Global warming, home runs, and the future of America's pastime

Christopher W. Callahan[1,2*], Nathaniel J. Dominy[1,3], Jeremy M. DeSilva[1,3], & Justin S. Mankin[1,2,4,5]

[1]Program in Ecology, Evolution, Environment and Society, Dartmouth College, Hanover, NH

[2]Department of Geography, Dartmouth College, Hanover, NH

[3]Department of Anthropology, Dartmouth College, Hanover, NH

[4]Department of Earth Sciences, Dartmouth College, Hanover, NH

[5]Ocean and Climate Physics, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY

*Corresponding author, Christopher.W.Callahan.GR@dartmouth.edu

Abstract

Home runs in baseball—fair balls hit out of the field of play—have risen since 1980, driving strategic shifts in gameplay. Myriad factors likely account for these trends, with some speculating that global warming has contributed via a reduction in ballpark air density. Here we use observations from 100,000 Major League Baseball games and 220,000 individual batted balls to show that higher temperatures substantially increase home runs. We isolate human-caused warming with climate models, finding that >500 home runs since 2010 are attributable to historical warming. Several hundred additional home runs per season are projected due to future warming. Adaptations such as building domes on stadiums or shifting day games to night games reduce temperature's effects on America's pastime. Our results highlight the myriad ways that a warmer planet will restructure our lives, livelihoods, and recreation, some quantifiable and easily adapted to, as shown here, many others, not.

Capsule

We show that global warming has increased home runs in baseball by reducing gametime air density. Without gameplay adaptations, future warming will intensify this effect alongside other climate impacts.

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-22-0235.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Main Text

Home runs are an exhilarating component of baseball, occurring, typically, when a batter hits the ball entirely out of the field of play, scoring one or more runs for their team. A single home run can alter the outcome of a game, causing elation or despair for hundreds of thousands of fans, and the powerful hitters who produce scores of home runs annually are among the most valued athletes in professional sports. Major League Baseball (MLB), the highest professional baseball organization of the United States, has witnessed a long-term increase in home runs per game since the 1980s (Fig. 1A). The recent surge from 2015-2019 raised concern from baseball officials, fans, and players that the dominance of home runs may undermine other compelling parts of the ballgame (Albert et al., 2018, 2019; Schreck & Sickles, 2018). Increases in home runs can alter baseball strategy, player acquisition and compensation, as well as the popular culture that sustains the sport. As such, understanding the source of home run trends is an important area of focus for baseball analysts.

In addition to factors such as the construction of the baseball, performance-enhancing drugs, advanced technology, analytics, and player training, climate change has been raised as a potential contributor to home run trends (Dykstra, 2012; Samenow, 2012). The ideal gas law tells us that air density is inversely proportional to temperature (Clapeyron, 1835). Ballistics tells us that the trajectory of a batted ball is influenced by temperature via its effect on density. All else being equal, warmer air is less dense, and a batted ball will carry farther (Adair, 2002; Albert et al., 2018; Bahill et al., 2009). The well-documented rise in home runs (Fig. 1A) has coincided with a long-term increase in gametime temperatures at baseball stadiums (Fig. 1B) and a resulting decrease in air density during games (Fig 1C). These trends have fueled spirited debate among commentators and sportswriters, including controversial arguments concerning the role of global warming (Dykstra, 2012; Samenow, 2012), but a formal analysis linking human-caused climate change and home run totals has not been performed to date.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

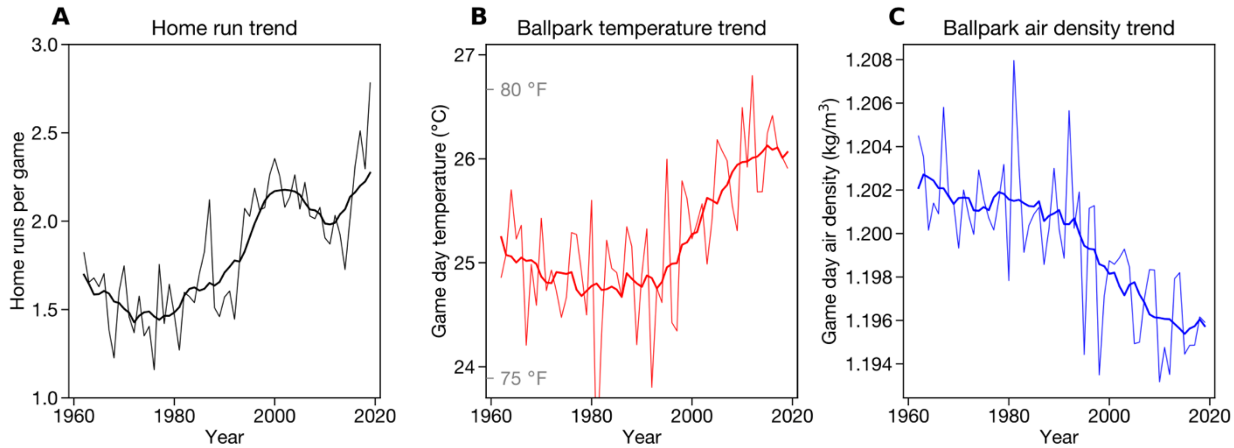

Fig. 1 | Observed trends in home runs, temperature, and air density during baseball games. A) Average number of home runs in Major League Baseball games from 1962-2019. B) Average daily high temperature (°C, y-axis and °F, internal y-axis) at baseball stadiums on game days from 1962-2019. C) Average air density (kg m^{-3}) at baseball stadiums on game days from 1962-2019, calculated using gameday temperature, vapor pressure, and sea level pressure. In all plots, the thin line is the actual time series and the bolded line is a 10-year running mean.

Observers have leveraged the trove of baseball data to highlight correlations between gameday temperatures and batted ball distances (Kraft & Skeeter, 1995) or home run totals (Koch & Panorska, 2013), but many confounding factors complicate a causal interpretation of these relationships. Home runs have likely risen due to advanced analytics and data on pitcher tendencies, training, and changes to the construction of baseballs (Albert et al., 2018, 2019; Schreck & Sickles, 2018; Vincent, 2007), all of which happen to coincide with the warming trend. The variable dimensions and elevations of ballparks may further confound associations between temperature and home runs by making home runs easier or more difficult within and across parks. Additionally, the baseball season coincides with the annual cycle of temperature, and home runs may increase with spring warming as players become more comfortable hitting after the offseason. As such, estimating a causal effect of temperature on home runs requires an identification strategy that separates temperature from these other, potentially unobservable, drivers.

Here we empirically estimate the relationship between gameday temperatures and home runs across >100,000 Major League Baseball games between 1962-2019 and >220,000 individual batted balls between 2015-2019. Our model is a Poisson regression, which estimates the number of home runs in a game as a function of the temperature during that game (Appendix, Eqn. 1). We use fixed effects to control for the spatially varying, long-term trending, and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

seasonally varying factors that could confound the temperature-home run relationship (Angrist & Pischke, 2008; S. Hsiang, 2016). The result from this framework is a plausibly causal estimate of the effect of an increase in temperature on the number of home runs hit in each game (Appendix). We then couple our empirical estimates with experiments from climate models that separate human-caused warming from natural climate variations. In conjunction with our empirical model relating temperature to home runs, these model experiments allow us to quantify the influence of historical climate change on home run totals. They also allow us to project how home runs may change in the future with warming.

A 1-°C increase in the daily high temperature on the day of a baseball game played in a stadium without a dome increases the number of home runs in that game by 1.96% (95% confidence interval [CI]: 1.5 – 2.4) (Fig. 2A). The effect is larger (2.4 % $°C^{-1}$) for games played in the early afternoon when temperatures are highest ("day games"), and smaller (1.7 % $°C^{-1}$) for games played in the evening when temperatures are milder ("night games"). These coefficients are outside the range of estimates produced when gameday temperatures are randomized within each park (Fig. 2B, Appendix), indicating that they are unlikely to be spurious (S. M. Hsiang & Jina, 2014).

The effects described above include all games played in open-air stadiums and in retractable-roof stadiums when the roof is open (Appendix), comprising 89% of our observations. In the remaining 11% of games, which are played under closed domes, we find only small and insignificant effects of temperature, as covered games are less exposed to ambient weather (Fig. 2B, Table S1).

Home runs respond similarly to temperature for both home and visiting teams independently, for subsets of the time period of analysis, and for alternative temperature data sets (Table S2, Fig. S1), again indicating the relationship is robust. Excluding years after 2000, in which the most recent home run and warming surges have occurred, does not alter our results, indicating that this recent period is not driving our identification. A negative binomial regression model, which relaxes the requirement for equal mean and variance in the Poisson model, yields similar results (Table S3). When we control for precipitation, relative humidity, or wind speed, the temperature effect remains (Table S4). Precipitation is unlikely to have an effect since games are usually canceled in inclement weather and relative humidity has only a minor effect on air density (Fig. S2). Consistent with expectation, we do not find these additional variables to have

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

significant effects. We also find no significant effect of wind speed (Table S4), potentially because we do not have high-quality within-park measurements of wind speed and direction. All standard errors are clustered at the park and year levels, which accounts for spatial and temporal autocorrelation in the processes generating home runs.

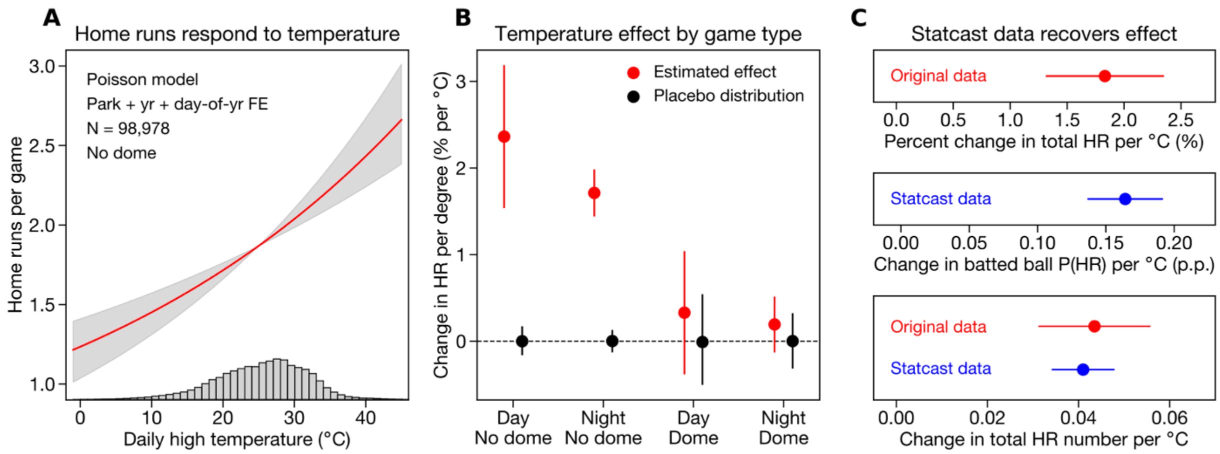

Fig. 2 | Empirical relationship between game day temperature and home runs. A) Relationship between temperature and home runs per game, derived from a Poisson model with park, year, and day-of-year fixed effects. Red line is the estimate and gray shading shows 95% confidence intervals, centered on the sample average temperature and home run number. Model is fitted to data from parks without domes. Lower histogram shows the density of temperature observations in the sample. B) Poisson regression coefficients showing the percent change in home runs per °C for day versus night games in parks with and without domes. Red estimates are from the actual data; black estimates are from a model where temperature values are randomized within park groups (same park, same home runs, different temperature). C) Comparison of results from the main data (red) and Statcast data (blue). Top row shows the main estimate of percent change in home runs per °C from all games in parks without domes. Middle row shows the percentage-point change in the probability of a batted ball being a home run per °C using a linear probability model fitted to the Statcast data (Appendix). Lower row shows the change in total number of home runs per game per °C for both data sources, converted to comparable estimates. In B and C, dots correspond to central estimates while lines span the 95% confidence intervals.

Higher temperatures may affect home runs through multiple complex pathways beyond air density, such as heat stress on pitchers (Howe & Boden, 2007). To clarify the mechanism, we perform two additional analyses. First, we directly allow air density to enter the regression alongside temperature (Table S5). Independently, both air density and temperature have significant effects with comparable magnitudes. When both are included in the model, however, the effect of temperature becomes small and not significant, while the effect of air density

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

remains (Table S5). These results imply that temperature's effect on home run occurs primarily by modifying air density.

To further confirm this mechanism, we analyze data from the Statcast system of high-speed cameras installed in MLB stadiums, which provides high-resolution tracking data on every fly ball from 2015 to 2019. These data allow us to control for the launch angle and launch speed of each batted ball (Albert et al., 2018). Such controls effectively hold the skill of the pitcher and hitter constant, comparing a ball leaving the bat at the same angle and speed on a warm day versus a cool day. As such, only the air density mechanism is left to account for changes in home run probability. Our game-level model (Appendix, Eqn. 1) yields a 1.83% increase in home runs per game per °C over 2015-2019 (Fig. 2C, top). We fit a linear probability model to the Statcast data to estimate the effect of temperature on the probability that a fly ball is a home run, controlling for launch angle and speed (Appendix, Eqn. 2). We find a 0.16-percentage-point increase in home run probability per °C (CI: 0.14 – 0.19) (Fig. 2C, middle). We can compare the game-level and batted ball-level estimates with dimensional analysis: Multiplying the game-level estimate of 1.83% by the average number of home runs in a game (2.38) yields 0.44 additional home runs for a 1-°C temperature increase (CI: 0.31 – 0.56). Multiplying the Statcast probability change of 0.16 p.p. per fly ball by the average number of fly balls in a game (24.98) yields a very similar estimate of 0.41 home runs (CI: 0.34 – 0.48) (Fig. 2C, bottom). These two estimates are similar in magnitude and statistically indistinguishable, supporting our conclusion that home runs increase with temperature due to reduced air density.

Comparing temperature differences in climate model experiments that include greenhouse gas forcing versus those that exclude it allows us to quantify changes in home runs due to human-caused warming (Appendix, Fig. 3). We find that human-caused climate change decreased home runs between 1962 and 1995 and increased them thereafter (Fig. 3A). Between 2010 and 2019, global warming led to an additional 58 home runs per year on average (ensemble standard deviation [SD]: 20 – 96) and 577 cumulatively (SD: 195 – 959). The high atmospheric loading of anthropogenic aerosols cooled regional climate from the 1960s through the 1980s (Eyring et al., 2021). Policy changes aiming to improve air quality succeeded, so greenhouse gas forcing has dominated recent climate changes, accelerating home runs. This acceleration, while statistically distinguishable, is small—an increase of some 1% relative to the total number of

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

home runs in 2019. Other factors such as changes in the height of the stitches on the baseball appear to have been more important in driving recent home run trends (Albert et al., 2018, 2019).

Despite the modest historical change, home runs will—absent changes in gameplay—increase with future warming (Fig. 3A). Each degree of global warming is associated with an additional ~95 home runs per baseball season (Fig. 3B). The magnitude and spatial pattern of warming, structural features like domes, and the fraction of day games determine ballpark-level responses (Fig. 3C, Table S6).

Warming associated with a high-emissions pathway (SSP5-8.5) results in an additional 192 home runs per year by 2050 (SD: 131 – 253) and an additional 467 by 2100 (SD: 337 – 598), almost a 10% increase relative to 2000-2019. Strong climate mitigation (SSP1-2.6) holds the increase to ~130 by 2100. In context, these increases are large: The home run "explosion" between 2015 and 2019 resulted in ~350 additional home runs per year (Schreck & Sickles, 2018), and warming can generate a home run surge that exceeds previous such surges. Mitigation reduces these effects: Limiting global warming to 1.5 °C instead of 2 °C would avoid an additional 1865 home runs cumulatively (SD: 961 – 2770). Limiting warming to 2 °C instead of 3 °C would avoid an additional 3891 (SD: 2434 – 5347) (Fig. 3B, inset).

In the absence of mitigation, MLB could adapt to limit warming's effects on home runs. Converting all day games into night games, for example, reduces exposure to daily high temperatures. While our main analysis holds the future frequency of day and night games fixed at the historical level for each park (Appendix), a season entirely comprised of night games would cut the effect of warming by 15% on average (Fig. 3B). Other structural adaptations such as adding domes to stadiums would also dampen the influence of climate change: On average, domed parks experience only 45% of the home run increase that non-domed parks do (Fig. 3C). Other unobserved adaptations such as altering pitching strategy, in-game management, or changes in bat composition may further blunt the effects of warming.

While changes in technology and player skill will undoubtedly shape the projections we show here, our results highlight that MLB will need to contend with climate change's influence on baseball performance. Steadily rising home runs may alter the incentives for player acquisition, offensive and defensive strategy, and public perception and engagement with the game, with consequences for the business of baseball and its on-field play.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

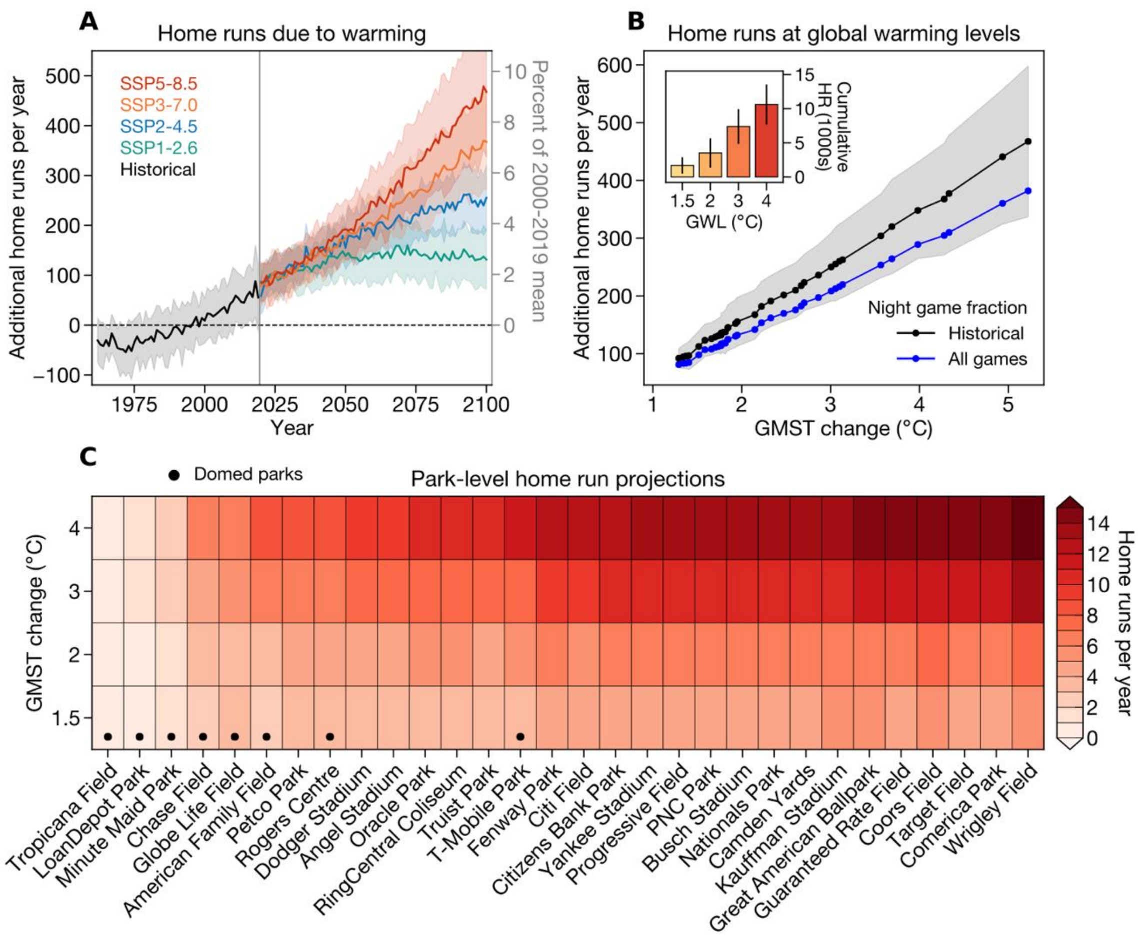

Fig. 3 | Home runs due to anthropogenic climate change. A) Change in total home runs per season due to anthropogenic climate change over the historical period (black) and across four Shared Socioeconomic Pathway (SSP) emissions scenarios: SSP1-2.6 (teal), SSP2-4.5 (blue), SSP3-7.0 (orange), and SSP5-8.5 (red). Solid lines show the multi-model ensemble mean; shading spans the mean plus or minus the ensemble standard deviation. B) Change in total home runs per year referenced to global mean surface temperature (GMST) change when the fraction of day and night games is held at historical averages for each park (black, Table S6) and when all games are night games (blue). Solid line shows the ensemble mean and shading spans the mean plus or minus the ensemble SD. Dots denote decadal averages of home runs and GMST change, following Pachauri and Meyer (2014). Inset shows cumulative additional home runs at each global warming level in thousands. Bar height shows ensemble mean and error bars span the mean plus or minus the ensemble standard deviation. C) Additional home runs per year for each individual park at global warming levels of 1.5, 2, 3, and 4 °C. Parks are ranked from left to right based on the number of additional home runs at 4 °C. Data are pooled across SSP scenarios in panels B and C.

Identifying, quantifying, and adapting to climate change requires data, and professional baseball is one of the most well-documented activities on the planet. While such data position us to assess warming impacts on home runs, they belie the far more difficult (and data-poor) task of

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

assessing the acute human risks that accompany such gameplay changes, such as increased heat stress for ballpark staff, patrons, and players alike, or wider risks to the unequal cities in which these ballparks reside. More broadly, our findings are emblematic of the widespread influence anthropogenic global warming has already had on all aspects of life. Warming will continue to burden the poorest and most vulnerable among us, altering the risks of wildfires, heat waves, droughts, and tropical cyclones (IPCC, 2022). Our results point to the reality that even the elite billion-dollar sports industry is vulnerable to unexpected impacts. Greenhouse gas mitigation and climate adaptation are a priority not only to reduce the large-scale loss and damage associated with extreme climate events, but also to avoid pervasive (and sometimes subtle) changes to recreation and leisure activities enjoyed by people.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

Appendix: Materials and Methods

Data

The data used for the empirical analysis is a panel dataset of the number of home runs in each MLB game matched with daily high temperatures on the day of the game, made up of 114,417 observations spanning 1962-2019.

Our primary source of baseball data is the Retrosheet game logs database (https://www.retrosheet.org/gamelogs/index.html), which provides summary data for all MLB games played, reaching back to 1871. We use game log data starting in 1962, the first year in which the current 162-game season was adopted, and continuing through the end of the 2019 season. Data from 2020 or 2021 are excluded due to the unusual conditions induced by the COVID-19 pandemic. For each game, we extract the total number of home runs as well as identifying information including date, park, and the names of the home and visiting teams. For some games, manually recorded temperature within each ballpark at game time is also available from the Retrosheet event files (https://www.retrosheet.org/game.htm).

The primary temperature data is drawn from the HadISD database v3.1.2 (Dunn, 2019; Dunn et al., 2012, 2016; Smith et al., 2011), which provides hourly data for weather stations across the United States. Using each baseball stadium's latitude and longitude, we find the five closest HadISD stations and average them, weighting by the inverse distance between each station and the stadium. We extract daily maximum and mean temperature, wind speed, vapor pressure, and sea level pressure from each station. Mean temperature, vapor pressure, and sea level pressure are used to calculate air density (Fig. 1C). Alternative temperature data come from the ERA5 reanalysis (C3S, 2017), where we select daily maximum temperatures from the grid cell closest to each park, and the Retrosheet game-time temperature data. Daily total precipitation data over 1997-2015 come from the Global Precipitation Climatology Project (Adler et al., 2017), again matched by the closest grid cell for each stadium.

Most baseball stadiums are open-air, but several have retractable or (in the case of Tropicana Field) fixed domes. In many of the games played in these stadiums, the Retrosheet data lists whether the dome was open or closed in each game. We estimate separate regressions for games played in the open air and games played under domes (Fig. 2B) given the different exposure to ambient temperature in each situation. We drop observations where the Retrosheet

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

data lists the dome status as “unknown,” which removes ~17% of observations in the retractable-roof parks but does not affect the data in open parks.

We download granular data on individual batted balls from 2015-2019 from the Statcast web portal (https://baseballsavant.mlb.com/statcast_search). We filter the data to all batted balls classified as “line drives” and “fly balls,” with the “outcome” as one of the following: home run, single, double, triple, field out, error, and sacrifice fly. We then match these batted ball data to our original gameday temperature data from HadISD, for a total of 223,337 observations.

Climate model data are drawn from the sixth phase of the Coupled Model Intercomparison Project (CMIP6) (Eyring et al., 2016; Gillett et al., 2016; O’Neill et al., 2016). We use daily maximum temperature data from the historical, historical-natural, and several Shared Socioeconomic Pathway (SSP) experiments: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. Comparison of temperatures in the historical and historical-natural experiments enable an analysis of human contributions to warming thus far, and therefore how home runs have changed due to historical human-caused warming (Gillett et al., 2016). The SSP scenarios enable an analysis of future changes in home runs due to projected warming conditioned on different assumptions about emissions mitigation (O’Neill et al., 2016). The SSP scenarios used here are the “Tier 1” experiments from the Scenario Model Intercomparison Project and span a wide range of future climate outcomes, from aggressive mitigation (SSP1-2.6) to very high emissions (SSP5-8.5) (Tebaldi et al., 2021). We analyze data from 14 models across the historical and SSP experiments, each with between 1 and 50 realizations, for a total of 403 simulations (Tables S7-S11). The combination of multiple models and multiple realizations from individual models allows us to sample uncertainty from both model structure and model representations of internal climate variability, while using multiple scenarios allows us to sample uncertainty from emissions trajectories.

Park-level temperatures are calculated from each climate model by finding the grid cell closest to each park’s latitude and longitude and calculating the spatial average across that grid cell and the eight grid cells surrounding it.

Regression analysis

The central goal of our analysis is to estimate a plausibly causal effect of temperature on home runs per game. We estimate the following Poisson fixed effects model that relates total

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

home runs in each game (HR) to the daily high temperature on the day of the game (Tx), with games indexed by park p, day-of-year d, and year y:

$$HR_{pdy} = exp\left(\beta Tx_{pdy} + \mu_p + \gamma_d + \delta_y + \epsilon_{pdy}\right) \tag{1}$$

The coefficient of interest is β, which identifies the percent change in home runs for a 1-°C change in the daily high temperature on the day of the game. μ is a park fixed effect, which accounts for time-invariant average differences between parks such as their physical dimensions, latitude, or altitude. γ is a day-of-year fixed effect, which accounts for the seasonality of home runs and temperature, capturing changes in the comfort level of batters after the offseason as temperatures warm. δ is a year fixed effect, which accounts for league-wide changes across years, such as alterations to the structure of the baseball itself and the long-term global warming trend. These fixed effect terms are equivalent to demeaning each term before identifying; a park fixed effect, for example, means that the home run and temperature terms are estimated using deviations from each park's long-term average number of home runs and temperature, respectively. We use a Poisson model because the home run data are integer values with an approximately exponential distribution.

We use two-way clustering for the standard errors on the estimates from Eqn. 1, clustering simultaneously at the park and year level to account for autocorrelation in the errors in both time and space. When uncertainty in the coefficients is sampled in the climate model analysis (see *Historical home run attribution* and *Future home run projections* sections), we generate a set of coefficients from a Gaussian distribution using the estimate of β as the mean and the clustered standard error on β as the standard deviation. We also randomize temperatures within parks in Fig. 2B to nonparametrically estimate a null distribution of coefficients, accounting for autocorrelation in home runs within parks across time.

Because higher temperatures could affect pitcher and hitter skill through mechanisms such as heat stress, air density may not be the only pathway by which global warming might alter home runs. To isolate the role of air density, we perform two analyses. Firstly, we estimate models with and without air density as a separate term (Table S5), which shows that when both temperature and air density are included, only air density has a statistically significant effect. Secondly, to control for batted ball characteristics, we analyze the Statcast data at the level of individual batted balls. These data include both the initial speed (sometimes called "exit velocity") and angle of line drives and fly balls as they come off the bat. Controlling for launch

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-22-0235.1.

speed and launch angle holds the characteristics of the batted ball constant, so that environmental variables like temperature can only affect home runs after the ball is hit.

We estimate the following linear probability model for the binary outcome that a given fly ball is a home run (*P*), with balls indexed by *b*, *A* representing launch angle, and *S* representing launch speed:

$$P_{bpdy} = \beta_1 Tx_{pdy} + \beta_2 A_{bpdy} + \beta_3 A^2_{bpdy} + \beta_4 S_{bpdy} + \mu_p + \gamma_d + \delta_y + \epsilon_{bpdy} \quad (2)$$

The coefficient of interest is β_1, which is the change in probability that a ball is a home run for a 1-°C change in temperature, after controlling for: (1) the fixed characteristics of the ballpark, time of year, and year (as in Eqn. 1); and (2) the initial characteristics of each batted ball before it must fly through the air and therefore be affected by air resistance. We use a quadratic in launch angle because batted balls that are both too low (line drives) and too high (pop-ups) are unlikely to be home runs, though using only a linear term for launch angle yields similar results for β_1. We cluster standard errors at the game level because the treatment (temperature) varies by game, not by individual batted ball (Abadie et al., 2017).

Historical home run attribution

To determine the contribution of historical global warming to observed home run trends, we compare the historical and historical-nat ("natural") simulations from CMIP6 over 1962-2019 (Table S7). Note that because the historical simulations extend only to 2014, whereas the natural simulations extend through 2019, we splice the historical simulations with years 2015-2019 from each model's corresponding SSP2-4.5 simulation, following the CMIP6 experimental protocol (Gillett et al., 2016).

We use a "delta method" to bias-correct the simulations (Diffenbaugh et al., 2021). For each day in each simulation, we take the difference between the historical and natural simulations. We then subtract this difference from the observed temperature on that day to calculate the "counterfactual" temperature. We then apply the regression coefficients to the observed and counterfactual temperatures on game days to calculate the difference in home runs due to warming.

We use distinct coefficients for day and night games and games played with and without domes (Table S1). For games where the status of the retractable-roof domes is "unknown" (~17% of games played in retractable-roof stadiums, ~3.1% of all games), we make the

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

conservative assumption that the dome was closed on those days. As such, our estimates of historical home runs due to warming are likely a lower bound on the true number.

Future home run projections

Future home runs are projected by comparing daily maximum temperatures from the SSP scenarios over 2020-2100 to the 2000-2019 average in each park. We preserve the seasonal cycle in the climatology by calculating the average 2000-2019 temperature for each day of year for each park from the HadISD data. This comparison quantifies how future home runs evolve relative to a counterfactual scenario in which temperatures remain at their early-21st-century average.

We use a similar delta method as in the historical analysis. We take the difference between the model-projected temperature in each game and subtract the model-simulated climatological 2000-2019 temperature for the corresponding day of year. We then add this difference to the observed 2000-2019 climatology. Additional home runs due to warming are calculated by comparing the bias-corrected model-projected temperature for each game day to the corresponding observed climatological temperature and applying the regression coefficients to calculate the effect of this difference on home runs.

Projections are referenced to global mean surface temperature (GMST) change where appropriate. GMST change is calculated as the average of global surface temperature relative to 1850-1900, spatially weighted by the cosine of latitude. Where we show global warming levels (GWLs), we define them as the first year that a thirty-year running mean of GMST change crosses each level, similar to the method used by the Intergovernmental Panel on Climate Change (Lee & Marotzke, 2021). As an analytical strategy, benchmarking projections to GWLs neglects the time scale and pace of future emissions. However, the quasi-linear relationship between temperature change and future home runs (Fig. 3b) supports our strategy of calculating cumulative home runs at GWLs regardless of the specific emissions scenario.

Uncertainty in the climate model analysis

In both the historical attribution and future projections, uncertainty is sampled with a Monte Carlo analysis. In the historical attribution analysis, we run 1000 simulations, where each

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

simulation randomly selects one climate model realization and one set of regression coefficients from their respective distributions.

In the case of the future projections, it is unknown how future baseball games will be distributed throughout the baseball season (approximately April through September). 81 games are played in each park in each regular season, but the effect of warming may depend on the times during the year that those games are played. As such, we simulate a series of future baseball seasons, randomly distributing 81 games on dates from April through September in each of the stadiums used by the 30 MLB teams (Table S6). These schedule simulations are performed 250 times, and in each iteration, a random set of the games are chosen as day games and the rest are chosen as night games. While the exact dates chosen as day and night games is random, the fraction of day and night games is set to equal the average 2000-2019 fraction in each park (Table S6). Similarly, for retractable-roof stadiums, the fraction of games where the dome is closed is set to equal the average 2000-2019 fraction (Table S6). These simulated schedules do not perfectly represent how MLB schedules will be set in the future, which is unknown, but they allow us to quantify the magnitude of the change in home runs driven by warming assuming similar gameplay conditions as today.

Then, for each SSP scenario, we run 250 simulations where each simulation randomly selects one set of regression coefficients from their distributions, one climate model realization, and one simulated schedule. With four scenarios, this yields 1000 simulations in total.

In both the historical and future cases, we adjust the sampling probability for models with multiple realizations so that each model is equally likely to be sampled and models with multiple realizations do not dominate the analysis. The coefficient samples are from a Gaussian distribution (see *Regression analysis*) and the schedule samples are from a uniform distribution.

Home run definition

As defined by Major League Baseball, a home run is any fair ball where the batter makes a complete circuit of the bases. This definition includes both out-of-play home runs (balls hit over the outfield fence, the vast majority of home runs) and inside-the-park home runs (fair balls inside the field of play where a player is able to run around all the bases before being put out). Increased temperatures and their effects on air density are likely affect out-of-play home runs

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

substantially more than inside-the-park home runs, since inside-the-park home runs do not necessarily need to fly several hundred feet in the air.

The Retrosheet game summary data we use does not distinguish between these two types of home runs, making it imperfect for isolating the effect of temperature on home runs through the mechanism of reduced air density. However, because inside-the-park home runs are extremely rare—less than 1% of all home runs—conflating these two types of home runs in the data should not substantially affect the analysis. Additionally, because the more granular Statcast data allows us to limit our analysis to only fly ball home runs, the estimates using those data are not subject to this limitation.

Acknowledgements

This work was supported by National Science Foundation Graduate Research Fellowship #1840344 to C.W.C. and grants from Dartmouth's Neukom Computational Institute and the Nelson A. Rockefeller Center to J.S.M.

Author contributions

C.W.C. conceived the study and performed the analysis. C.W.C. and J.S.M. designed the analysis and interpreted the results. N.J.D. and J.M.D. provided insights to aid the analysis and interpretation of results. C.W.C. and J.S.M. wrote the paper with all authors providing feedback and edits.

Competing interests

The authors declare no competing interests. The authors received no funding or endorsement from Major League Baseball or any affiliated organization.

Data and code availability

All data and code that support the findings of this study will be made available upon publication at github.com/ccallahan45/Callahan-et-al_ClimateBaseball_2023.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When Should You Adjust Standard Errors for Clustering?* (Working Paper No. 24003; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w24003

Adair, R. K. (2002). *The Physics of Baseball* (3rd ed.). HarperCollins.

Adler, R., Wang, J.-J., Sapiano, M., Huffman, G., Bolvin, D., Nelkin, E., & NOAA CDR Program. (2017). *Global Precipitation Climatology Project (GPCP) Climate Data Record (CDR), Version 1.3 (Daily)*. NOAA National Centers for Environmental Information.

Albert, J., Bartroff, J., Blandford, R., Brooks, D., Derenski, J., Goldstein, L., Hosoi, A., Lorden, G., Nathan, A., & Smith, L. (2018). *Report of the Committee Studying Home Run Rates in Major League Baseball*. Office of the Commissioner of Baseball.

Albert, J., Hosoi, A., Nathan, A., & Smith, L. (2019). *Preliminary Report of the Committee Studying Home Run Rates in MLB*. Office of the Commissioner of Baseball.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bahill, A. T., Baldwin, D. G., & Ramberg, J. S. (2009). Effects of Altitude and Atmospheric Conditions on the Flight of a Baseball. *International Journal of Sports Science and Engineering*, *3*(2), 109–128.

C3S. (2017). ERA5: Fifth generation of ECMWF atmospheric reanalysis of the global climate. *Copernicus Climate Change Service Climate Data Store*.

Clapeyron, E. (1835). Mémoire sur la puissance motrice de la chaleur. *Journal de l'École Polytechnique*, *XIV*, 153–190.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

Diffenbaugh, N. S., Davenport, F. V., & Burke, M. (2021). Historical warming has increased U.S. crop insurance losses. *Environmental Research Letters*, *16*(8), 084025. https://doi.org/10.1088/1748-9326/ac1223

Dunn, R. J. H. (2019). *HadISD version 3: Monthly updates*. Met Office Hadley Centre.

Dunn, R. J. H., Willett, K. M., Parker, D. E., & Mitchell, L. (2016). Expanding HadISD: Quality-controlled, sub-daily station data from 1931. *Geoscientific Instrumentation, Methods and Data Systems*, *5*(2), 473–491. https://doi.org/10.5194/gi-5-473-2016

Dunn, R. J. H., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., & Vose, R. S. (2012). HadISD: A quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate of the Past*, *8*(5), 1649–1679. https://doi.org/10.5194/cp-8-1649-2012

Dykstra, P. (2012, April 30). Is Global Warming Causing More Home Runs in Baseball? *Scientific American*. https://www.scientificamerican.com/article/global-warming-more-baseball-home-runs/

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Eyring, V., Gillett, N., Achuta Rao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., & Sun, Y. (2021). Human Influence on the Climate System. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 423–552).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

Gillett, N., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B., Stone, D., & Tebaldi, C. (2016). The Detection and Attribution Model Intercomparison Project (DAMIP v1. 0) contribution to CMIP6. *Geoscientific Model Development*, *9*, 3685–3697.

Howe, A. S., & Boden, B. P. (2007). Heat-Related Illness in Athletes. *The American Journal of Sports Medicine*, *35*(8), 1384–1395. https://doi.org/10.1177/0363546507305013

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, *8*, 43–75.

Hsiang, S. M., & Jina, A. S. (2014). The causal effect of environmental catastrophe on long-run economic growth: Evidence from 6,700 cyclones. *National Bureau of Economic Research Working Paper*.

IPCC. (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability*. Cambridge University Press.

Koch, B. L. D., & Panorska, A. K. (2013). The Impact of Temperature on Major League Baseball. *Weather, Climate, and Society*, *5*(4), 359–366. https://doi.org/10.1175/WCAS-D-13-00002.1

Kraft, M. D., & Skeeter, B. R. (1995). The Effect of Meteorological Conditions on Fly Ball Distances in North American Major League Baseball Games. *The Geographical Bulletin*, *37*(1), 40–48.

Lee, J.-Y., & Marotzke, J. (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., &

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.

Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Pachauri, R. K., & Meyer, L. A. (2014). *Climate Change 2014: Synthesis Report.* IPCC.

Samenow, J. (2012, April 30). Tim McCarver isn't crazy: The home run and global warming connection. *Washington Post*. https://www.washingtonpost.com/blogs/capital-weather-gang/post/tim-mccarver-may-not-be-crazy-the-home-run-and-global-warming-connection/2012/04/30/gIQA1hI1rT_blog.html

Schreck, L., & Sickles, E. (2018). Offensive Explosion: Trends in Hitting Production in Major League Baseball. *The Baseball Research Journal*, *47*(2), 19–28.

Smith, A., Lott, N., & Vose, R. (2011). The Integrated Surface Database: Recent Developments and Partnerships. *Bulletin of the American Meteorological Society*, *92*(6), 704–708. https://doi.org/10.1175/2011BAMS3015.1

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., … Ziehn, T. (2021). Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293. https://doi.org/10.5194/esd-12-253-2021

Vincent, D. (2007). *Home Run: The Definitive History of Baseball's Ultimate Weapon*. Potomac Books. https://www.nebraskapress.unl.edu/potomac-books/9781597970358

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-22-0235.1.