A Statistical Study of PITCHf/x Pitched Baseball Trajectories

Alan M. Nathan*
Department of Physics, University of Illinois, Urbana, IL 61801
February 8, 2008

A statistical study of PITCHf/x trajectories is presented with two goals: (1) to determine the accuracy of the constant-acceleration fit that is used to parametrize the actual trajectory and determine various quantities of interest, such as the release velocity and break; and (2) to estimate the size of the random deviations of the coordinate measurements from their true values and their effect on the derived quantities.

I. INTRODUCTION AND SUMMARY OF RESULTS

The PITCHf/x system uses two cameras to track pitches between pitcher and batter, determining the coordinates of the ball x(t), y(t), z(t) at times t in 1/60-sec intervals. The resulting trajectory is a nine-parameter (or 9P) fit corresponding to constant acceleration in each of the three coordinates. The 9P fit is an approximate solution to the exact equations of motion. All quantities reported in the PITCHf/x data base, such as the pitch speed, the location of the pitch as it crosses the plate, the break (or pfx) of the pitch, etc., are derived from the fitted trajectory rather than from the original data. The question naturally arises as to the validity of the 9P fit and its effect on the derived quantities. It is also of interest to ask how large are the random fluctuations of the measured coordinates about the exact values and how these fluctuations propagate into the derived quantities. Both these questions are addressed in this note. Before proceeding with details of the study, it is useful to summarize the primary conclusions:

- For a broad selection of non-knuckleball pitches, the 9P fit does an excellent job reproducing the exact trajectory and the derived quantities.
- The intrinsic precision of PITCHf/x system in determining the coordinates of a pitched baseball is of order 1 inch, resulting in pitch-to-pitch deviations of the pfx break parameters of order 2.0-2.5 inches.

II. VALIDITY OF THE 9P FIT

We begin by investigating the validity of using the 9P fit as an approximation to the actual trajectory. As stated in the previous section, all of the reported quantities from the PITCHf/x tracking, such as the initial speed, the position of the ball as it crosses home plate, and the break, are derived from the 9P fit to the data. The nine parameters are the three initial positions x_0, y_0, and z_0; the three initial velocities v_{x0}, v_{y0}, v_{z0}; and the three (constant) accelerations a_x, a_y, and a_z. Here the coordinates refer to the usual PITCHf/x coordinate system, where the origin is at the point of home plate, $\hat{y}$ points towards the pitcher, $\hat{z}$ points vertically upward, and $\hat{x} = \hat{y} \times \hat{z}$ (i.e., the x axis points to the catcher's right).[1]

To investigate the overall accuracy of this method, an exact trajectory $r(t)$ (r refers generically to x, y, or z) is calculated by numerically solving the equations of motion

$$\begin{aligned} \ddot{x} &= -KC_D v v_x - KC_L v v_y \sin\phi \\ \ddot{y} &= -KC_D v v_y + KC_L v\,(v_x \sin\phi - v_z \cos\phi) \\ \ddot{z} &= -KC_D v v_z + KC_L v v_y \cos\phi - g\,, \end{aligned} \quad (1)$$

for given initial conditions. Here g is the acceleration due to gravity (32.174 ft/s^2), C_D and C_L are the drag and lift coefficients, respectively, and $K = 5.44\times10^{-3}$ ft^{-1} is a numerical factor.[2] The spin axis is assumed

*Electronic address: a-nathan@uiuc.edu
[1] For an explanation of the PITCHf/x system, see the web site webusers.npl.uiuc.edu/~a-nathan/pob/pitchtracker.html
[2] The factor K in defined in webusers.npl.uiuc.edu/~a-nathan/pob/Analysis.pdf.

TABLE I: Pitched ball parameters for the initial part of the study, with positions in ft., velocities in ft/s, and the ϕ in degrees.

x_0	y_0	z_0	v_{x0}	v_{y0}	v_{z0}	C_D	C_L	ϕ
2.7	50.	6.1	-7.	-133.	-6.	0.46	0.18	137.
2.6	50.	6.1	-8.	-135.	-6.	0.45	0.17	171.
2.7	50.	6.2	-7.	-127.	-4.	0.41	0.09	234.
2.8	50.	6.5	-5.	-108.	+1.	0.45	0.18	322.

to lie in the $x - z$ plane, making an angle ϕ with the x axis, with a sign such that $\phi = 90°$ corresponds to the spin pointing upward, along the z axis. Nine parameters are needed for the calculations: three initial positions, three initial velocities, and C_D, C_L, and ϕ. The trajectory is calculated at 1/60-sec intervals, with t=0 corresponding to the release point at $y = y_0$ and with the trajectory terminating at the front of home plate $y = 1.4$ ft.

A glance at the structure of Eq. 1 shows that the acceleration due to aerodynamic forces (the terms proportional to K) are proportional to the *square* of the velocity. As the drag reduces the magnitude of the velocity, the magnitudes of the accelerations also decrease. For a pitched baseball, the non-constancy of the acceleration is not expected to be a serious problem, since the velocity only varies by about 10% over the short flight distance between pitcher and home plate. Said differently, a constant-acceleration parametrization should be an excellent approximation to the actual trajectory. The purpose of this study is to quantify this statement.

Four different pitches from the arsenal of left-handed-pitcher Jon Lester were studied, with parameters given in Table I. The exact trajectory $r(t)$ is fitted to the 9P constant-acceleration function. An example of such a fit is shown in Fig. 1 (positions) and 2 (velocities) for the pitch parameters given in the first line of Table I. The following discussion refers explicitly to the first line but applies equally well to the other pitches in Table I. The fits are excellent, with the residuals (exact minus fit) typically less than about 1 inch over the entire flight path. In fact, for the transverse coordinates (x and z), the largest residuals are less than 0.05 ft (0.6 inches) and for y they are about twice as large. The shape of the residuals is exactly that expected from the neglect of a cubic term corresponding to a constant rate of change of acceleration. The consequences of assuming constant acceleration can also be studied by looking at the velocity as a function of time (see Fig. 2). For constant acceleration, the velocity would be linear in time. However, since the aerodynamic forces are proportional to the square of the velocity (see Eq. 1), which is reduced during flight because of the drag, the magnitude of the acceleration must decrease in time, leading to curvature in $v(t)$. As a consequence of fitting $v(t)$ to a straight line, the initial and final velocities are either both overestimated or both underestimated, but the average velocity is estimated nearly correctly. For the pitch shown, the largest deviation is about 0.4 ft/s (0.3 mph). For all the pitches in Table I, the deviation of the other derived quantities from their exact value were small in inconsequential. We conclude that the 9P fit works extremely well over the full trajectory and does not lead to any serious errors in the derived quantities.[3]

III. EFFECT OF RANDOM ERROR ON THE DERIVED PARAMETERS

We next investigate the effect of random measurement error on the derived parameters. To this end, a trajectory $r_M(t)$ simulating the PITCHf/x measurements is calculated by adding to the exact trajectory $r(t)$ a random number sampled from a Gaussian distribution with zero mean and standard deviation σ_r. This procedure is followed for each coordinate and at each time.[4] Next, the standard 9P constant-acceleration fit is applied to $r_M(t)$, to generate a fitted trajectory $r_F(t)$. Then the fitted trajectory is refitted to the exact equations of motion, Eq. 1, to obtain fitted values of C_D, C_L, and ϕ, for comparison with the values that were used to calculate the exact trajectory. These procedures were followed for the pitches listed in Table

---

[3] This conclusion may not apply to knuckleball pitches, which were not investigated in this study.

[4] In this study, only the effects of random deviations of the measured from the actual coordinate are investigated. An investigation of the effect of systematic deviations (for example, due to a miscalbration of a camera) will be reported in a separate note.

I. For each pitch, 2500 simulated trajectories were generated. All calculations assumed σ_r=1 inch; that is, the root-mean-square (rms) deviation of the measured from the actual coordinates is 1 inch for each of x, y, and z. The results of these investigations are shown in Figs. 3-7.

Fig. 3 shows the difference between the exact and fitted values of C_D, C_L, ϕ, and the initial velocity v_0. In each case, the mean value is close to zero, indicating that there is no systematic deviation from the exact value determined from the 9P fit. The spread about the mean value is the result of the random errors introduced into the simulated trajectory $r_M(t)$. The rms deviations of the fitted parameters are directly proportional to σ_r. For σ_r=1 inch, they are 0.03 for C_D and C_L, 10° for ϕ, and 0.4 mph for v_0. Are these rms values a reasonable representation of actual data? To answer this question, we specifically consider data for C_D and C_L. For a narrow range of velocities, C_D (and to a lesser extent C_L) should be approximately constant, especially for speeds over 90 mph. Any spread in experimental values of C_D and C_L are likely due to random experimental error. In Fig. 3 are plotted actual data for C_D and C_L taken from an analysis of a large set of PITCHf/x data from 2007 games played in Toronto. The values presented for C_D and C_L are for the narrow range of v_0=90-92 and 95-97 mph, respectively. In each case the mean value has been subtracted to facilitate comparision with the simulation. The rms value is 0.29 and 0.32 for C_D and C_L, respectively. These values are quite close to the simulation, leading us to conclude that the PITCHf/x data are consistent with $\sigma_r \approx 1$ inch, at least for the Toronto venue.

Having established that σ_r=1 inch is a reasonable estimate of the random measurement error, we next ask how that error propagates into other quantities of interest. Fig. 4 shows the difference between the exact and fitted values of x_f, z_f (the location of the pitch as it crosses home plate) and pfx_x, pfx_z (the "break", or deviation of the pitch from a straight-line trajectory due to the Magnus force). Once again, the mean values are close to zero indicating no large systematic errors. For x_f and z_f, the rms values are small, ≈0.55 inches. That is, from pitch to pitch, there are experimental uncertainties of order 0.5 inch in the determination of the location of the pitch as it crosses the plane of home plate. We do not consider these deviations to be significant. The rms values are much larger for pfx_x and pfx_z, ≈2.3 inches, meaning that from pitch to pitch the break has experimental uncertainties in the range in the range 2-2.5 inches. The pfx_x and pfx_z values for each of the four pitches in Table I are presented Fig. 5 and 6, respectively. As a reality check, the values of pfx_z for the Toronto data are also shown in Fig. 6. The rms value of the data (2.0 inches) is comparable to that of the simulation (2.3 inches). We conclude that our statistical study reasonably well predicts the random error in the pfx values.

As expected, the inferred values of pfx_x, C_D, and pfx_z are perfectly correlated with the inferred values of the accelerations, a_x, a_y, and a_z, respectively, as shown in Fig. 7, which also shows a histogram of acceleration values. The rms of the accelerations (generically denoted by σ_a) are independent of the accelerations themselves and scale linearly with σ_r; that is, doubling σ_r will double the σ_a, as well as the rms values of the derived quantities C_D, C_L, pfx_x, and pfx_z. Simple statistical considerations[5] show that the rms values of the pfx distributions are expected to depend only weakly on the initial velocity (for a given tracking distance and frame rate), an expectation that is verified by the similarity among the different pitches in Fig. 5 and 6.

We conclude that the random measurement errors in the PITCHf/x system lead to random pitch-to-pitch errors in the derived values of pfx_x and pfx_z of order 2-2.5 inches. Of course, when averaging these quantities over N pitches, the error in the determination of the mean values are reduced by $1/\sqrt{N}$.

---

[5] For example, see P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 2003), pp. 127-132.

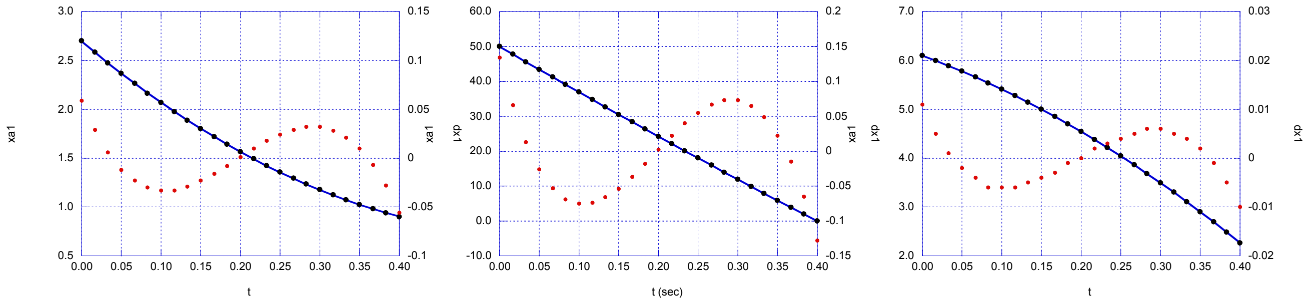


FIG. 1: Comparision between the exact trajectory (points) and the 9P constant-acceleration fit (solid curve) for each of the three coordinates. The dashed curves show the residuals (exact minus fit). Note that the trajectory uses the left-hand scale and the residuals the right-hand scale. The input parameters are those in the first line of Table I.

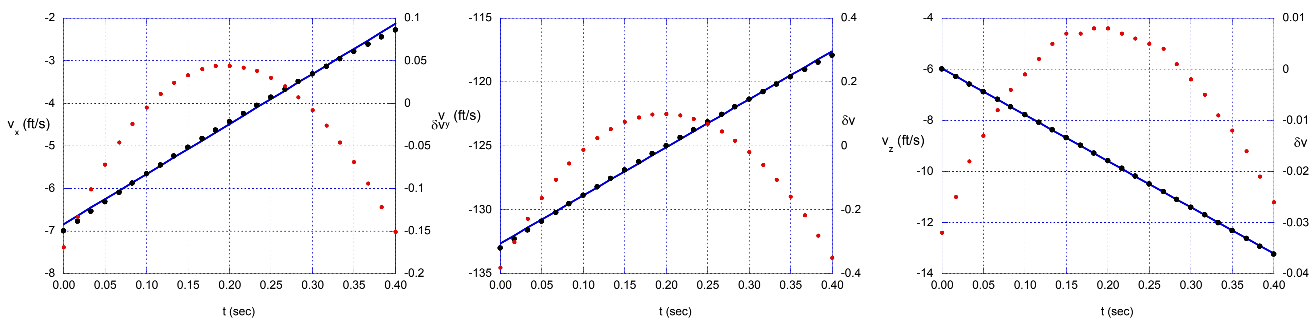
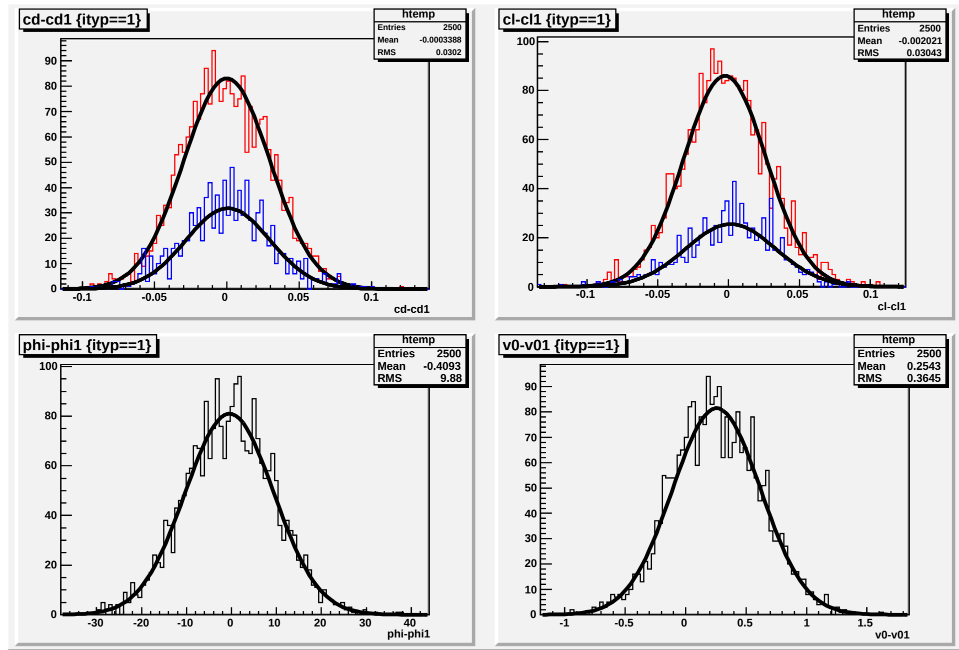


FIG. 2: Comparision between the exact velocity (points) and that determined from the 9P constant-acceleration fit (solid curve) for each of the three coordinates. The dashed curves show the residuals (exact minus fit). Note that the velocities use the left-hand scale and the residuals the right-hand scale. The input parameters are those in the first line of Table I.



]

FIG. 3: Histograms of difference between the fitted and exact values of C_D (upper left), C_L (upper right), and ϕ (lower left, in degrees), and v_0 (lower right, in mph). The C_D and C_L plots show both the simulation (red) and actual data (blue). These results are for the first pitch listed in Table I. For the actual data, the histograms show the difference between the actual value and the mean value. The curves are Gaussian fits to the histograms.

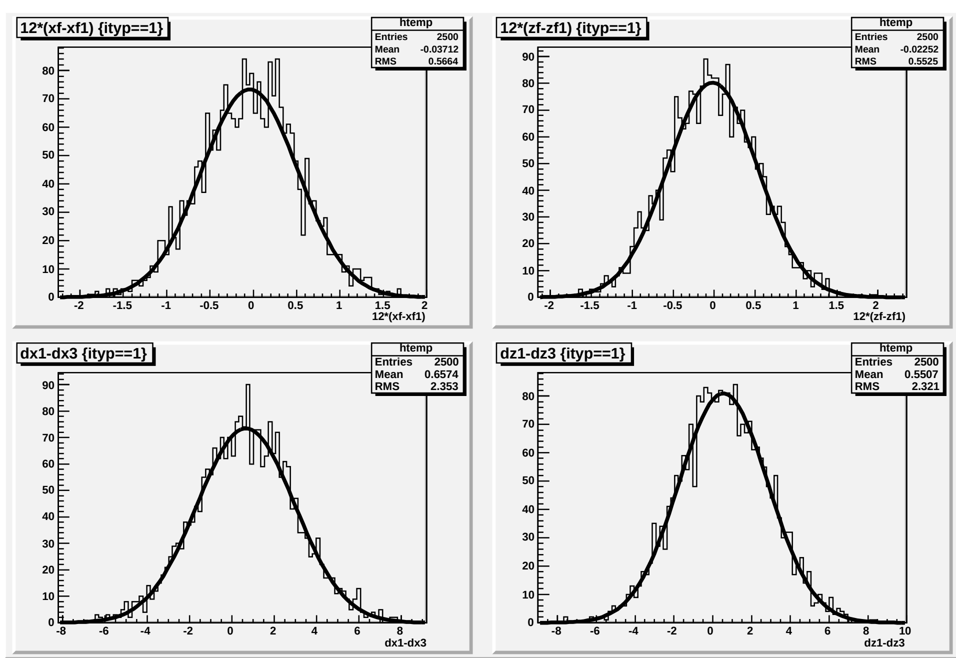


FIG. 4: The difference between fitted and exact values for various quantities as follows: x (upper left) and z (upper right) location of the pitch as it crosses home plate; pfx_x (lower left) and pfx_z (lower right), the break due to the Magnus force. All values are in inches. These results are for the first pitch listed in Table I. The curves are Gaussian fits to the histograms.

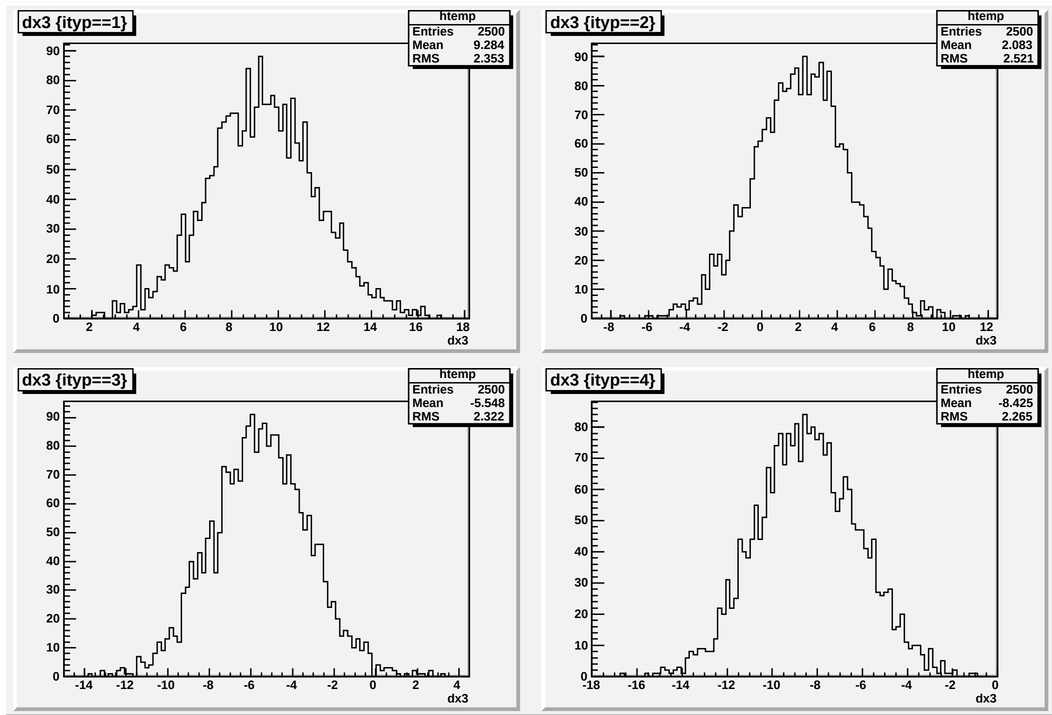


FIG. 5: Histogram of the difference between exact and fitted values of pfx_x for each of the four pitches in Table I.

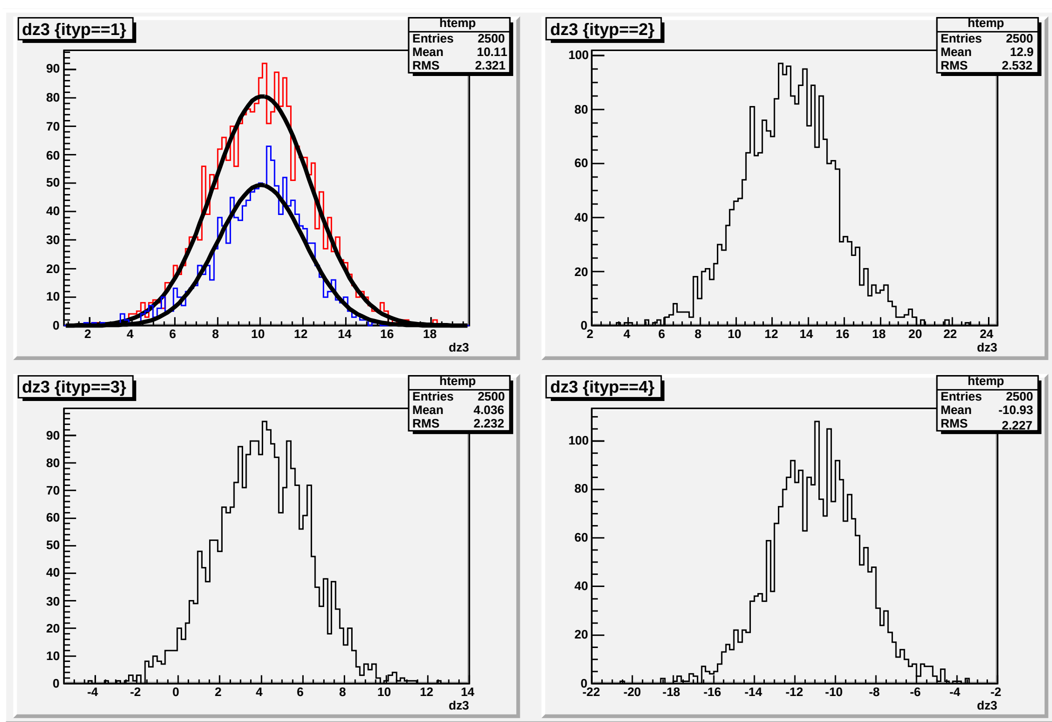


FIG. 6: Histogram of the difference between exact and fitted values of pfx_z for each of the four pitches in Table I. The upper left plot shows both the simulated (blue) and measured (red) values of pfx_z from the Toronto data, the latter assuming v_0 >94 mph, along with Gaussian fits.

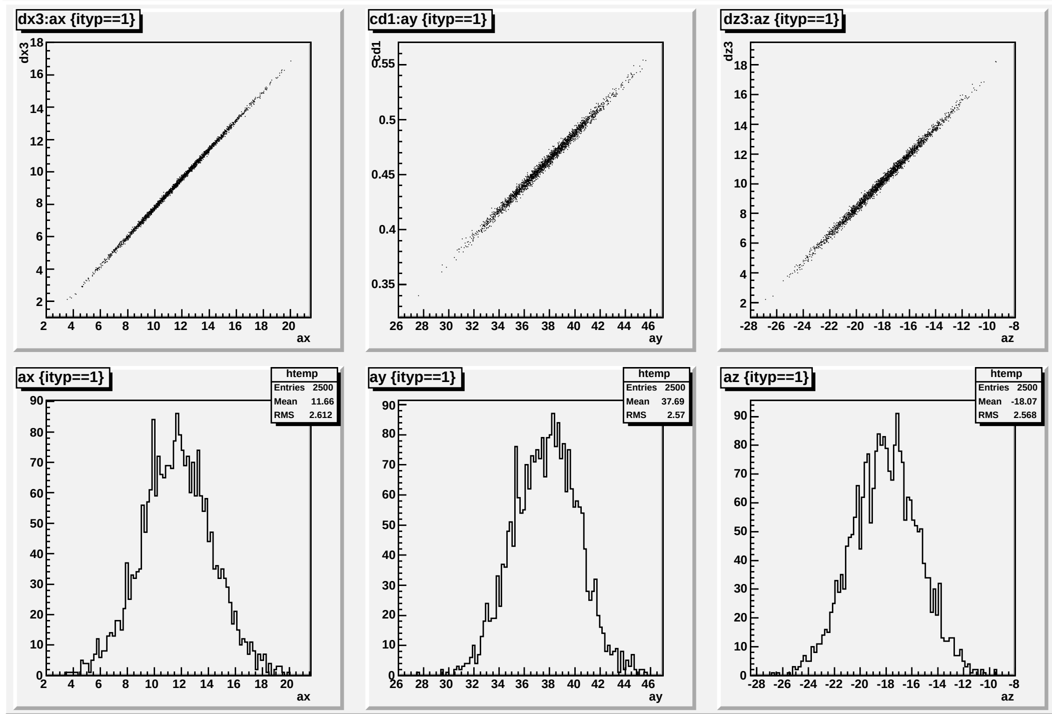


FIG. 7: Upper plots: Scatter plots of pfx_x vs. a_x (left), C_D vs. a_y (center), and pfx_z vs. a_z (right). Lower plots: Histograms of a_x (left), a_y (center), and a_z (right). These results are for the first pitch listed in Table I. The units of pfx are inches and of a are ft/s^2.