The Effect of Temperature on Home Run Production (Revisited)

Alan M. Nathan*

Department of Physics, University of Illinois, Urbana, IL 61801

(Dated: April 12, 2023)

Back in 2017 I wrote an article examining the effect of temperature on home run production. The motivation was the first two games of the 2017 World Series, where the temperature in Los Angeles was over 90°, raising the obvious question about how that elevated temperature might affect home runs. In my highly simplified analysis, I considered a typical long fly ball resulting in a home run, with exit velocity of 104 mph and launch angle of 28°, which travels about 400 ft. Using my Trajectory Calculator, which is based on analysis of Statcast fly ball distances, I concluded that a 1°F rise in temperature results in about 0.33 ft additional distance on such a fly ball. Then using Statcast data to determine the probability of a home run for a given distance, I found that the increased distance resulted in an increase in home run production by about 1%. That is an incredibly simple result: 1°F leads to 1% more home runs, or if you prefer, 1°C results in 1.8% more home runs.

The basic physics is straightforward. When the baseball travels through the air, it experiences many, many collisions with air molecules, losing a tiny bit of energy with each encounter. Of course, there are lots of encounters, all resulting in a loss of forward speed of the ball. This is the phenomenon known popularly as air resistance, or more technically as "drag". When the temperature is higher, the air molecules move faster and sort of repel each other, reducing the air density. In effect, the ball encounters fewer of these molecules, resulting in less loss of speed, longer distances, and more home runs.

That is how things stood until the week of April 3, 2023, when I was deluged with reporters–mostly science writers–wanting me to comment on a new study that was about to be published on April 7 in the Bulletin of the American Meteorological Society. The article entitled Global warming, home runs, and the future of America's Pastime was written by a

* a-nathan@illinois.edu

group of researchers from Dartmouth, with lead author Christopher Callahan. One result of interest was their estimate that nearly 600 home runs since 2010 can be attributed to global warming. Another interesting result is their use of climate models to estimate that by the end of the century, several hundred home runs per season will be attributable to global warming. However, the result that interested me most was that, based on their analysis of Statcast data from 2015-2019, they reached almost the exact conclusion I had reached in 2017 despite using a very different analysis technique: 1°C rise in temperature results in 1.8% more home runs.

In view of the Dartmouth work and given the simplicity of my own previous analysis, I decided to return to this issue, using a much more elaborate technique than I used in 2017. My interest is not so much the climate change implications but rather the question of how day-to-day changes in temperature affect home run production. While temperature might affect the game in many ways (e.g., pitcher and/or batter fatigue, changes to the coefficient of restitution of the ball, etc.), I am only interested in how it affects the distance the baseball travels once it leaves the bat. Therefore the variables I will concentrate on are the launch parameters (exit velocity and launch angle) and the temperature.

I considered the Statcast data from the period 2017-2022. Using a Generalized Additive Model (GAM) applied to these data, I did a logistic regression to find the dependence of home run probability on exit velocity, launch angle, gametime temperature, and year. I ignored any explict effects due to variation of the drag coefficient of the baseball, C_D. Instead any year-to-year variation in C_D will be taken into account implicitly by using the year as an additional covariate in the fitting. Only games played in open air were included.

The predicted number of home runs for each season agrees perfectly with the actual number. An example of applying the GAM model to the 2022 data is shown in Fig. 1. Note that the variation in home run probability with temperature is greatest in the mid-range of probabilities and very small when the probability is close to zero or one. This is exactly as one expects. For example, a ball hit over 110 mph and 27° launch angle will always be a home run, regardless of the variation of temperature within normal bounds. In this regard, I note that the Dartmouth group, while including a number of additional covariates in their fitting, used linear rather than logistic regression, which can be problematic when calculating probabilities that are bounded by 0 and 1.

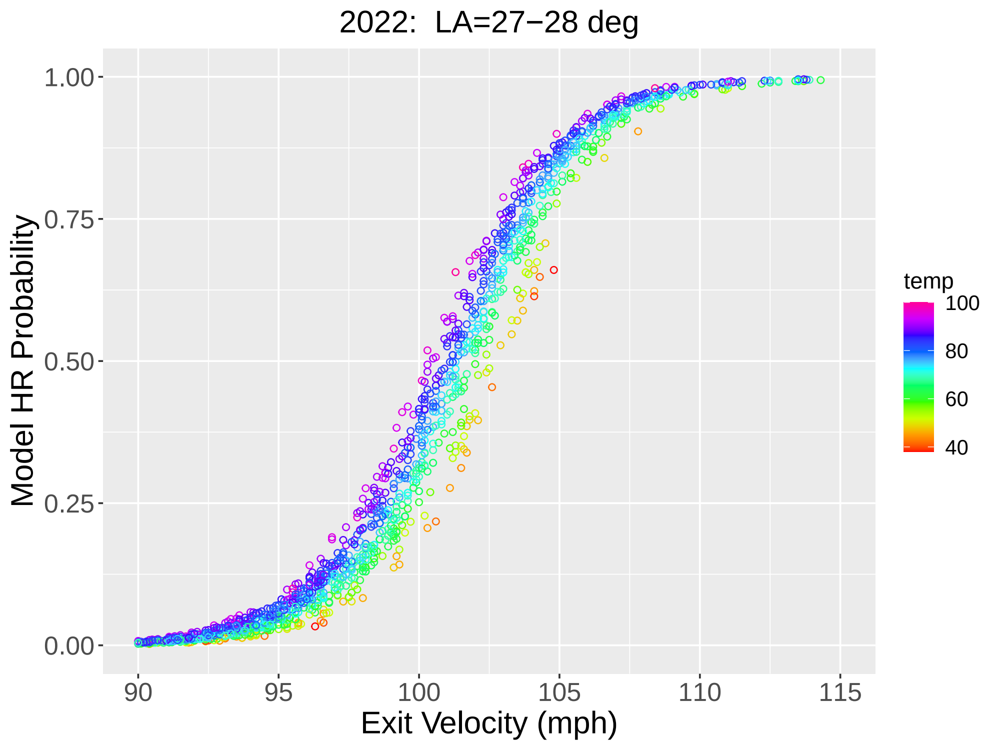

FIG. 1: Application of the Generalized Additive Model to the 2022 Statcast data, where the fitted HR probability is plotted vs. exit velocity, with launch angles fixed in the range 27°-28° and with the points color-coded by gametime temperature. These data clearly show that higher temperature leads to greater probability, especially for the mid-range of probabilities. For data at much larger exit velocity, the fly balls are "no doubters" so that the temperature has very little effect on the HR probability. For much lower exit velocities, the fly balls are "no wayers" and are also not affected by temperature.

The next step is key: I used the same GAM model described above to predict home runs for a modified data set, which is identical to the original data set except that the gametime temperatures were artificially increased or decreased by 5°F. That allows me to obtain the expected number of home runs with modified temperatures, based on the same distribution of launch parameters. The results of this analysis are shown in Table I and Fig. 2. The proverbial "bottom line" agrees perfectly with both my own earlier analysis and with the Dartmouth analysis: A 1°F rise in temperature results in a 1% increase in home runs.

TABLE I: Result of applying the model Statcast data for each season, with temperature modified by ΔT. The last column is the percent change in home runs per °F, with the bottom line the average over all six seasons.

Season	Actual HR	ΔT=0	ΔT=+5	ΔT=-5	ΔHR/°F
2017	5224	5226	5476	4981	0.95
2018	4742	4742	4986	4509	1.01
2019	5550	5552	5810	5297	0.92
2020	1913	1913	2005	1823	0.95
2021	4970	4971	5201	4742	0.92
2022	3538	3538	3725	3359	1.04
					0.96

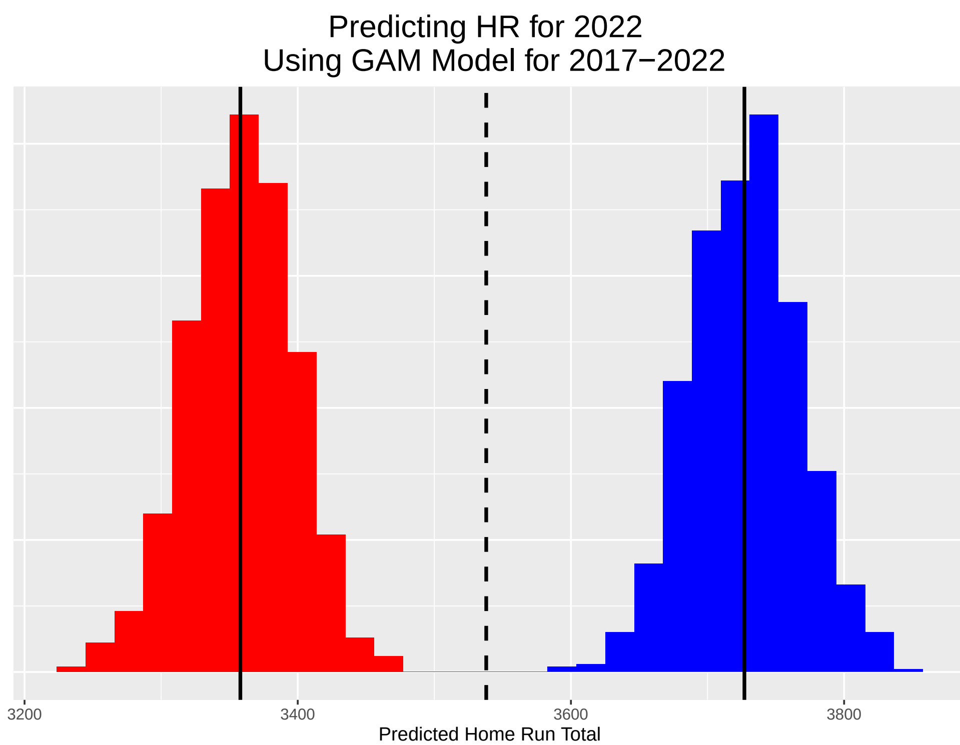

FIG. 2: Distributions of predicted home runs found by applying the GAM model for home run probabilities to the 2022 data. The histograms are Monte Carlo calculation with 1000 iterations, using the fitted probabilities applied to the temperature-modified data (blue/red for ±5°F. The actual number of home runs is given by the dashed line. The mean values of each histogram are indicated by the solid lines and given by the numbers shown in Table I

To investigate the robustness of the result, the following additional studies were done.

- Instead of changing the temperature by a constant $\pm 5^\circ$, I sampled for each day from a normal distribution, with mean of $\pm 5^\circ$ and a variety of standard deviations. The net result did not change.

- Rather than fit to all six seasons simultaneously, I fit to each season individually. The net result did not change.

- Rather than fit to all parks simultaneously, I fit to each park individually. The net result did not change.

The final result that a 1°F rise in temperature results in a 1% increase in home runs, has now been confirmed by three indepedent studies, all using very different analysis techniques. This result quantifies what baseball people have known all along: The baseball carries better when the temperature is higher, leading to more home runs. This effect will be most pronounced when comparing cold 40° days in April with hot 90° days in August, with the latter producing on average 50% more home runs than the former, all other things the same. That's a huge effect.

How this result plays into the climate change discussion is way beyond my area of expertise. Nevertheless, the worst-case prediction of nearly additional 500 home runs per year by the end of the century would mean a temperature rise of approximately 5°C. If that were to occur, I would expect far more serious consequences than a rise in home runs.

One final word: As I complete this article, I learn from Jim Albert's blog that he has done his own analysis using a negative binomial regression model to estimate the effect of temperature on home runs. He finds a result completely in accord with all the other studies.