Statistically bias-corrected and downscaled climate models underestimate the adverse effects of extreme heat on U.S. maize yields

David C. Lafferty¹ , Ryan L. Sriver¹ , Iman Haqiqi² , Thomas W. Hertel², Klaus Keller^{3,4} & Robert E. Nicholas⁴

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. Here, we use a multi-model ensemble of statistically bias-corrected and downscaled climate models, as well as the corresponding parent models from the Coupled Model Intercomparison Project Phase 5 (CMIP5), to drive a statistical panel model of U.S. maize yields that incorporates season-wide measures of temperature and precipitation. We analyze uncertainty in annual yield hindcasts, finding that the CMIP5 models considerably overestimate historical yield variability while the bias-corrected and downscaled versions underestimate the largest weather-induced yield declines. 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.

¹Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ²Department of Agricultural Economics, Purdue University, West Lafayette, IN, USA. ³Department of Geosciences, The Pennsylvania State University, University Park, PA, USA. ⁴Earth and Environmental Science Institute, The Pennsylvania State University, University Park, PA, USA. ✉email: davidcl2@illinois.edu

Understanding and managing climate risk hinges on a quantitative description of the human and Earth system dynamics as well as their associated uncertainties¹. 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 analysis². In addition, GCMs exhibit systematic biases relative to observations that require projections to be appropriately corrected and interpreted³.

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 economy⁴, energy demand⁵, and human populations⁶; accounting for climate change in regional infrastructure planning^{7, 8}; and devising climate change adaptation strategies from the national level⁹ to the city and urban areas¹⁰.

Although generally regarded as adding value to the raw GCM outputs¹¹, 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^{12, 13}, the physical plausibility of results if methods are applied without expert knowledge¹⁴, the resulting representation of atmospheric or hydrologic variability^{15, 16}, and the applicability of methods in a changing climate^{17, 18}. It is important to understand how these uncertainties might propagate through impact 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^{19–21} or evaluating the efficacy of existing ones^{22–24}. However, fewer perform an end-to-end analysis that incorporates sector-specific models of climate impacts.

Here, we focus on agriculture. The agricultural sector has already been negatively impacted by climate change²⁵, and broad challenges are expected to continue throughout this century²⁶. From national policy to local farmer-level decision-making, reliable climate projections are crucial to identifying the risks that climate change poses to agriculture and developing subsequent adaptive measures.

Specifically, we analyze how uncertainties from one statistical bias-correction and downscaling technique propagate into simulated agricultural yields. We focus on county-level rainfed maize yields in the U.S. and employ a well-established statistical panel model²⁷. In keeping with a substantial literature^{27–30}, we account for the influence of temperature on yields via growing degree days (GDDs), which are beneficial for yields, and extreme degree days (EDDs), which are indicative of heat stress and severely reduce yields. Further relying on previous works³¹, we also include a quadratic in season-total precipitation. This model can reproduce yield records from the United States Department of Agriculture (USDA) with good accuracy (see the “Methods” section). We then drive the yield model with a multi-model ensemble of bias-corrected and downscaled climate models from the NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) dataset³², as well as with the corresponding parent GCMs from CMIP5³³. By comparing yield hindcasts over the historical period (1956–2005) 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. We corroborate the findings of previous works that highlight the usefulness of bias-correction for crop modeling^{34–38}. However, these studies focus primarily on process-based agricultural models and, if included, dynamic downscaling techniques. To our knowledge, ours is one of the first studies to investigate the effects of bias-correction and (statistical) downscaling on agricultural yields simulated by a statistical panel model.

Results

Hindcast evaluation of the national yield distribution. There are considerable differences in the yield outcomes simulated by each of the CMIP5 models, but most overestimate the variability of historical yields. Conversely, although the bias-corrected and downscaled models more closely match observations, they

Fig. 1 National-level evaluation of NEX-GDDP and CMIP5 maize yield hindcasts. Panel **a** shows historical (1956–2005) national-level log-yield distributions from each CMIP5 model (orange), NEX-GDDP model (purple), and the observational data (black). Panel **b** shows percentage differences between each individual NEX-GDDP model and the observational data, for our chosen summary statistics: the mean, standard deviation (SD), 10th percentile (q10), and minimum (Min.). The boxes extend to the 25th and 75th percentile models, and the whiskers extend the full range of models. Panel **c** shows the equivalent to **b** but for the CMIP5 models.

overestimate average yields and tend to underestimate yield variability as well as the severity of the largest historically observed yield declines. Figure 1a shows the simulated national-level probability distribution of log yields for each model, including both the original GCMs and the downscaled and bias-corrected versions. The national-level distribution is constructed by aggregating county-level yields using yearly production weights. In Fig. 1 and throughout this work, we remove the state-specific time trends and county-level fixed effects so that our analysis only considers the component of yields resulting from climate or weather (see the “Methods” section). Additionally, in order to facilitate a clearer diagnosis of the associated uncertainties, we compare model-simulated yields against yields simulated by an observational dataset rather than actual USDA records and choose the same observational dataset that was used to bias-correct and downscale the original models. From each model’s probability distribution in Fig. 1a, we compute four-summary statistics—the mean, standard deviation, 10th percentile, and minimum—and compare them with the corresponding observed values, shown via percentage differences in Fig. 1b and c.

Figure 1 highlights two important points, one relating to each ensemble. First, the vastly different outcomes simulated by the CMIP5 models suggest that some form of bias-correction is required before the corresponding yield projections can be used in decision-making contexts. For each summary statistic, the range of the CMIP5 models comfortably encompasses the historical value, and many models over- or under-estimate summary statistics by more than 100%. If the range of future projected yields is similarly broad, decision-makers in the agricultural sector, including farmers, insurance companies, or public sector agencies, may conclude that the CMIP5 models are under-confident and do not provide sufficiently precise information upon which to act. Figure 1 also demonstrates, as expected, that the NEX-GDDP models generally align much more closely with observations. However, the observed mean and minimum of historical yields lie outside of the NEX-GDDP ensemble range, indicating that in contrast to the CMIP5 models, NEX-GDDP may be over-confident with respect to certain aspects of the yield distribution. Of particular importance to decision-makers here is the lowest observed yield (“Min.” in Fig. 1), which represents the most extreme weather-induced yield drop throughout the historical period. Figure 1b shows that the downscaled and bias-corrected models underestimate this magnitude by between 63% and 233%, with the median model underestimating by 162%. Interestingly, the NEX-GDDP models give a much better representation of the 10th percentile, highlighting the importance of how extremes are defined. Based on a simple empirical analysis, the 10th percentile corresponds to approximately the 1-in-10-year event, while the lowest observed yield corresponds to approximately the 1-in-50-year event (from a 50-year sample). From a risk management perspective, it is important to properly characterize the distribution tails and thus the magnitude of these events³⁹, but exactly how reliable the models need to be at different thresholds ultimately depends on the decision context of the end-use application.

Hindcast evaluation at the county level. To what extent are the national-level results replicated across individual U.S. counties? We find that for many counties, particularly in and around the Corn Belt region encompassing historically high-yield counties (Fig. S1), conclusions from the national level hold. Figure 2 shows a map of numeric and percentage differences of the median-performing NEX-GDDP model relative to observations. The median-performing model is selected for each summary statistic

after taking an unweighted average over all counties in our sample. It gives an intuitive picture of how the overall ensemble performs in the sense that as many other models perform better than it as do worse, for each summary statistic. An extended analysis that evaluates all models simultaneously can be found in the Supplementary Information (Fig. S2). Figure 2a shows that the median-performing NEX-GDDP model overestimates the mean historical yield for almost all counties in the U.S. The other summary statistics show a more latitudinal gradient, with the median-performing model underestimating the length of the tails (Fig. 2b and c) and the magnitude of bulk variability (Fig. 2d) of more northern counties but matching or overestimating the observed values for warmer southern counties. For many counties in Illinois and Iowa, which are among the highest-production counties in the U.S. (Fig. S1), the median NEX-GDDP model underestimates the standard deviation by up to 50% and considerably underestimates the severity of the largest yield drop, which explains much of the findings of Fig. 1. Although we have only shown one model for each summary statistic here, the other NEX-GDDP models exhibit very similar behavior (Fig. S2).

The equivalent plots for the CMIP5 ensemble, shown in Figs. S3 and S4, again reveal a wide range of simulated yield outcomes, but most of the parent models overestimate historical yield variability and the associated tail statistics across most maize-growing counties in the U.S. Thus, we again conclude that while bias-correction and downscaling improve the representation of historical yields, it also potentially produces an overly narrow distribution and leads to an underestimation of aspects of the tails. Further, we find that this holds across many high-yield counties in the U.S., which is why this bias persists at the national level.

Hindcast evaluation of climate variables. What is the source of these biases 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 is 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 struggles to reproduce the observed distribution of EDDs. It performs better with the more moderate GDDs and with season-total precipitation.

Figure 3 shows a map of percentage differences of the median-performing NEX-GDDP model relative to observations, for each climate variable in the yield model and for each of our selected summary statistics. As demonstrated by Fig. 3a–d, the median NEX-GDDP model captures the observed GDD distribution quite well over the entire eastern half of the U.S., with only some biases in the standard deviation. In contrast, the models struggle to adequately represent EDDs (Fig. 3e–h), with considerable biases across all summary statistics that follow the latitudinal gradient displayed in Fig. 1. Recall that EDDs negatively impact yields, so that an underestimate of the historical mean (as in Fig. 3e) leads to an overestimate of average yields (as in Fig. 2a). This is also why we test the upper tails of the EDD distribution (the 90th percentile in Fig. 2f and maximum in Fig. 2g), since it corresponds to lower yields. For season-total precipitation (Fig. 3i–l), there is less evidence of a clear bias that would affect maize yields—the models tend to underestimate the amount of precipitation in the historically driest year (Fig. 3k) and tend to overestimate historical inter-annual variability (Fig. 3l). As we did for the yield evaluation, we also include here an extended analysis in the Supplementary Information that evaluates all models simultaneously, showing that our conclusions for the

Fig. 2 County-level evaluation of NEX-GDDP maize yield hindcasts. County-level differences between the median-performing NEX-GDDP model and observational data for selected summary statistics from the historical distribution of maize log-yields. For each summary statistic, the median-performing model (reported at the bottom of each panel) is selected after taking an unweighted average over all counties east of the 100th meridian west and comparing to the equivalent observational data. Numeric differences are reported in panels **a–c** because for some counties the observational value is very close to zero, which artificially inflates the percentage difference. For the standard deviation (panel **d**), we report percentage differences.

median-performing model hold across the ensemble (Figs. S5–S7). Thus, we argue that the primary driver of bias in the NEX-GDDP maize yield hindcasts is the underlying poor representation of EDDs, particularly throughout the high-production Corn Belt region.

As might be expected, the CMIP5 models give a somewhat more mixed picture. For EDDs, we find that the models tend to overestimate historical variability and the tail statistics (Figs. S8 and S10). Many of the CMIP5 models exhibit a good representation of GDDs, other than for the standard deviation (Fig. S9). Similarly, the CMIP5 models capture season-total precipitation quite well (Fig. S11). The fact that both ensembles struggle the most with EDDs, defined as an integrated measure of temperature above 29 °C, highlights the challenge of effectively representing climatological extremes. One aspect of this challenge revealed here is the difficulty in reproducing the observed spatial patterns: both the CMIP5 and NEX-GDDP ensembles are better able to reproduce the statistics of EDDs for hotter southern counties but struggle for colder northern ones. However, while the CMIP5 models tend to simulate northern counties that are too hot (measured by the historical average or upper tail) and with too high inter-annual variability (measured by the standard deviation), these shortcomings are over-corrected by the bias-correction and downscaling of NEX-GDDP.

Additional checks. We argue that biases in the hindcast simulations of U.S. maize yields result primarily from a poor underlying representation of EDDs. Here, we report on three additional checks that further validate this hypothesis and elucidate the origins of this bias. First, we perform two checks related to correlational effects between the two temperature variables in the yield model and the precipitation variable. In the observations, season-total precipitation is negatively correlated with both GDDs and EDDs with Spearman rank correlations of -0.11 ($P = 0.43$) and -0.23 ($P = 0.11$), respectively. We find that the models from both ensembles also exhibit negative correlations but tend to overstate their magnitudes (Figs. S12 and S13). This implies that in the models, hot years are relatively more likely to be dry and thus yields more likely to be lower. Despite this effect, the NEX-GDDP models still underestimate the severity of yield losses, pointing to the biases in EDDs as a more important factor. Indeed, we construct a yield model that only includes the temperature-derived variables, then perform an identical evaluation to that of Fig. 1, and find that the same biases persist and in some cases are enhanced (Fig. S14).

Second, we construct six additional models that do not use degree day variables, but growing season averages of daily maximum, minimum, and mean temperature, and with/without including precipitation (see Supplementary Methods Eqs. (1)–(8) for a detailed list of these models). Corroborating the findings of several previous works, we find that the temperature models all give a worse fit to the USDA data compared to the degree-day models (Table S1). However, the temperature models are generally better able to replicate the corresponding observationally simulated mean, minimum, and 10th percentile (Fig. S15). There is no meaningful improvement in replicating inter-annual variability measured by the standard deviation (Fig. S15). We also perform a two-sample Kolmogorov–Smirnov test for each model, with the null hypothesis that the simulated and observed samples of national-level yields are both drawn from the same reference distribution. For the degree-day models, the null hypothesis is rejected ($P < 0.01$) for all climate models from both ensembles (Fig. S16). In most cases, the null hypothesis is not rejected (at a 1% significance threshold) when using the temperature models coupled with NEX-GDDP. When the temperature models are coupled to CMIP5, the null hypothesis is rejected in many instances, highlighting again the importance of some form of bias-correction before estimating yields, even when using highly temporally aggregated temperature variables. This check is consistent with the interpretation that the NEX-GDDP biases found in our baseline degree-day model, particularly the underestimation of the mean and the low yield tail, stem from the degree day variables.

Finally, we highlight that biases in the representation of EDDs ultimately result from biases in the underlying daily temperature time series but are amplified by the integration process that is used when transforming to EDDs. We can rule out that the biases arise from the spatial aggregation in translating from the native grid of the NEX-GDDP models to the county level since we find that the biases persist at the model grid-scale (Fig. S17). We then perform a more qualitative check based on a representative time series of maximum and minimum temperature, and the corresponding accumulation of EDDs, shown in Fig. 4. The time series is taken from a single grid point lying within McLean County, Illinois, which is the highest-producing county over our historical sample. Figure 4a and b show two periods of continually high temperatures, beginning in early July, that fall outside of the NEX-GDDP ensemble range—particularly for the maximum temperature time series (Fig. 4b). As can be seen in Fig. 4c, the models’ underestimation of these extreme temperatures is then integrated into an underestimation of EDDs, from

Fig. 3 County-level evaluation of NEX-GDDP predictor variable hindcasts. County-level differences between the median-performing NEX-GDDP model and observational data, for selected summary statistics of each climate variable in the yield model: growing degree days (top row) (**a–d**), extreme degree days (middle row) (**e–h**), and season-total precipitation (bottom row) (**i–l**). For each summary statistic and variable, the median-performing model (reported at the bottom of each panel) is selected after taking an unweighted average over all counties east of the 100th meridian west and comparing to the equivalent observational data. For extreme degree days, we report on the 90th percentile (**f**) and maximum (**g**) rather than the 10th percentile and minimum since extreme degree days negatively impact yields.

which the models never recover. The result is an overestimation of yields for that year. It is likely that these temperature biases stem from the bias-correction procedure; however, since the NEX-GDDP dataset is only available after both bias-correction and downscaling have been applied, this is difficult to test rigorously in the current analysis. The equivalent time series plot for the CMIP5 models (Fig. S18) shows that both minimum and maximum temperature observations fall within the ensemble range all throughout the season. Consequently, accumulated EDDs also lie within the ensemble range, but this comes at the

expense of many CMIP5 models considerably overestimating end-of-season EDDs—some by a factor of 3 or more (Fig. S18c). Figure 4 (and Fig. S18) again demonstrate the difficulty in adequately representing climatological extremes, particularly when aiming to capture the downstream impacts on human or natural systems that may exhibit threshold responses or other nonlinear sensitivities.

Yield projections. Decision-makers care about the reliability of future projections, for which adequate past performance is

Fig. 4 Representative time series of historical temperature and extreme degree days. Daily time series of minimum temperature (a), maximum temperature (b), and the corresponding accumulation of extreme degree days (c) for a single representative grid point (40.375°N, −88.875°E) in March through August 1980. This grid point falls within McLean County, Illinois, which was the highest maize-producing county in the U.S. over 1956–2005. The NEX-GDDP model hindcasts are shown in gray and the observations in red.

Fig. 5 Projected national-level yield distributions and associated return periods. Panel a shows RCP 8.5 projections of mid-century (2030–2059) national-level log-yield distributions. The NEX-GDDP models are shown in purple and the CMIP5 models in orange; the historical observations are also shown as the black boxplot, for reference. Panel b shows the corresponding 1, 5, 10, and 20-year return period values for each model. The historical values are shown for comparison as black points. In the absence of future yearly county-level production shares, we aggregate to the national level by using the mean 1980–2020 weights. The boxes extend to the 25th and 75th percentile models, and the whiskers extend the full range of models.

generally regarded as a necessary but insufficient condition⁴⁰. A reasonable hypothesis, based on the tendency of the NEX-GDDP models to give overly narrow historical yield distributions, is that the projections are similarly overconfident.

We find that there are considerable differences between the yield projections of the NEX-GDDP and CMIP5 models. Figure 5a shows the projected national-level distributions for each model in both ensembles during 2030–2059. We use the

relatively high-emissions representative concentration pathway (RCP) 8.5 scenario⁴¹ to emphasize the differences between the ensembles. One robust response is that all model projections show an increase in yield variability and a shift to lower yields. However, as with the historical hindcasts, the NEX-GDDP models exhibit reduced variability relative to the original CMIP5 models. Although these projection differences alone do not demonstrate conclusively that the bias-corrected and downscaled models are overconfident, they do illustrate the importance of end-users' modeling choices, particularly in light of the results discussed in previous sections.

To demonstrate how projection differences might affect decision-making, we consider how the different distributions affect estimates of yield return periods. Return periods give the average time between (usually rare and damaging) events and are used often as part of risk analysis⁴². In our case, we estimate the average number of years between large national-level yield drops driven by adverse weather or climate. Figure 5b shows that the estimated return period magnitudes from the CMIP5 models are systematically larger than from NEX-GDDP. For example, consider the 20-year event: the median NEX-GDDP model projects that weather and climate will contribute only 0.25 to the overall (log-)yield, with an ensemble range of 0.1–0.4. The median CMIP5 model projects –0.29 and the ensemble range is –2.4 to 0.54. To give these estimates some context, the 20-year event based on historical observations is 0.36. This presents a challenge for decision-makers in deciding which product to rely on, given that low-probability, tail events can have a disproportionately negative impact. It might pay to hold commodity stocks under NEX-GDDP but not CMIP5. Alternatively, investing in adaptation measures might pass a cost–benefit test under CMIP5, but not NEX-GDDP.

If end-users believe that the NEX-GDDP and CMIP5 models are over- and under-confident, respectively, they may wish to use one of many methods aimed at constraining or otherwise modifying ensemble projections. There is a growing literature on model weighting or selection techniques that can account for model dependence and historical performance such that important ensemble metrics are optimized⁴³. Such schemes could be applied here but would bring additional assumptions, modeling choices, and potential pitfalls⁴³.

Discussion

Our main results broadly agree with the existing literature. Previous studies focusing on North America find that observational measures of interannual temperature variability fall within the CMIP5 ensemble range⁴⁴ and that the CMIP5 models simulate a wide range of historical and future meteorological extremes^{45–48}. To our knowledge, there has not been a comprehensive evaluation of the NEX-GDDP models over North America. However, analyses of NEX-GDDP over other regions^{49, 50} and analyses of very similar downscaling and bias-correction procedures^{51–53} demonstrate the much-improved representation of temperature and precipitation. This includes, in many instances, the representation of extremes as often defined by a standard set of indices⁵⁴. Where our analysis differs and expands on this literature is by using an evaluation framework that is specific to a given outcome of interest—in this case, U.S. maize yields and their negative response to temperatures above 29 °C. Our results show that such a framing can lead to a different interpretation regarding the performance of downscaling and bias-correction, particularly for extremes.

Our study also has important caveats that point to future research needs. First, we consider projections based on a relatively simple statistical downscaling and bias-correction method. The

bias-correction performs a univariate quantile mapping that corrects the statistical moments of the GCM output for each variable by pooling daily (± 15 days) data over all observational years. The spatial disaggregation (downscaling) step performs a bilinear interpolation of the GCM-generated fields that conserves the spatial structure of the observational data and the relative changes simulated by the GCMs. Our conclusions are of limited relevance to potentially more skillful bias-correction and downscaling techniques, which may give a better representation of atmospheric variability relative to the NEX-GDDP approach^{55, 56}.

Second, our conclusions may not apply to other econometric models that use different predictor variables or on different spatial or temporal scales. As described earlier, a yield model based on seasonal temperature averages rather than the integrated degree day variables does not suffer from the same biases (Fig. S15). Studies that rely on highly temporally averaged predictors⁵⁷ are thus unlikely to underestimate future impacts. However, econometric analyses for which the outcome variable is sensitive to temperature extremes—including, for example, for mortality rates⁵⁸ or electricity consumption⁵⁹—may underestimate the most extreme impacts owing to underlying biases in the climate data. In general, we would argue that performing a hindcast test targeted specifically at the outcome variable, as done here for maize yields, is a worthwhile endeavor. It may uncover biases that are not immediately apparent otherwise.

The third caveat of this work is the relative simplicity of the yield model. Several recent studies have highlighted the importance of soil moisture in simulating maize yields, which may or may not be strongly correlated with season-total precipitation depending on factors such as runoff, drainage, and irrigation^{60, 61}. We also neglect complex and potentially important intra-seasonal varying effects^{62, 63}, as well as several other factors that are typically included in process-based models, such as solar radiation, wind speed, humidity, and farm management practices. Although a growing number of comparative studies find similar impacts from statistical and process-based models^{28, 64, 65}, it is difficult to extrapolate our results to say anything meaningful about driving process-based models with downscaled and bias-corrected climate data. Finally, we assume no future adaptation in yield projections (meaning that coefficients are held fixed) and we do not account for other factors that may affect yields such as increased atmospheric carbon dioxide concentrations⁶⁶ or changing levels of air pollution⁶⁷. Recall that Fig. 5 shows the “weather only” contribution to yields above (or below) the technological time trend, meaning that even in years with large drops, the absolute yield could still be higher than today's levels. Thus Fig. 5 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.

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 the 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 econometric yield model and minimizing the influence of other factors, we find that the NEX-GDDP ensemble underestimates the largest historical yield drops driven by adverse weather. We can attribute this effect to the process of bias-correction and downscaling of the original GCM outputs. Through a series of additional checks, we also provide qualitative evidence that this bias stems from the underlying representation of temperature extremes in the form of EDDs. 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, the historical performance of bulk metrics and extremes, and possible overconfidence.

Given the caveats discussed in the preceding section, this work raises important questions about the use of bias-corrected and downscaled climate information, particularly when degree-day variables are used to simulate agricultural outcomes. Although we focus only on U.S. maize, other major crops such as soy, cotton, and wheat are often modeled using very similar econometric approaches^{68, 69}. NEX-GDDP is also one of the few downscaled and bias-corrected products with a global spatial extent, and we find that the biases uncovered here extend to other important agricultural regions, including Europe, China, Brazil, and sub-Saharan Africa (Figs. S19–S21). Without careful consideration of the representation of temperature extremes and its propagation through subsequent models, the use of bias-corrected and downscaled climate projections may lead to underestimates of the severity of impacts and consequently, poor decisions. This work also 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 daily precipitation, with projections run under RCPs 4.5 and 8.5. From the varying spatial resolutions of the parent CMIP5 models (see Table S2), all downscaled outputs have a constant lateral resolution of 0.25°. The bias-correction and downscaling were performed via the Bias-Correction Spatial Disaggregation algorithm³² (applied at the daily timescale) using observations from the Global Meteorological Forcing Dataset (GMFD)⁷⁰. Further details can be found in the Supplementary Methods. Academic studies that employ the NEX-GDDP dataset include refs. ^{5, 71–74}.

Yield model specification and evaluation. The concept of degree days is a widely used measure for the cumulative heat a crop is exposed to over the length of the growing season, which in this work is held constant as March through August. Degree days are derived for each grid cell by sinusoidally interpolating the diurnal temperature cycle using maximum and minimum daily temperatures³¹. GDDs are then calculated as the area under this temperature (T) curve subject to the bounds $10^{\circ}\text{C} < T \leq 29^{\circ}\text{C}$, summed over the growing season. EDDs are defined similarly, but with $T > 29^{\circ}\text{C}$. These bounds are specific to maize²⁷. 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. Crops derive most of their water from soils, not directly from precipitation; water availability is hence mediated through complex biophysical and hydrological processes. However, there is a limited historical record of spatially comprehensive soil moisture measurements, leading to the use of precipitation as a predictor variable and (often poor) proxy for water availability. Precipitation fields are also inherently noisier than temperature fields in climate reanalysis products, further complicating the extraction of the response function. Here, to keep the yield model simple and fast, we follow previous works and include a quadratic in season-total precipitation^{75, 76}.

Our resulting baseline yield model is then

$$\log Y_{i,t} = f_s(t) + c_i + \beta_1 \text{GDD}_{i,t} + \beta_2 \text{EDD}_{i,t} + \beta_3 P_{i,t} + \beta_4 P_{i,t}^2 + \epsilon_{i,t}. \quad (1)$$

Here, i denotes the spatial index (each U.S. county east of the 100th meridian west), t denotes the temporal index (each year), Y is the yield in bushels per acre, GDD denotes growing degree days, EDD denotes extreme degree days, and P denotes total in-season precipitation. A quadratic, state-specific time trend, $f_s(t)$, is included to account for technological progress and a county-level constant term, c_i , is included to account for time-invariant group fixed effects such as soil quality. The residual error is given by ϵ and assumed to be Gaussian, independent, and identically distributed. Standard errors on the regression coefficients are clustered at the state level. Other modeling choices are possible here, including Bayesian methods that would allow a more sophisticated approach to uncertainty in the regression coefficients. Additionally, the fit residuals are correlated in space (Fig. S22). More sophisticated treatments of the residual structure are possible, but we choose a simple model-fitting approach for parsimony.

Given these caveats, we fit this model using 1956–2005 as the training period. This corresponds to the longest timeframe for which the USDA, GMFD, and climate model data are all available. 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 . For the training period, with 96,142 observations throughout 2119 counties, the overall R^2 is 0.81 and the Within R^2 is 0.25. The Within R^2 gives the fraction of variance explained within counties, i.e., not including the fixed effects (the time trend and constant term). It hence gives a goodness-of-fit measure that incorporates only the climate and weather variables.

As discussed in the main text, we analyze the ‘weather only’ part of yields, given by

$$\log Y'_{i,t} = \beta_1 \text{GDD}_{i,t} + \beta_2 \text{EDD}_{i,t} + \beta_3 P_{i,t} + \beta_4 P_{i,t}^2.$$

This ensures that the historical distributions are stationary and allows us to isolate the effects of weather and climate. It also means that we do not need to estimate the future time trend.

We additionally specify and evaluate a total of eight models, which are detailed in the Supplementary Methods. For fit comparison purposes, we compute the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for each of them. Our baseline model described above has the lowest AIC and BIC scores of all models (Table S1).

Evaluation metrics. 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 evaluate in a manner that is insensitive to the temporal dimension and instead focuses on the long-term distributions. For each model in each ensemble, we construct the empirical distribution function of log-yield and compare the chosen summary statistics to the equivalent statistics from the observational data. All quantiles are calculated empirically. The sample size for most counties is 50 (1956–2005) although some counties are missing yield records for some years. Throughout this work, we neglect the uncertainty associated with calculating quantiles from a finite sample, since it is considerably smaller than the model-to-model or inter-ensemble variations that we focus on.

Data availability

The raw data that support the findings of this study is available as follows: NEX-GDDP climate data is available from NCCS (<https://www.nccs.nasa.gov/>); CMIP5 simulations are available from PCMDI (<https://pcmdi.llnl.gov/>); historical yields records are available from the USDA National Agricultural Statistics Service (<https://www.nass.usda.gov/>). The processed data to reproduce all figures in this study is available at <https://doi.org/10.5281/zenodo.5143434>.

Code availability

The code to reproduce all figures in this study is available at <https://doi.org/10.5281/zenodo.5143434>.

Received: 16 February 2021; Accepted: 5 August 2021;

Published online: 20 September 2021

References

1. Intergovernmental Panel on Climate Change. *Climate Change 2014—Impacts, Adaptation and Vulnerability: Part A: Global and Sectoral Aspects: Working Group II Contribution to the IPCC Fifth Assessment Report: Vol. 1: Global and Sectoral Aspects* (Cambridge University Press, 2014).
2. Huber, V. et al. Climate impact research: beyond patchwork. *Earth Syst. Dyn.* 5, 399–408 (2014).

3. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1107 (2009).
4. Hsiang, S. et al. Estimating economic damage from climate change in the United States. *Science* **356**, 1362–1369 (2017).
5. Ruijven, B. J., van, C., E. D. & Wing, I. S. Amplification of future energy demand growth due to climate change. *Nat. Commun.* **10**, 2762 (2019).
6. Xu, C., Kohler, T. A., Lenton, T. M., Svenning, J.-C. & Scheffer, M. Future of the human climate niche. *Proc. Natl Acad. Sci. USA* **117**, 11350 (2020).
7. 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).
8. Anderson, J. et al. Progress on incorporating climate change into management of California's water resources. *Clim. Change* **87**, 91–108 (2008).
9. Martinich, J. & Crimmins, A. Climate damages and adaptation potential across diverse sectors of the United States. *Nat. Clim. Change* **9**, 397–404 (2019).
10. Arnbjerg-Nielsen, K., Leonardsen, L. & Madsen, H. Evaluating adaptation options for urban flooding based on new high-end emission scenario regional climate model simulations. *Clim. Res.* **64**, 73–84 (2015).
11. 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. *Bull. Am. Meteorol. Soc.* **92**, 1181–1192 (2011).
12. Dixon, K. W. et al. Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Clim. Change* **135**, 395–408 (2016).
13. Teutschbein, C. & Seibert, J. Is bias correction of regional climate model (RCM) simulations possible for non-stationary conditions? *Hydrol. Earth Syst. Sci.* **17**, 5061–5077 (2013).
14. Maraun, D. et al. Towards process-informed bias correction of climate change simulations. *Nat. Clim. Change* **7**, 764–773 (2017).
15. 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**, RG3003 <https://agupubs.onlinelibrary.wiley.com/action/showCitFormats?doi=10.1029%2F2009RG000314> (2010).
16. 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).
17. Lopez-Cantu, T., Prein, A. F. & Samaras, C. Uncertainties in future U.S. extreme precipitation from downscaled climate projections. *Geophys. Res. Lett.* **47**, e2019GL086797. <https://doi.org/10.1029/2019GL086797> (2020).
18. Maraun, D. Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. *J. Clim.* **26**, 2137–2143 (2013).
19. Walton, D. B., Sun, F., Hall, A. & Capps, S. A hybrid dynamical–statistical downscaling technique. Part I: Development and validation of the technique. *J. Clim.* **28**, 4597–4617 (2015).
20. Hertig, E. & Jacobeit, J. A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area. *J. Geophys. Res. Atmos.* **118**, 520–533 (2013).
21. Sippel, S. et al. A novel bias correction methodology for climate impact simulations. *Earth Syst. Dyn.* **7**, 71–88 (2016).
22. 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. Clim.* **28**, 6938–6959 (2015).
23. Eum, H.-I., Cannon, A. J. & Murdock, T. Q. Intercomparison of multiple statistical downscaling methods: multi-criteria model selection for South Korea. *Stoch. Environ. Res. Risk Anal.* **31**, 683–703 (2017).
24. Teutschbein, C. & Seibert, J. Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. *J. Hydrol.* **456–457**, 12–29 (2012).
25. Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G. & Lobell, D. B. Anthropogenic climate change has slowed global agricultural productivity growth. *Nat. Clim. Change* **11**, 306–312 (2021).
26. IPCC, 2019: Summary for Policymakers. In: *Climate Change and Land: An IPCC Special Report On Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems* (eds Shukla, P. R. et al.) (2019). <https://www.ipcc.ch/srccl/chapter/summary-for-policymakers/>. In press.
27. Schlenker, W. & Roberts, M. J. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proc. Natl Acad. Sci. USA* **106**, 15594 (2009).
28. Schauburger, 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. *Agric. 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. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
34. Glotter, M. et al. Evaluating the utility of dynamical downscaling in agricultural impacts projections. *Proc. Natl Acad. Sci. USA* **111**, 8776–8781 (2014).
35. Ceglar, A. & Kajfež-Bogataj, L. Simulation of maize yield in current and changed climatic conditions: addressing modelling uncertainties and the importance of bias correction in climate model simulations. *Eur. J. Agron.* **37**, 83–95 (2012).
36. Liu, M. et al. What is the importance of climate model bias when projecting the impacts of climate change on land surface processes? *Biogeosciences* **11**, 2601–2622 (2014).
37. Liu, D. L. et al. Propagation of climate model biases to biophysical modelling can complicate assessments of climate change impact in agricultural systems. *Int. J. Climatol.* **39**, 424–444 (2019).
38. Laux, P. et al. To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data. *Agric. For. Meteorol.* **304**, 108406 (2021).
39. Bernard, L., Semmler, W., Keller, K. & Nicholas, R. *Improving Climate Projections to Better Inform Climate Risk Management* <https://doi.org/10.1093/oxfordhb/9780199856978.013.0002> (2015).
40. Knutti, R. Should we believe model predictions of future climate change? *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **366**, 4647–4664 (2008).
41. Vuuren, D. Pvan et al. The representative concentration pathways: an overview. *Clim. Change* **109**, 5 (2011).
42. Sriv, 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).
43. Herger, N. et al. Ensemble optimisation, multiple constraints and overconfidence: a case study with future Australian precipitation change. *Clim. Dyn.* **53**, 1581–1596 (2019).
44. Lewis, S. C. & Karoly, D. J. Evaluation of historical diurnal temperature range trends in CMIP5 models. *J. Clim.* **26**, 130715122904005 (2013).
45. Yao, Y., Luo, Y., Huang, J. & Zhao, Z. Comparison of monthly temperature extremes simulated by CMIP3 and CMIP5 models. *J. Clim.* **26**, 130513145307007 (2013).
46. Sriv, 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).
47. Hogan, E., Nicholas, R. E., Keller, K., Eilts, S. & Sriv, R. L. Representation of US warm temperature extremes in global climate model ensembles. *J. Clim.* **32**, 2591–2603 (2019).
48. Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. & Bronaugh, D. Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **118**, 1716–1733 (2013).
49. Alvaro, A.-D., Gabriel, A., Flavio, J., Roger, T. & Aaron, W. Extreme climate indices in Brazil: evaluation of downscaled earth system models at high horizontal resolution. *Clim. Dyn.* **54**, 5065–5088 (2020).
50. Raghavan, S. V., Jina, H. & Shie-Yui, L. Evaluations of NASA NEX-GDDP data over Southeast Asia: present and future climates. *Clim. Change* **148**, 503–518 (2018).
51. Bürger, G., Murdock, T. Q., Werner, A. T., Sobie, S. R. & Cannon, A. J. Downscaling extremes—an intercomparison of multiple statistical methods for present climate. *J. Clim.* **25**, 4366–4388 (2012).
52. Maurer, E. P. & Hidalgo, H. G. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrol. Earth Syst. Sci.* **12**, 551–563 (2008).
53. Gutmann, E. et al. An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resour. Res.* **50**, 7167–7186 (2014).
54. Dunn, R. J. H. et al. Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *J. Geophys. Res.* **125**, e2019JD032263 (2020).
55. Abatzoglou, J. T. & Brown, T. J. A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.* **32**, 772–780 (2012).
56. Pierce, D. W., Cayan, D. R. & Thrasher, B. L. Statistical downscaling using localized constructed analogs (LOCA)*. *J. Hydrometeorol.* **15**, 2558–2585 (2014).

57. Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).
58. Carleton, T. A. et al. *Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits* <http://www.nber.org/papers/w27599> (2020).
59. Auffhammer, M., Baylis, P. & Hausman, C. H. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proc. Natl Acad. Sci. USA* **114**, 1886–1891 (2017).
60. Haqiqi, I., Grogan, D. S., Hertel, T. W. & Schlenker, W. Quantifying the impacts of compound extremes on agriculture. *Hydrol. Earth Syst. Sci.* **25**, 551–564 (2021).
61. 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).
62. Ortiz-Bobea, A., Wang, H., Carrillo, C. M. & Ault, T. R. Unpacking the climatic drivers of US agricultural yields. *Environ. Res. Lett.* **14**, 064003 (2019).
63. Lesk, C., Coffel, E. & Horton, R. Net benefits to US soy and maize yields from intensifying hourly rainfall. *Nat. Clim. Change* **10**, 819–822 (2020).
64. Lobell, D. B. & Asseng, S. Comparing estimates of climate change impacts from process-based and statistical crop models. *Environ. Res. Lett.* **12**, 015001 (2017).
65. Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B. & Schlenker, W. Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environ. Res. Lett.* **12**, 095010 (2017).
66. 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).
67. Lobell, D. B. & Burney, J. A. Cleaner air has contributed one-fifth of U.S. maize and soybean yield gains since 1999. *Environ. Res. Lett.* **16**, 074049 <https://iopscience.iop.org/article/10.1088/1748-9326/ac0fa4> (2021).
68. Rising, J. & Devineni, N. Crop switching reduces agricultural losses from climate change in the United States by half under RCP 8.5. *Nat. Commun.* **11**, 4991 (2020).
69. Tack, J., Barkley, A. & Nalley, L. L. Effect of warming temperatures on US wheat yields. *Proc. Natl Acad. Sci. USA* **112**, 6931–6936 (2015).
70. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **19**, 3088–3111 (2006).
71. Missirian, A. & Schlenker, W. Asylum applications respond to temperature fluctuations. *Science* **358**, 1610–1614 (2017).
72. Ahmadiipour, A., Moradkhani, H. & Svoboda, M. Centennial drought outlook over the CONUS using NASA-NEX downscaled climate ensemble. *Int. J. Climatol.* **37**, 2477–2491 (2017).
73. Thilakarathne, M. & Sridhar, V. Characterization of future drought conditions in the Lower Mekong River Basin. *Weather Clim. Extrem.* **17**, 47–58 (2017).
74. Obradovich, N., Tingley, D. & Rahwan, I. Effects of environmental stressors on daily governance. *Proc Natl Acad. Sci. USA* **115**, 8710–8715 (2018).
75. 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).
76. Butler, E. E. & Huybers, P. Adaptation of US maize to temperature variations. *Nat. Clim. Change* **3**, 68–72 (2013).

Acknowledgements

We thank the editor and three anonymous reviewers whose constructive comments greatly improved the manuscript. This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics, Contract No. DE-SC0016162 as well as the Penn State Center for Climate Risk Management. Computations for this research were performed on the Pennsylvania State University's Institute for Computational and Data Sciences Advanced CyberInfrastructure (ICDS-ACI). Climate scenarios used were from the NEX-GDDP dataset, prepared by the Climate Analytics Group and NASA Ames Research Center using the NASA Earth Exchange, and distributed by the NASA Center for Climate Simulation (NCCS). We thank Murali Haran for insightful discussions. All errors and opinions are those of the authors and not of the funding entities.

Author contributions

D.C.L.: Conceptualization, methodology, formal analysis, visualization, writing—original draft, writing—review & editing. R.L.S.: Conceptualization, methodology, writing—review & editing, data curation, supervision, funding acquisition. I.H.: Conceptualization, methodology, writing—review & editing. T.W.H.: Conceptualization, methodology, writing—review & editing, supervision, funding acquisition. K.K.: Conceptualization, methodology, writing—review & editing, funding acquisition. R.E.N.: Conceptualization, methodology, writing—review & editing, data curation, funding acquisition.

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/s43247-021-00266-9>.

Correspondence and requests for materials should be addressed to David C. Lafferty.

Peer review information *Communications Earth & Environment* thanks James Rising and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Clare Davis. Peer reviewer reports are available.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021