M., Magorokosho, C. & Vivek, B. Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nat Clim Change* **1**, 42–45 (2011).

30. Lobell, D. B. *et al.* The critical role of extreme heat for maize production in the United States. *Nat Clim Change* **3**, 497–501 (2013).

31. D’Agostino, A. L. & Schlenker, W. Recent weather fluctuations and agricultural yields: implications for climate change. *Agr Econ* **47**, 159–171 (2016).

32. Thrasher, B., Maurer, E. P., McKellar, C. & Duffy, P. B. Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol Earth Syst Sci* **16**, 3309--3314 (2012).

33. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *B Am Meteorol Soc* **93**, 485–498 (2012).

34. Bernard, L., Semmler, W., Keller, K. & Nicholas, R. Improving Climate Projections to Better Inform Climate Risk Management. (2015) doi:10.1093/oxfordhb/9780199856978.013.0002.

35. Rigden, A. J., Mueller, N. D., Holbrook, N. M., Pillai, N. & Huybers, P. Combined influence of soil moisture and atmospheric evaporative demand is important for accurately predicting US maize yields. *Nat Food* **1**, 127–133 (2020).

36. Haqiqi, I., Grogan, D. S., Hertel, T. W. & Schlenker, W. Quantifying the Impacts of Compound Extremes on Agriculture and Irrigation Water Demand. *Hydrology Earth Syst Sci Discuss* **2020**, 1--52 (2020).

37. Knutti, R. Should we believe model predictions of future climate change? *Philosophical Transactions Royal Soc Math Phys Eng Sci* **366**, 4647–4664 (2008).
38. Vuuren, D. P. van et al. The representative concentration pathways: an overview. *Climatic Change* **109**, 5 (2011).

39. Herger, N. et al. Calibrating Climate Model Ensembles for Assessing Extremes in a Changing Climate. *J Geophys Res Atmospheres* **123**, 5988–6004 (2018).

40. Sriver, R. L., Lempert, R. J., Wikman-Svahn, P. & Keller, K. Characterizing uncertain sea-level rise projections to support investment decisions. *Plos One* **13**, e0190641 (2018).

41. Abatzoglou, J. T. & Brown, T. J. A comparison of statistical downscaling methods suited for wildfire applications. *Int J Climatol* **32**, 772–780 (2012).

42. Pierce, D. W., Cayan, D. R. & Thrasher, B. L. Statistical Downscaling Using Localized Constructed Analogs (LOCA)*. *J Hydrometeorol* **15**, 2558–2585 (2014).

43. Buzan, J. R. & Huber, M. Moist Heat Stress on a Hotter Earth. *Annu Rev Earth Pl Sc* **48**, 1–33 (2020).

44. Abramowitz, G. et al. ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth Syst Dynam* **10**, 91–105 (2019).

45. Moore, F. C., Baldos, U. L. C. & Hertel, T. Economic impacts of climate change on agriculture: a comparison of process-based and statistical yield models. *Environ Res Lett* **12**, 065008 (2017).

46. Eiras-Barca, J. et al. Changes in South American hydroclimate under projected Amazonian deforestation. *Ann Ny Acad Sci* **1472**, 104–122 (2020).

47. DeAngelis, A. et al. Evidence of enhanced precipitation due to irrigation over the Great Plains of the United States. *J Geophys Res* **115**, (2010).

48. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *J Climate* **19**, 3088–3111 (2006).

49. Missirian, A. & Schlenker, W. Asylum applications respond to temperature fluctuations. *Science* **358**, 1610–1614 (2017).

50. Ahmadalipour, A., Moradkhani, H. & Svoboda, M. Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble. *Int J Climatol* **37**, 2477–2491 (2017).

51. Thilakarathne, M. & Sridhar, V. Characterization of future drought conditions in the Lower Mekong River Basin. *Weather and Climate Extremes* **17**, 47–58 (2017).
52. Obradovich, N., Tingley, D. & Rahwan, I. Effects of environmental stressors on daily governance. *Proc National Acad Sci* **115**, 8710--8715 (2018).

53. Diffenbaugh, N. S., Hertel, T. W., Scherer, M. & Verma, M. Response of corn markets to climate volatility under alternative energy futures. *Nat Clim Change* **2**, 514--518 (2012).

54. Butler, E. E. & Huybers, P. Adaptation of US maize to temperature variations. *Nat Clim Change* **3**, 68--72 (2013).

55. Conley, T. G. GMM estimation with cross sectional dependence. *Journal of Econometrics* **92**, 1--45 (1999).
(top) National-level distribution of yield anomalies, modeled by each multi-model ensemble: NEX-GDDP is shown in purple and CMIP in orange. The observational yield distribution is shown as the gray histogram.

(bottom) Sampling distribution of each summary statistic: standard deviation (b), median absolute deviation (c), and magnitude of largest negative yield shock (d). NEX-GDDP is shown in purple and CMIP in orange; the single observed value is denoted by the black bar. Corresponding tail-area probabilities are also stated.
Tail-area probabilities of modeled maize yields for the NEX-GDDP ensemble measured against observationally driven yields. Results are shown for standard deviation (a), median absolute deviation (b), and the magnitude of the largest negative yield shock (c). Stippling indicates tail-area probabilities less than 0.01 or greater than 0.99.

Figure 2
Figure 3

Tail-area probabilities for each climate variable used in the yield model, calculated from the NEX-GDDP ensemble. The left column shows the results for standard deviation; the right column for median absolute deviation (MAD). Climate variables are organized by row: growing degree days (top), extreme degree days (middle), and precipitation (bottom). Stippling indicates tail-area probabilities less than 0.01 or greater than 0.99.
Figure 4

(a) RCP8.5 projections of U.S. yield variability for mid-century (2040-2059; full lines) and late century (2080-2099; dashed lines). The NEX-GDDP projections are shown in purple and the CMIP projections in orange; historical observations are also included for reference as the gray histogram. (b) Corresponding return periods for increasingly large magnitudes of negative yield shocks. In the absence of future yearly county-level production shares, we aggregate to the national level by using the mean 1960-2016 weights.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NCCSIMaizeUncertaintyBCSD.pdf