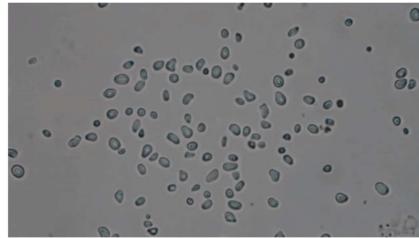


Characterization of Pollen Grain Motion in Water



Linda Oster, Sasha Bakker, Faizah Siddique

Department of Physics
University of Massachusetts Amherst 01003



Vid link: <https://www.youtube.com/watch?v=d9h4SnwSnf8>

Our presentation is about the Characterization of Pollen Grain Motion in Water.

I worked on this experiment with Linda Oster and Faizah Siddique

The video at the bottom is the one in which we are analyzing. Here, you see a top view of pollen grains suspended in water that are wiggling over time

Background and Motivation

1827: Botanist Robert Brown observes pollen grains moving *randomly* in water

1905: Albert Einstein argues the movement is governed by atoms and molecules

1908: Jean Perrin experimentally confirms Einstein's equations

Since then: biological applications (diffusion w/in the body), prediction of weather patterns, stock market behavior

Newburgh, R., Peidle, J., & Rueckner, W. (2006). Einstein, Perrin, and the reality of atoms: 1905 revisited. *American Journal of Physics*, 74(6), 478-481. doi:10.1119/1.2188962

Ramaswamy, S. (2000). Pollen Grains, Random Walks and Einstein. *Resonance*, 5(3), 16-34. doi:10.1007/bf02838997

The Editors of Encyclopaedia Britannica. (2020, April 13). Jean Perrin. Retrieved August 29, 2020, from <https://www.britannica.com/biography/Jean-Perrin>

1827: Botanist Robert Brown observes pollen grains moving randomly in water

1905: Albert Einstein argues the movement is governed by atoms and molecules, proving the existence of kinetic particle theory. Specifically, he calculated the Mean Squared Distance of particles suspended in a liquid.

1908: Jean Perrin experimentally confirms Einstein's equations for brownian motion and estimates the size of water molecules and atoms

Objective

Research Question: Can we confirm that pollen grains in water exhibit brownian motion?

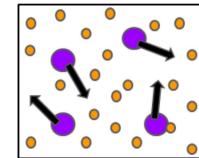
We expect:

- The step lengths to be randomly and normally distributed
- The probability of step direction equal for all directions
- The path should be a random walk
 - No spatial or temporal correlation

Theory

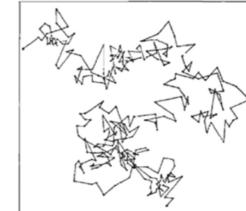
Kinetic Theory of Matter: Atoms constantly vibrate, elastically collide

$$\frac{1}{2} mv^2 = \frac{3}{2} k_B T$$



Stoke's Law: The drag force exerted on a spherical object in a viscous fluid is determined by the radius 'r', viscosity η , and velocity 'v'

$$F = 6\pi\eta rv$$



Brownian Motion: Particles suspended in fluid undergo random motion

Newburgh, R., Peidle, J., & Rueckner, W. (2006). Einstein, Perrin, and the reality of atoms: 1905 revisited. *American Journal of Physics*, 74(6), 478-481.
doi:10.1119/1.2188962

Ramaswamy, S. (2000). Pollen Grains, Random Walks and Einstein. *Resonance*, 5(3), 16-34. doi:10.1007/bf02838997

Kinetic Theory of Matter: Atoms and molecule are constantly vibrating and undergoing elastic collisions. By this process, an increase in temperature leads to an increase in Kinetic Energy.

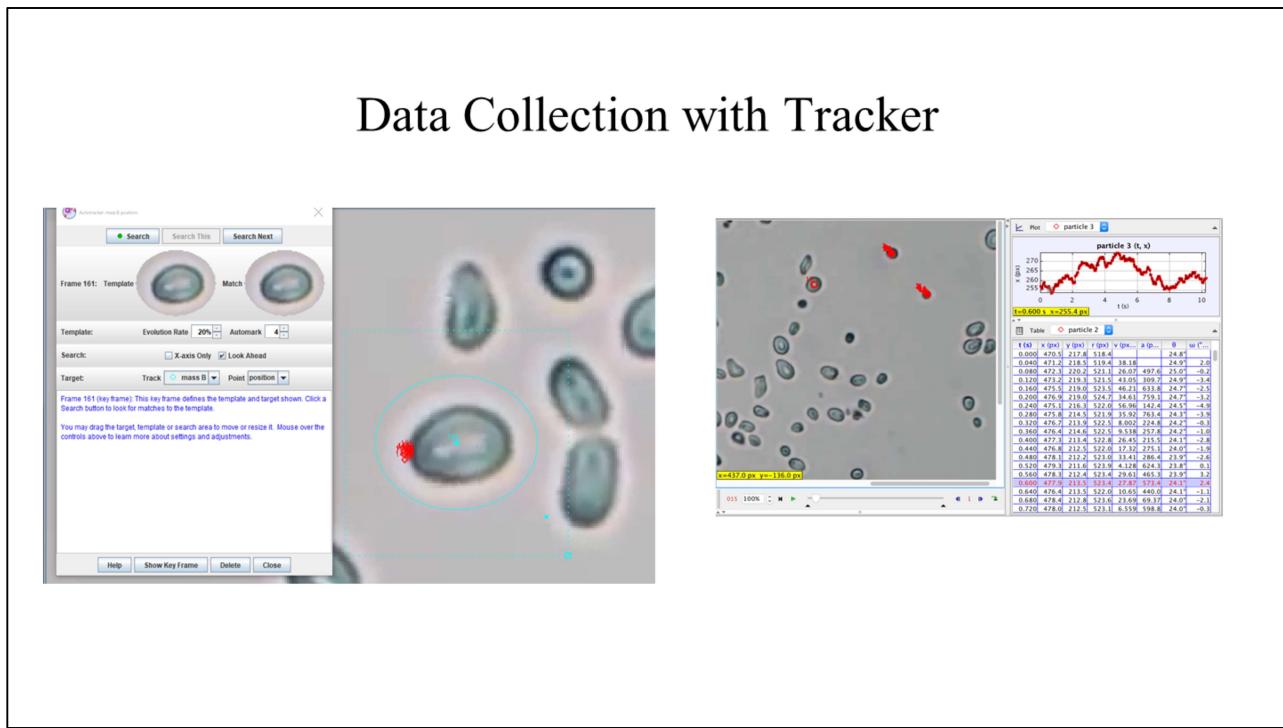
Stoke's Law: A frictional force is exerted on a spherical objects with very small Reynolds number in a viscous fluid. Meaning, viscous forces are dominant in Stoke's regime with the motion characterized by smooth, constant fluid motion.

Brownian Motion: Particles suspended in a fluid have a 'random walk path' due to the collisions between the particles and the surrounding fluid molecules.

Methods and Preview

1. Tracked 28 particle trajectories using Tracker 5.1.5
 - a. Collected x, y, t for each particle
2. Calculated D coeff. 2 ways
 - a. Histogram method
 - b. MSD
3. Plotted probability for step direction in X, Y
4. Viewed effect of time interval on D coeff.
5. Explored relationship of D with:
 - a. Particle area
 - b. Circularity
 - c. High and low density regions

Data Collection with Tracker



Autotracker → creates 1 or more template images and searches each frame for the best match to that template.

The best match is one with the best match score (# inversely proportional to the sum of the squares of the RGB differences between the template and match pixels).

Compared with nearby scores to determine an interpolated **sub-pixel best match** position.

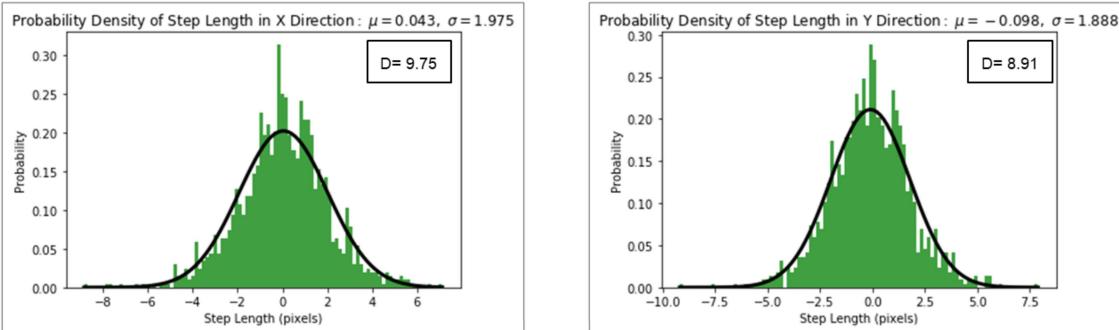
A **target** defines the position at which points are marked relative to the template when matches are found. The target may be offset from the template. The specific **track** and **point** marked at the target position must be selected prior to autotracking.

Probability Density of 1D Displacement:

$$P(\Delta) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\Delta^2/2\sigma^2} \quad \sigma = \sqrt{2D\tau}$$

$$D = \sigma^2 / 2\tau$$

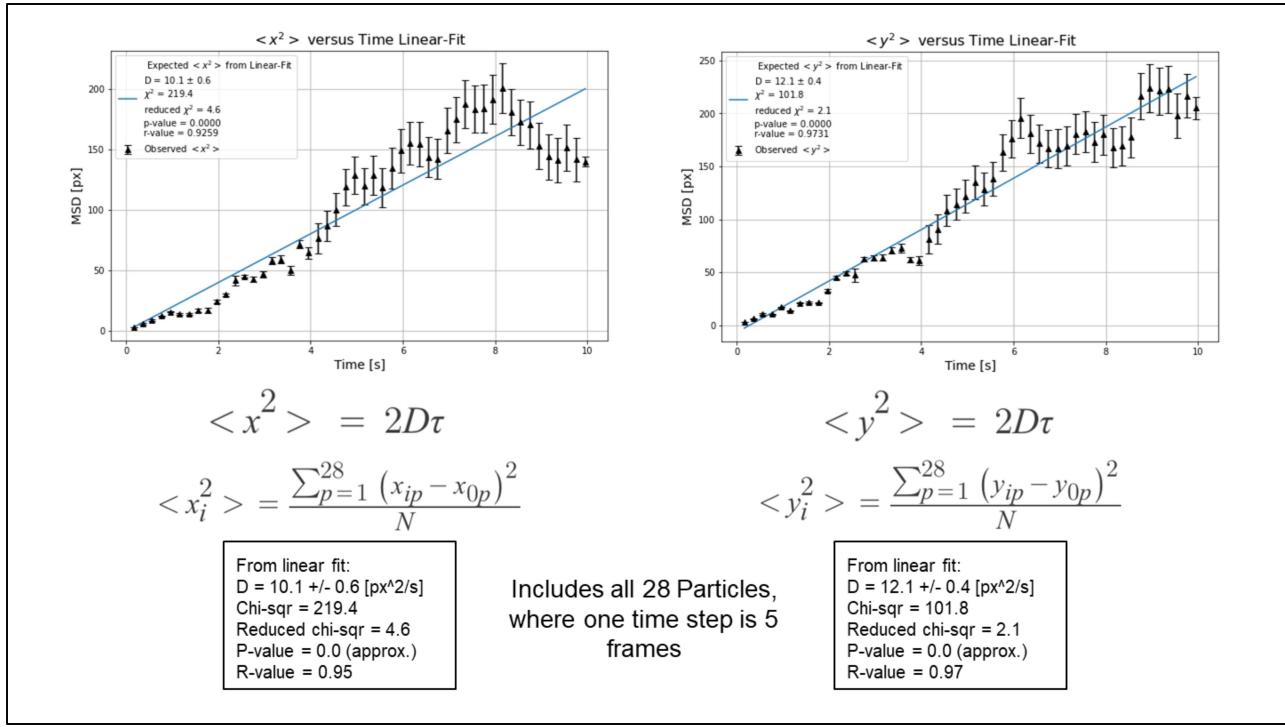
Includes all 28 Particles,
where one time step is 5 frames



Histogram Method:

- Step length is defined as the difference in x or y position of a particle between two frames
- We make a histogram of all of the step length for every particle in both x and y and verify that it is normally distributed
- We make histograms for every individual particle of the step length in the x and y direction. For these histograms, we compute the standard deviation which allows us to calculate the D coefficients in x and y for every individual particle.
- While this is the histogram for all 28 particles, we also made histograms for all the particles individually, and used sigma to calculate diffusion constants for each of the 28 particles. These values will be discussed later.

We picked 5 frames per interval for our time step because it allows for a larger range of step lengths. If we chose just one frame per interval then the step lengths would be 1 pixel or less which is within the uncertainty of tracking the particle's position.



MSD Method:

- We compute the squared distance from the initial position for x and y. We do this for every individual particle 'p'
- From these squared distances, we can compute the mean squared distance in x and y at every time
- We then plot the MSD vs. time and the slope of the linear fit is the dimensional coefficient 2 multiplied by the diffusion coefficient
- We then use the slope to solve for the diffusion coefficient.
- This is done by Taking the slope from plots of individual particles to get the diffusion coefficient of each individual particle.

Statistics:

- The reduced chi-square for these data signifies this linear fit is good for most of the observations
- The p-value is less than 0.05 thus we can reject the null hypothesis that the slope is zero
- The correlation coefficient is close to one which signifies an almost perfect positive linear relationship

Large Error Bars with Increasing Time: Since our MSD values are calculated by subtracting the initial position from its current position and squaring that difference, the uncertainty would scale in the same way. Since these MSD values get larger with time, so will their uncertainties.

Mean Square Displacement (MSD) versus Time Fits

$$\langle r^2 \rangle = 4D\tau$$

$$\langle r_i^2 \rangle = \frac{\sum_{p=1}^{28} (x_{ip} - x_{0p})^2 + (y_{ip} - y_{0p})^2}{N}$$

From linear fit:

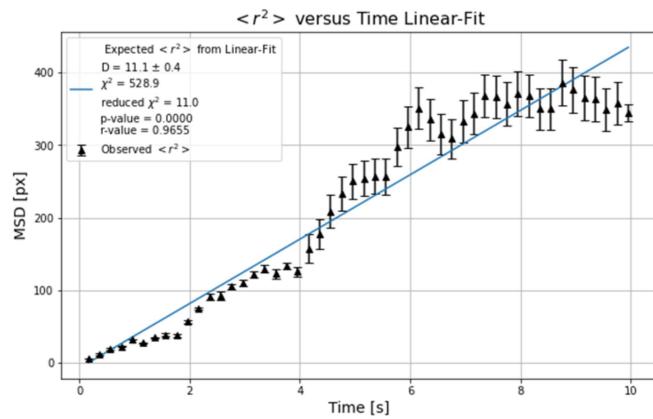
$D = 11.1 \pm 0.4 \text{ [px}^2/\text{s}]$

Chi-sqr = 528.9

Reduced chi-sqr = 11.0

P-value : 0.0 (approx.)

R-value: 0.97



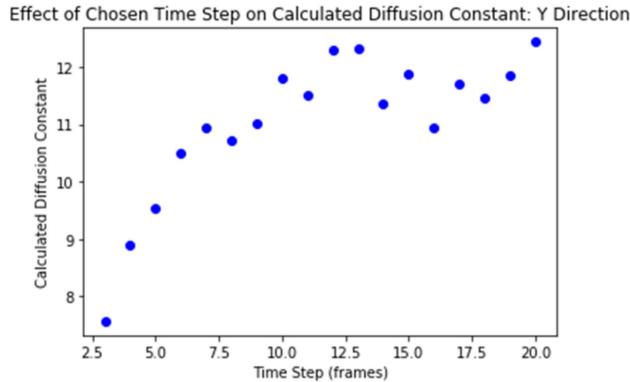
MSD Method:

- We compute the squared distance from the initial position for r . We do this for every individual particle 'p'
- From these squared distances, we can compute the mean squared distance in r at every time
- We then plot the MSD vs. time and the slope of the linear fit is the dimensional coefficient 4 multiplied by the diffusion coefficient
- We then use the slope to solve for the diffusion coefficient.
- This is done by either summing over all particles to get a diffusion coefficient for the system

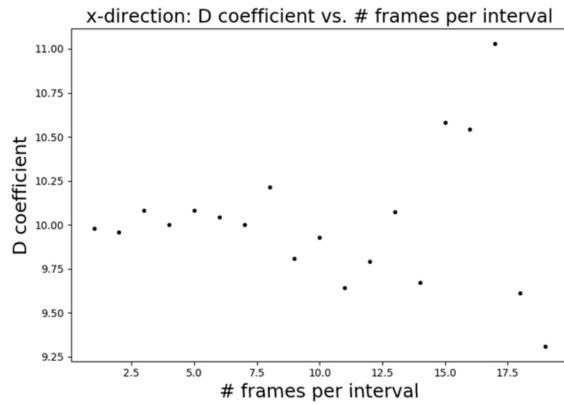
Statistics:

- The reduced chi-square for this data signifies that the linear fit has not fully captured the data
- The p-value is less than 0.05 thus we can reject the null hypothesis that the slope is zero
- The correlation coefficient is close to one which signifies an almost perfect positive linear relationship

Histogram Method



MSD Method



Both Methods: Calculated the D coefficient from all 28 particles in the system at different time steps. This “time step” signifies the number of frames per interval. For example, we have been using 5 frames per interval which means that we only use every 5th frame in the video to calculate the different values of x^2 , y^2 , r^2 .

Histogram Method: We see that the calculated diffusion coefficient levels off to a diffusion coefficient of approx 12 for larger step size

MSD Method: We see that the calculated diffusion coefficient seems to vary about an approximate mean of 10 near the beginning and its accuracy decreases (more spread in the value) as the time step increases

D is invariant for the Histogram method approx $\tau > 10$

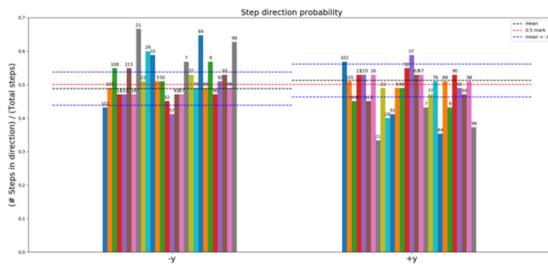
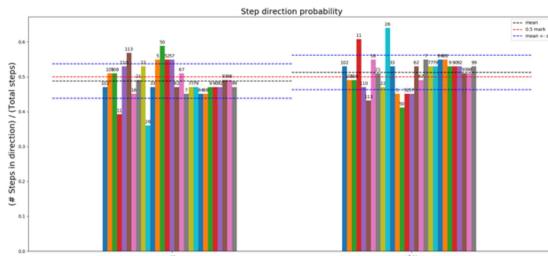
D is invariant for the MSD method approx $\tau < 8$

Time invariance is important because it signifies the motion as Brownian (random walk).

Our choice of $\tau = 5$ has an approximate D coeff. Of 10 for both of the methods

Probability of Step Direction

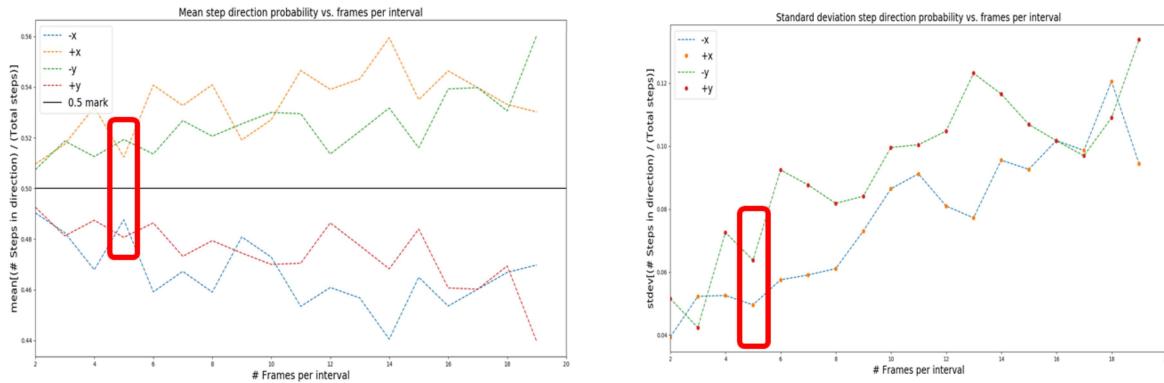
$$P(\text{direction}) = \frac{\# \text{ steps in (+ or -) direction}}{\text{Total steps of particle in } x \text{ or } y}$$



The probability of the step direction is defined as the number of steps in the positive or negative direction of x or y, divided by the total steps of the particle in x or y

- With the interval of 5 frames, we made bar plots of the step direction probability for each particle in the directions -x, +x, and -y, +y
- For a particle undergoing random motion, we would expect a particle's motion in the x-direction to be in the +x 50% of the time and in the -x 50% of the time
- We would expect a particle's motion in the y-direction to be in the +y 50% of the time and in the -y 50% of the time
- Most of these probabilities fall within one standard deviation of 50%. We do not expect these probabilities to be exactly 50% because we only have 50 data points and thus the deviations in the particle motion are more obvious. Also, the means of the step direction probabilities are within +/- 2% of the theoretical expectation of 50%

Probability of Step Direction: At different frame intervals

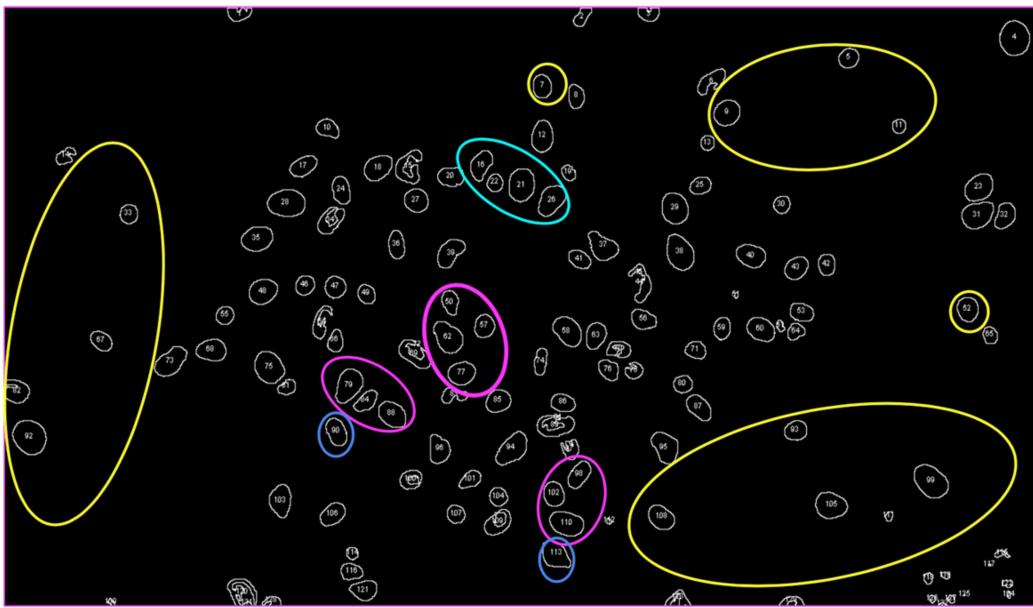


At 5 frames per interval the means of the step direction probabilities are within +/- 2% of the theoretical expectation of 50%. We see that as the number of frames per interval increases, this percent difference from the mean step direction probability slowly increases.

At 5 frames per interval the standard deviations of the step probabilities are approximately 0.065 for Y and 0.050 for X. We see that as the number of frames per interval increases, the spread of the step direction probabilities for the particles increases.

This justifies our choice of 5 frames per interval because the spread of the probabilities about their means is fairly low and the means themselves are close to the theoretical expectation of 50%. Along with this, 5 frames per interval allows the particle step lengths to vary from 0 to 6 pixels.

Data Collection in ImageJ



Yellow= alone

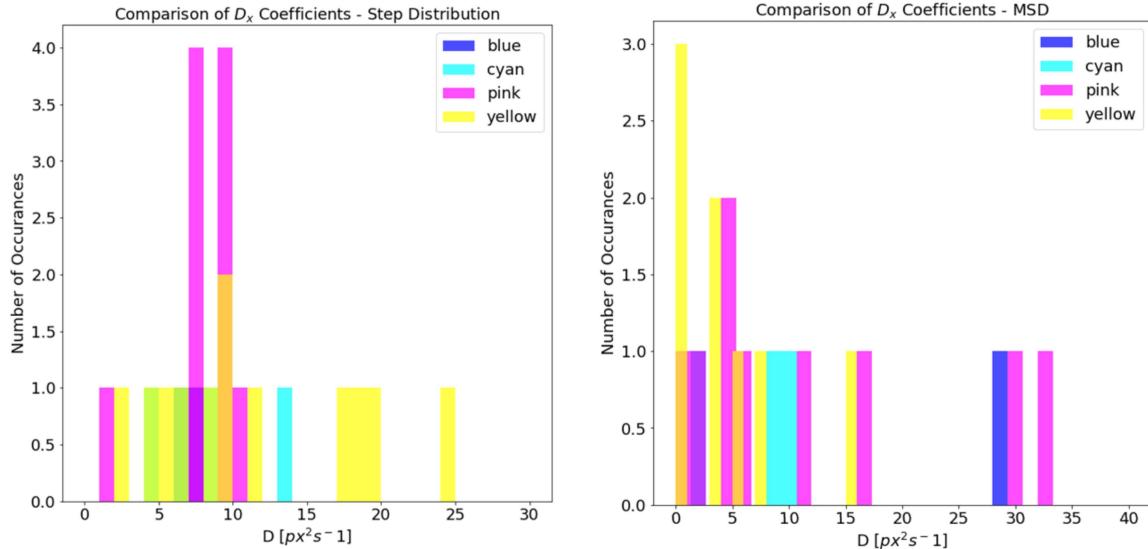
Pink: constrained on all sides

Light blue: Constrained on some sides

Dark blue: constrained on one side

ImageJ was used to number the particles in one frame, and collect their structural information (such as perimeters, and areas) → this was used to calculate the circularity. Areas → circle area → circle radius

Comparisons of Diffusion Constants Between Regions



Comparing the D values in the different colored regions (and also between the different methods - MSD, step distribution), we see the values of D_x and D_y are not consistent between the regions. This suggests a possibility that density does not have an affect on the diffusion. Since, if there was a relationship between density region and the D coefficient, we would expect the colored sections to be close to one another rather than spread out over a wide range of D values.

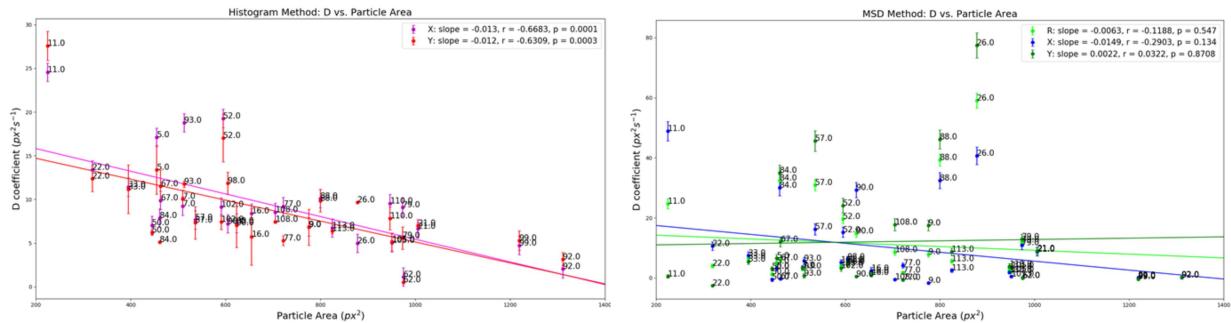
Yellow= alone

Pink: constrained on all sides

Light blue: Constrained on some sides

Dark blue: constrained on one side

D vs. Particle Area



$$\langle r^2 \rangle = 4D\tau = \frac{k_B T \tau}{3\pi r \eta}$$

We would expect that if particles are on the scale of a diameter less than 0.1 microns they exhibit brownian motion, and thus their area or size should not affect their diffusion. However, we do not know the actual size of the pollen grains in microns (only in pixels) because we do not know the camera's zoom on these particles.

In the histogram method the larger diameter/area signifies a larger drag force.
(Stoke's law)

In the MSD method Stoke's was used to derive this value of D which assumes that all particles are perfectly spherical, and thus this is why the plot has no relationship between particle area and D.

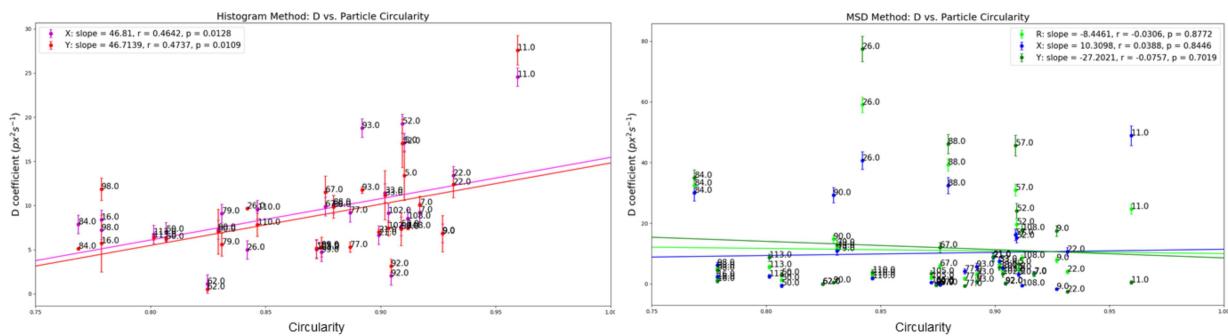
Histogram Method:

X: slope = -0.013, r = -0.6683, p = 0.0001 # moderate downhill
Y: slope = -0.012, r = -0.6309, p = 0.0003 # moderate downhill

MSD Method:

R: slope = -0.0062, r = -0.1188, p = 0.547 # no linear relationship
X: slope = -0.0149, r = -0.2903, p = 0.134
Y: slope = 0.0022, r = 0.0322, p = 0.8708

D vs. Particle Circularity



Circularity = $4 * \pi * \text{area} / \text{perimeter}^2$

Circularity = 1 is a perfect circle

Circularity = 0 is a line (no area)

These particles have a low high aspect ratio since the circularity $\sim (0.75, 1)$. We would expect brownian motion to be insensitive to shape under a low aspect ratio due to its derivation assuming spherical particles (sphere has circularity = 1).

In the MSD method Stoke's was used to derive this value of D which assumes that all particles are perfectly spherical, and thus this is why the plot has no relationship between particle circularity and D.

Histogram Method:

X: slope = 46.81, r = 0.4642, p = 0.0128 # moderate uphill
 Y: slope = 46.7139, r = 0.4737, p = 0.0109 # moderate uphill

- Intercepts
 - X: -31.4
 - Y: -31.9

MSD Method:

R: slope = -8.4461, r = -0.0306, p = 0.8772 # no linear relationship

X: slope = 10.3098, r = 0.0388, p = 0.8446

Y: slope = -27.2021, r = -0.0757, p = 0.7019

- This result appears to make more sense because for particles that are close to circles they should have D independent of circularity

Uncertainties

Tracked particles multiple times, calculated difference in measured x and y positions for the same particle

→ on average, Tracker measured x and y positions with an uncertainty of ± 0.65

→ Uncertainty in Histogram Method: $D \pm 1$

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$
$$D = \sigma^2 / 2\tau$$

→ Uncertainty in MSD Method:

- $(\pm 0.1, \pm 20)$ for MSD of x, y, r
- $\pm 0.6, \pm 0.4, \pm 0.4$ for D of x, y, r

We tracked particles multiple times and calculated the difference in measured x and y positions for the same particle.

- On average, Tracker measured x and y positions with an uncertainty of ± 0.65 px
- This uncertainty was used and propagated through our MSD calculations

Histogram Method: When propagating the x, y position error through the standard deviation with single variable error propagation, and then through the calculation of D coeff., we get approximately ± 1 as our uncertainty on D for this method

MSD Method: When propagating the x, y position error through the x difference and y difference between frames calculation we use two variable error propagation. Once we have these uncertainties we use single variable error propagation to compute the uncertainty on the MSD values.

Large Error Bars with Increasing Time: Since our MSD values are calculated by subtracting the initial position from its current position and squaring that difference, the uncertainty would scale in the same way. Since these MSD values get larger with time, so will their uncertainties.

The uncertainty on the D reported for the MSD method is from the standard error on the linear fit.

Conclusions

Research Question: Can we confirm that pollen grains in water exhibit brownian motion?

We expect:

- The step lengths are randomly and normally distributed
- The probability of step direction is equal for all directions
- Random walk?
 - No spatial correlation: density regions do not affect D coeff.
 - Possible temporal correlation for MSD method on D coeff.
 - This temporal correlation is different for the 2 methods

Which Method?

- The histogram method appears more realistic in describing the system

More realistic in the sense that it acknowledges these pollen grains are not all perfectly circular

Next Steps

- Compute a theoretical value of D to compare to our results
- Find the difference in the D values between the two methods for the individual particles
- Compare these D coeffs. From other studies
- Find other methods to determine whether there is temporal correlation
- Use a larger data set

Thank you!

Questions?

Characterization of Pollen Grain Motion in Water

¹Faizah Siddique, ¹Sasha Bakker, and ¹Linda Oster
¹*Department of Physics, University of Massachusetts Amherst*

ABSTRACT

A study is performed to confirm whether pollen grains exhibit Brownian motion. The analysis is executed on a video of aspherical shaped pollen grains suspended in a water medium. Positional measurements were collected using Tracker 5.1.5 and analyzed for grains in spatially different density regions of the video. The step lengths in both the X and Y coordinates was shown to be randomly and normally distributed, suggesting the particles exhibited Brownian motion. The mean square displacement (MSD) and histogram probabilities of step length in both spatial coordinates were used to compute diffusion constants. It is shown that there exists no spatial correlation in computed diffusion constant values for different density regions in either method, and that a temporal correlation exists for the MSD method. By studying particle properties, such as circularity and area, in relation to the diffusion constants in different density regions, it is shown that the histogram probability distribution method is the more reliable method when studying Brownian motion.

I. INTRODUCTION

The study of the random motion of small particles in a medium has been investigated across many years by multiple scientists. In 1828, botanist Robert Brown was the first to observe that pollen grains moved randomly in water, and since then, Brownian motion became the nomenclature to describe the random behavior of small particles suspended in a medium. Today, the behavior of Brownian motion can be observed in biological systems (diffusion of chemicals within the body), as well as in weather systems, and even the stock market [1].

Albert Einstein's discoveries on relativity and light quanta has overshadowed his PhD dissertation from 1905 on the treatment of fluctuations in motion of suspended particles. In his study, Einstein argued that the fluctuations in movement were governed by the atoms and molecules themselves. Specifically, Einstein performed a statistical analysis on the motion of the particles in which he calculated the MSD (Mean Square Displacement) of these particles. He found that the MSD is related to Avogadro's number and also devised a physical explanation for Brownian motion, which led to the acceptance of the kinetic theory of matter [2].

Following Einstein's discoveries, Jean Perrin performed a series of experiments during the twentieth century, similar to that of Einstein's, in which he calculated the MSD of suspended particles. Perrin confirmed Einstein's observation and the molecular-kinetic theory of matter [2].

A. Brownian Motion and Kinetic Theory of Matter

The ideal gas law, $pV = \mu RT$ (where pV is the product of the pressure and volume of a gas, μ is the moles of a dilute gas at temperature T , and R is the Ideal Gas constant) can be used to determine the value of Avogadro's number, N_A . The kinetic theory of gases suggests the value of R can be written in terms of Boltzmann's con-

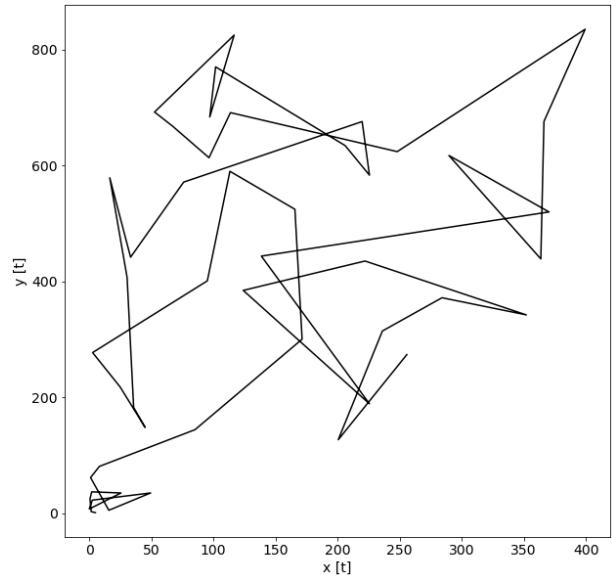


FIG. 1. The random walk of pollen grain particle 50 from this analysis (procedures discussed at depth below)

stant, k_B , and N_A as $R = k_B N_A$ [2].

The kinetic molecular theory also states that temperature is a measure of the mean molecular kinetic energy $K = \frac{3}{2}k_B T$. However, the only value uniquely defined in the temperature scale is absolute zero; the temperature scale itself is arbitrary. Measuring the value of k_B independently allows for a better method of calculating the value of N_A [2].

Einstein was able to show that a particle exhibiting Brownian motion in a fluid of viscosity η , temperature T , in time t would travel a distance of $R(t) = \sqrt{Dt}$ where D is the diffusion constant. Hence, an observation of a particle's trajectory under a microscope would allow for the measurement of $R(t)$, which, in turn, would allow for a calculation of k_B , thus eventually leading to a value for

N_A . Einstein was able to determine the quantity N_A by observing sugar molecules under a microscope [2].

B. Stoke's Law

In 1851, G. G. Stokes showed that the frictional coefficient, Γ , opposing the motion of particles through a fluid with viscosity η , can be written in terms of the particle radius a as $\Gamma = 6\pi\eta a$. Using this value for Γ , and writing an equation balancing the gravitational force with the force exerted on a particle from a displaced fluid, the following relationship can be written [2],

$$\frac{nF}{\Gamma} = -D \frac{dn}{dz} \quad (1)$$

where $\frac{dn}{dz}$ is a z-dependent profile for the concentration of particles. The osmotic pressure, which is the partial pressure of Brownian particles, can be written simply as a function of depth $p_{osm}(z) = n(z)k_B T$. The osmotic force per unit volume is then resolved to $f_{osm} = -k_B T \frac{dn}{dz}$. Balancing this with the force of gravity per unit volume, the following equation is devised [2],

$$n(z)F = n(z)(Mg - W) = k_B T \frac{dn}{dz} \quad (2)$$

Comparing (1) and (2) leads to Einstein's famous result for Brownian motion in 1-dimension,

$$D = \frac{k_B T}{\Gamma} = \frac{k_B T}{6\pi\eta a} \quad (3)$$

This equation allows for the extraction of k_B from measurements of diffusion. Details of the exact derivation of this expression can be found in [2] and [1].

II. PURPOSE

The purpose of this analysis was to determine whether the motion of pollen grains suspended in a medium can be characterized by Brownian motion, and whether it can be described by the interactions detailed in the sections above. The goal was to also make quantitative measurements of the value of the diffusion constant using different methods, and to determine whether D depends on density or other particle properties such circularity or surface area. A replication of Brownian motion, using the methods highlighted above, will further confirm that diffusion can be used to determine important constants in physics such as k_B .

III. METHODS AND PROCEDURES

In this analysis, 28 aspherically shaped pollen grains, likely observed under a microscope, suspended in a water medium were tracked as a function of time in at a

frame rate of 25 fps (frames per second). This video was found on YouTube and does not provide any information about the length scale of the video or what plant these pollen grains originate from)[3]. Thus, the analysis uses lengths in units of pixels to represent the data. Using Tracker 5.1.5, the individual $x(t)$, $y(t)$ positions were collected for each of the 28 particles in 5 frame intervals. The choice of picking 5-frame intervals for the position is discussed at length in Sec. IV.A and IV.F. The units of time in fps were converted to seconds, and the positional data for the particles was used to calculate the diffusion constant in two different ways. These values were and compared by plotting the probability density of step lengths taken by the particle in either X or Y and retrieving the diffusion constant from the standard deviation of the normally distributed probability density curve, as well as by employing Einstein's relationship of MSD as a function of time. If the probability distribution plots of step lengths in both directions are normally and randomly distributed, there would be reason to suggest the motion of the pollen grains is governed by Brownian interactions. The values of D individually calculated for each particle were compared for different particle areas, and circularity to determine which method of computing the value D was more reliable. The values of D computed by both methods were qualitatively compared for different density regions of the video to determine whether a relationship exists between particle density and D .

A. Tracking Particle Positions with Tracker 5.1.5

Tracker 5.1.5, an Open Source Physics (OSP) video analysis and modeling tool that allows for manual and automatic tracking of objects, was used in collecting positional data for this analysis [4]. The auto-tracking feature on Tracker 5.1.5 allows for the tracking of the X and Y in pixel positions of each particle at 25 fps for the duration of the video. Tracker's auto-tracking feature implements a 'best match score' process to track particles in subsequent frames by creating a template image and searching each frame for the best 'match score'. The best match score is defined to be the number that is inversely proportional to the sum of the squares of the RGB differences between the template and the match pixels. The template is compared with nearby scores to determine an interpolated best pixel match position [4]. A target is placed on the particle prior to tracking. Its cross-hairs are centered at matches relative to the template. For this analysis, the targets were placed at the centers of each particle. The positional X and Y data (measured with respect to the center of the image on Tracker 5.1.5) were exported as text files to be used for the rest of the analysis.

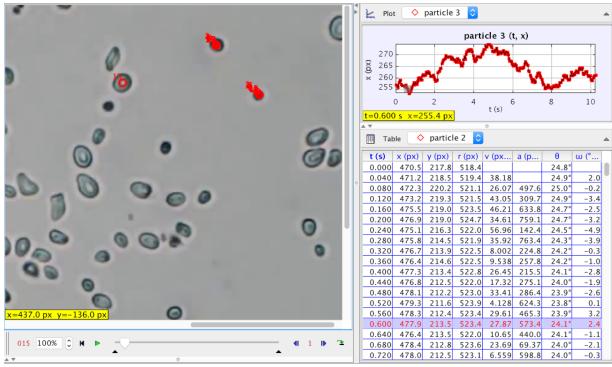


FIG. 2. Tracker auto-tracking feature dumping particle X, Y for every time step.

B. Data Collection with ImageJ

Image processing program ImageJ, an Open Source image processing program designed for scientific multi-dimensional image analysis [5], was used to number the particles of the system by first creating a binarized image of the first frame of the video. ImageJ was also used to collect other particle structural information such as particle perimeter and area that were used in the Sec. IV.D of the analysis to compute circularity.

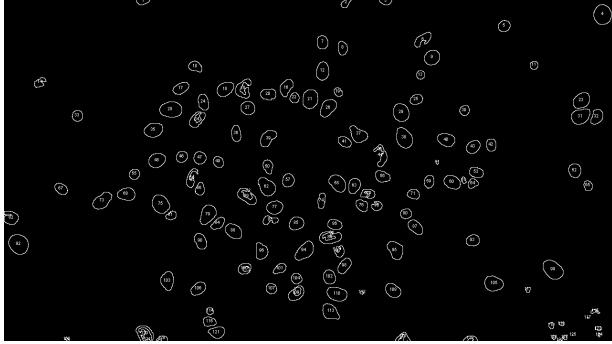


FIG. 3. ImageJ binary image of first time step with particles numbered.

C. Probability Density of 1-Dimensional Displacement

A method for determining the diffusion constant is by studying the step lengths for many particles in the X and Y positions [6]. The probability distribution density for finding 1-D displacement Δ is,

$$P(\Delta) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{\Delta^2}{2\sigma^2}} \quad (4)$$

where the standard deviation of this Gaussian distribution function is $\sigma = \sqrt{2Dt}$, and t is the time of each

measurement. The diffusion constant D can be determined by fitting Eq. (4) to a normalized histogram of the steps lengths of the particles in the X and Y positions [6].

D. Mean Square Displacement (MSD)

Mean square displacement values were computed separately in X and Y pixels for the group of 28 particles and were plotted as a function of time in seconds. Working backward from Einstein's Eq. (3), the diffusion constant can be rewritten in terms of MSD in X and Y as,

$$\langle x_j^2 \rangle = 2Dt = \frac{k_B T t}{3\pi a\eta} \quad (5)$$

where x_j is either X or Y defined for each time interval t . Here, the left side of Eq. (4) gives the computed MSD, for the specified direction and time interval. MSD is represented by the mean of the squared differences in position between time interval t and $t = 0$ for $N = 28$ particles,

$$\langle x_{jt}^2 \rangle = \frac{\sum_{p=1}^{28} (x_{jtp} - x_{j0p})^2}{N} \quad (6)$$

The diffusion constant was also computed for the total displacement MSD $\langle r^2 \rangle$ versus t where,

$$\langle r_t^2 \rangle = 4Dt = \frac{\sum_{p=1}^{28} (x_{tp} - x_{0p})^2 - (y_{tp} - y_{0p})^2}{N} \quad (7)$$

Similar to $\langle x_{jt}^2 \rangle$, $\langle r_t^2 \rangle$ represents the total MSD of the group of 28 particles for every 5-frame time step interval t . The slope of the $\langle r_t^2 \rangle$ versus t plot allows for the calculation of the total MSD diffusion constant.

E. Probability of step direction for X-Y coordinates

A metric to determine whether the pollen grain motion is reminiscent of a random walk is to verify whether the step in the positive or negative direction for X-Y coordinates is random. For a particle undergoing random motion in 1-D, the particle is expected to travel in the positive direction half of the time, and in the negative direction half of the time. Meaning, the probability for a 1-D positional step in the positive or negative direction is predicted to be 50% at each time step. Thus, the probability of stepping in the positive or negative direction is computed as the number of steps in the positive or negative direction divided by the total steps in the 1D coordinate. At the time interval of five frames, the mean of the step direction probability for the particles should be approximately 50%, for the motion to be considered

random. This mean is not expected to be exact because there are only fifty data points, and thus the deviations in the particle motion are more obvious with this small sample size per particle.

IV. RESULTS

A. Probability of Step direction Results

The 1-D step direction probabilities for X-Y coordinates were plotted as separate bar charts in Fig. 14 and Fig. 15, with every individual bar signifying a particle. The mean probabilities of the bars for both the X and Y coordinates were within 3% and 2% of the theoretical probability of 50%. In the X coordinate, 2 of the 28 particles fall outside two standard deviations from the theoretical probability, and 4 of the 28 particles fall outside this condition in the Y coordinate.

The result of increasing the number of frames per interval is that the mean step probability for each direction +X, -X, +Y, -Y deviates further from the theoretical expectation (Fig. 16). For example, the mean probability in the -Y direction is approximately 51% at 2 frames per interval, then increases to 56% at 19 frames per interval. Thus, as the number of frames per interval increases, the probability of stepping in the positive or negative direction deviates from theoretical, and the particle will appear to have a biased movement in either the positive or negative direction. Thus the choice of 5 frames per interval is justified, since this selection is close to the theoretical expectation.

B. Diffusion Constant Computed with 1-Dimensional Displacement Probability Distribution

The steps lengths for X and Y , computed for each particle, by taking the difference in particle positions for every 5 frames. The step lengths for all 28 particles were plotted separately as histograms by their 1D direction in Fig. 4 and Fig. 5. These distributions were shown to be randomly and normally distributed. Due to this, the histogram data were fitted to functions of the form Eq. (4), and the diffusion constant was extracted from the standard deviation of the fits.

The standard deviation for the histogram of X -step lengths and corresponding diffusion constant were found to be $\sigma=1.975$, $D=9.8\pm1$ [px^2s^{-1}]. The standard deviation of the Y -step lengths and corresponding diffusion constant were found to be $\sigma=1.888$, $D=8.9\pm1$ [px^2s^{-1}]. The uncertainty calculations on the measurements of D with this method are discussed in Sec.V.

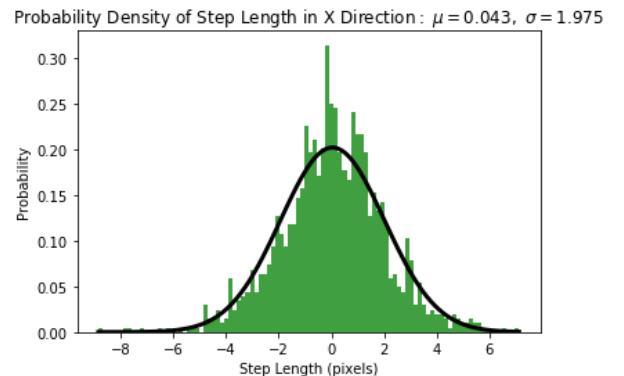


FIG. 4. Probability density of step length in X-direction.

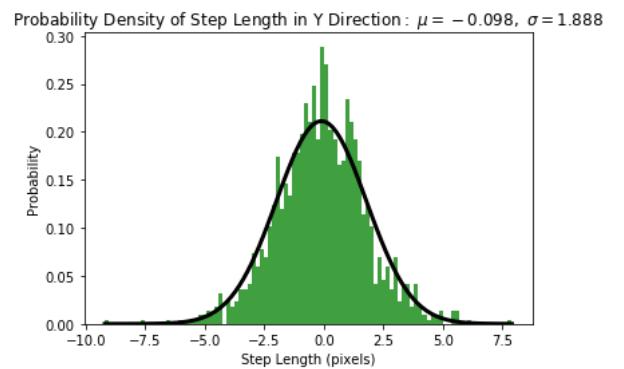


FIG. 5. Probability density of step length in Y-direction.

C. Diffusion Constant Computed with MSD

The diffusion constant can be extracted from the slopes of the MSD versus time plots. The individual MSD plots and linear fits are shown on Fig. 6 and Fig. 7, along with their linear fitting parameters and uncertainties. Here, the MSD values are plotted for 5-frame intervals.

The diffusion constant from the linear fit for $\langle x^2 \rangle$ versus t is found to be $D = 10.1 \pm 0.6$ [px^2s^{-1}], while the diffusion constant from the $\langle y^2 \rangle$ versus t fit is found to be $D = 12.1 \pm 0.4$ [px^2s^{-1}]. The chi-square value and reduced chi-square values for both the $\langle x^2 \rangle$ and $\langle y^2 \rangle$ fits are found to be, $\chi^2 = 219.4$, $\tilde{\chi}^2 = 4.6$, and $\chi^2 = 101.8$, $\tilde{\chi}^2 = 2.1$. These values suggest the linear fits are good for most of the data points in Fig. 6 and Fig. 7. The p-values of approximately zero for both fits, suggesting the null hypothesis can be rejected because it is less than the significance level of $\alpha = 0.05$. The null hypothesis rejection means it is statistically significant that the slope is nonzero. The r-values of 0.95 and 0.97 similarly suggests an almost perfect positive linear relationship between $\langle x^2 \rangle$ and t , and $\langle y^2 \rangle$ and t .

The comparatively large errorbars on the data during the later times on both Fig. 6 and Fig. 7 are due to the wider distribution of the positions of the 28 particles in total over time. This effect on the increase in error of

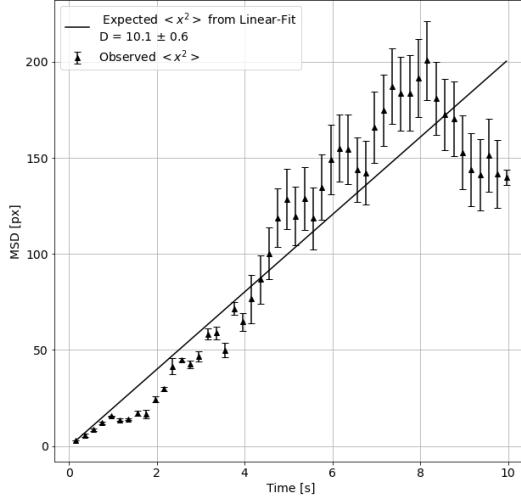


FIG. 6. $\text{MSD} < x^2 >$ versus Time Linear-Fit. The following parameters were computed for the fit: $\chi^2=219.4$, $\tilde{\chi}^2=4.6$, p-value=0.0, r-value=0.93

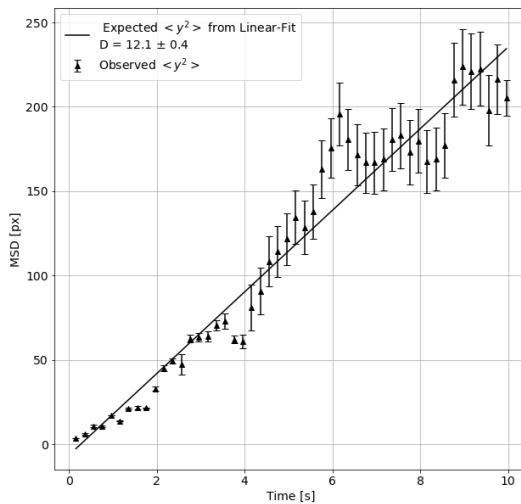


FIG. 7. $\text{MSD} < y^2 >$ versus Time Linear-Fit. The following parameters were computed for the fit: $\chi^2=101.8$, $\tilde{\chi}^2=2.1$, p-value=0.0, r-value=0.97

MSD calculations over time is not evident in the method outlined in Sec. III.C and is one reason to suggest that the method outlined in Sec. III.C may pose as a better alternative to calculating values of the diffusion constant for the system. Another reason to suggest the histogram method as the better alternative is the over weighting affect that the initial data points have on the MSD calculations for the particles at later times. Since, the values

for the MSD are calculated by first subtracting the initial x_0 and y_0 positions for each particle from their positions at time t , such that the earlier points of the data are more heavily weighted.

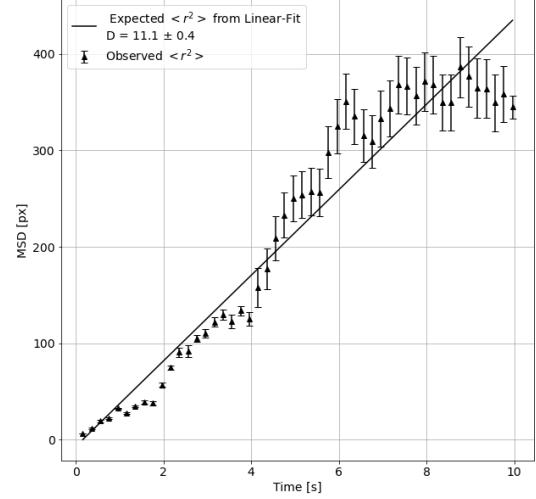


FIG. 8. $\text{MSD} < r^2 >$ versus Time Linear-Fit. The following parameters were computed for the fit: $\chi^2=528.9$, $\tilde{\chi}^2=11.0$, p-value=0.0, r-value=0.97

The diffusion constant from the linear fit for $< r^2 >$ versus t is found to be $D = 11.1 \pm 0.4 \text{ [px}^2\text{s}^{-1}\text{]}$. The chi-square and reduced chi-square values are $\chi^2=528.9$, $\tilde{\chi}^2=11.0$ and signifies that the linear fit does not fully represent the data. However, the p-value of approximately 0.0 and r-value of 0.97 suggests that the null-hypothesis can be rejected and that there is an almost perfect positive linear relationship between $< r^2 >$ and t . There are once again larger error bars on the data in Fig. 8 due to the wider distribution of the total 28 particles over time.

An unusual observation to note is the larger $\tilde{\chi}^2$ value for the $< r^2 >$ fit in Fig. 8 compared to that of the $< x^2 >$ and $< y^2 >$ fits in Fig. 6 and Fig. 7. Reinspecting the data from Fig. 6 and Fig. 7 shows that there is a substantial deviation that appears around $t=1.89$ in both plots. This is unusual given that calculations of $< x^2 >$ and $< y^2 >$ are expected to be uncorrelated. This suggests a possibility of an external force acting on the system at around $t=1.89$.

D. Diffusion Constant versus Particle Area and Circularity for Both Methods

When plotting the diffusion constant versus particle area or circularity, it is expected that larger particles will have a larger drag force according to Stoke's Law, such that the diffusion relies on the size and shape of the

molecule. First, the relationship between D and particle area is considered. The histogram method Fig. 19 produced a moderate negative linear relationship for both X and Y with a slope of -0.013 s^{-1} and -0.012 s^{-1} , respectively. The p-values of both linear models are smaller than the 0.05 threshold, signifying the hypothesis of a nonzero slope is supported. Thus, this relationship is statistically significant. On the other hand, the MSD method Fig. 17 produced fits with p-values larger than the threshold, signifying that the hypothesis of zero slope cannot be rejected. Now, the relationship between D and particle circularity is considered. These particles have a low aspect ratio with circularity within $(0.75, 1)$. In Eq. 3, the diffusion is inversely proportional to the frictional coefficient for the spherical geometry. When comparing this coefficient between a sphere and an ellipsoid, the ellipsoidal geometry should produce a larger coefficient, and, in turn, produce a smaller diffusion coefficient. The histogram method Fig. 20 produced a moderate positive linear relationship for both X and Y with a slope of $46.81 \text{ px}^2\text{s}^{-1}$ and $46.71 \text{ px}^2\text{s}^{-1}$, respectively. The p-values of both linear models are smaller than the 0.05 threshold, signifying the hypothesis of a nonzero slope is supported, meaning, the relationship is statistically significant. In contrast, the MSD method Fig. 18 produced fits with p-values larger than the threshold, such that the hypothesis of zero slope cannot be rejected.

E. Comparison of Individual Diffusion Constants for Different Regions and Different Methods

The individual D values computed for each particle from both methods highlighted in Sec. III.C and III.D were compared qualitatively at different density regions of the video. The segments of the video were divided as follows Fig. 9:

- Brown - Particle not surrounded on all sides by other particles.
- Grey - Particle constrained on all sides by other particles.
- White - Particle constrained on some sides by other particles.
- Blue - Particle constrained on only one side by another particle.

Based on these classifications, the values of the individual diffusion constants were plotted as histograms for the different methods in Fig. 10 and Fig. 11.

It can be inferred, by comparing the individual diffusion constants plotted in the histograms for both methods, that there does not exist a relationship between the values of D and the particle region density. The values of D fall in such a wide range of values (i.e. the grey region for both methods shows the D value range between approximately 2 to $33 \text{ px}^2\text{s}^{-1}$) such that it is difficult to

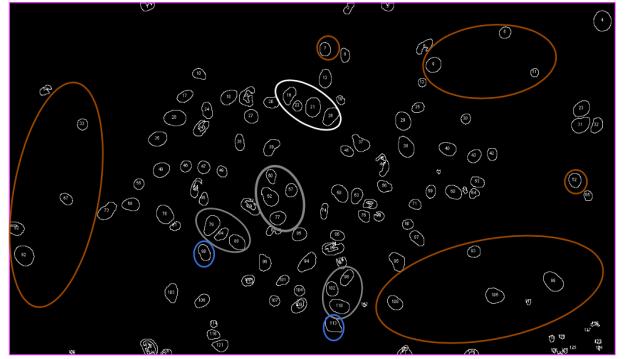


FIG. 9. Different density regions distinguished by colors.

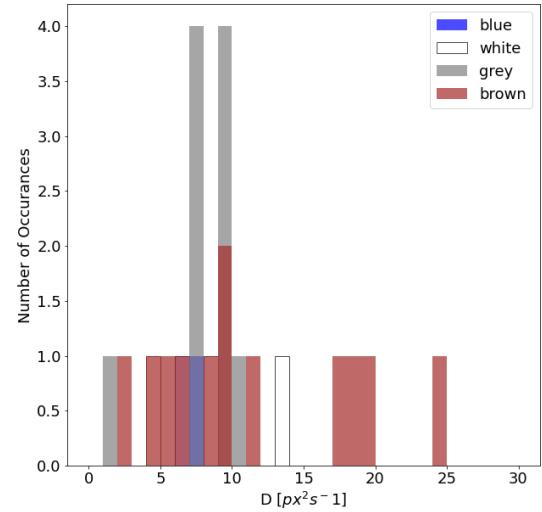


FIG. 10. Comparison of D constants using the X-direction step distribution.

suggest density has an effect on diffusion. The same colored peaks on the histograms would have been closer to one another in range, had the density affected the value of D measured from both methods.

F. Diffusion Constants Versus Time

The diffusion constant was computed for all 28 particles in the system at different time steps, signifying the choice of the number of frames per interval. This was done separately for both the Probability Density of 1-Dimensional displacement method and the MSD method. These values of D were then plotted against the number of time steps.

Fig. 12 reveals that the values of D from the probability density method are smallest for the lowest time interval and increase until approximately the time step of 12 frames, before leveling off to approximately a sin-

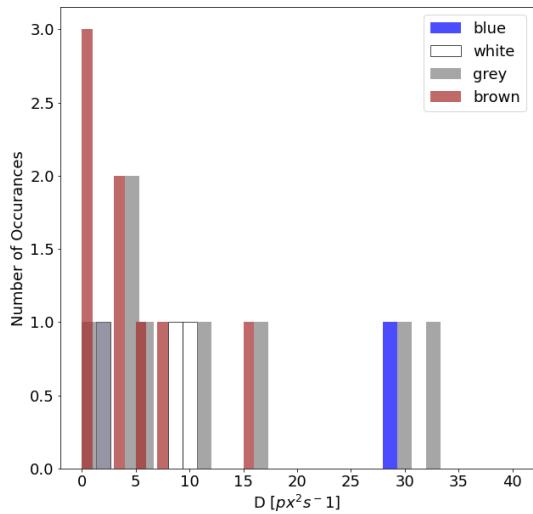


FIG. 11. Comparison of D constants using the X-direction MSD fits.

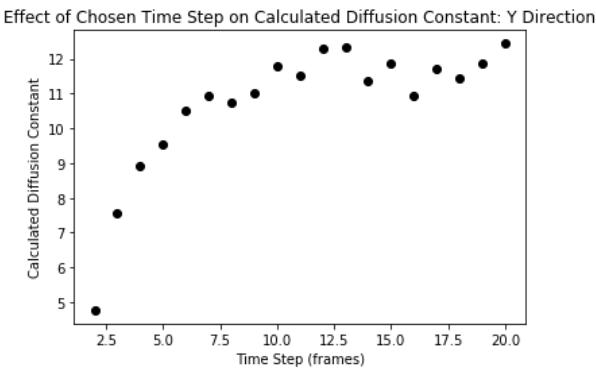
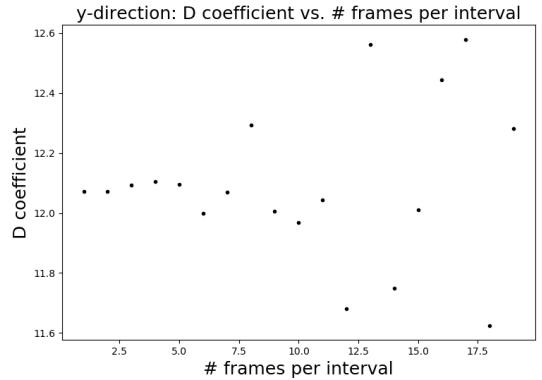


FIG. 12. One-dimensional diffusion constants in the Y coordinate computed from the probability density method, with use of all 28 particles, plotted against the number of frames per interval.

gular value. On the other hand, Fig. 13 reveals that the values of D begin at a stable value of approximately $10 \text{ px}^2 \text{s}^{-1}$. This value of D begins to randomly deviate at about 8 frames per second, with the magnitude of deviations increasing with the number of time steps.

The choice of five frames per interval is justified based on the results of these plots for two reasons. First, both of the relationships produce a diffusion constant of approximately 10 at 5 frames per interval which makes the constants comparable in magnitude between the two methods. The second justification is that there is minimal temporal correlation with D at the smaller number of frames per interval with the MSD method.



results in an unavoidable bias for the particle to move in a certain direction for this method of computing D .

VI. CONCLUSIONS

The step length distribution for both X-Y directions in Fig. 4 and Fig. 5 are shown to be both randomly and normally distributed, as expected for pollen grain particles exhibiting Brownian motion. For a 5-frame time interval, the probability of stepping in either $\pm x$ or $\pm y$ is approximately equal for all directions and all particles. From the comparison of D values in the different particle density regions of the video, there appears to be a zero spatial dependence relation (as evident in the histograms in Fig. 10 and Fig. 11), signifying particle density does not influence the effect of diffusion. However, a possible temporal correlation may exist for the MSD method for computing D since there exists an over-weighting effect on the initial data points for MSD calculations for the particles. If the initial positions of the particles change, so will the MSD calculations for those particles. Similarly, as the length of the stepping time intervals change, the step length histograms will deviate from Gaussian due to the decrease in the total number of data points analyzed. In other words, the normal fit of the probability distribution function for step-lengths will be less accurate. Similarly, Stoke's theorem assumes the particles to be spherical.

Due to this fact, it is expected that a relationship exists between the computed D values and individual particle properties. From Sec. IV.D, it was shown that there exists no relationship between the MSD method D constants and particle areas or circularities. In contrast, for the step-length histogram method, there appears to be a relationship between D constants and particle area or circularity. Based on these observations, it can be concluded that the step-length distribution method is a more reliable model in studying diffusion of pollen grains in a fluid.

For future work, it would be beneficial to compute a theoretical value of the diffusion constant for this particular system in order to make comparisons to the experimental data. It would also be useful to also make comparisons with the experimental results from other similar studies as well as to further study the effect of temporal correlation. This would be helpful for understanding the differences in D values computed for both MSD and step-length distribution methods.

ACKNOWLEDGMENTS

We would like to express our sincere acknowledgement to our instructor, Dr. Chen Wang, for providing us the guidance and resources we needed to be successful in this project.

-
- [1] P.-J. . R. W. Newburgh, R., American Journal of Physics **74**, 478 (2006).
 - [2] S. Ramaswamy, Resonance **5**, 16 (2000).
 - [3] koshu endo,.
 - [4] Physlets, (n/a).
 - [5] A.-C. I. . F. E. Schindelin, J., Fiji: an open-source platform for biological-image analysis, Nature methods **9**(7), 676 (2012).
 - [6] A.-M. L. J. . S. C. Nakroshis, P., Measuring boltzmann's constant using video microscopy of brownian motion, American Journal Of Physics **71**, 568 (2003).

VII. ADDITIONAL FIGURES

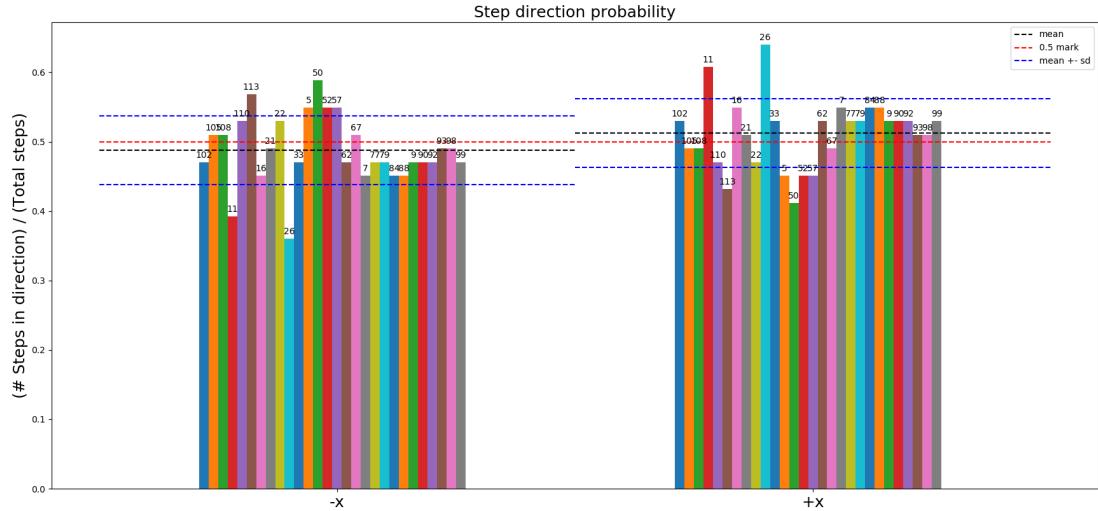


FIG. 14. X-Direction Step=5 Probabilities for all 28 Particles.

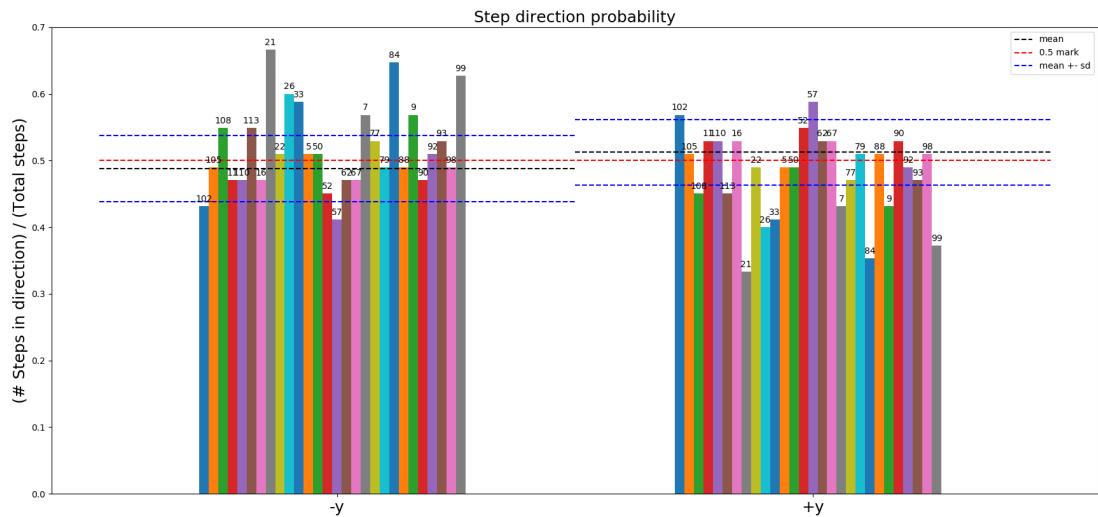


FIG. 15. Y-Direction Step=5 Probabilities for all 28 Particles.

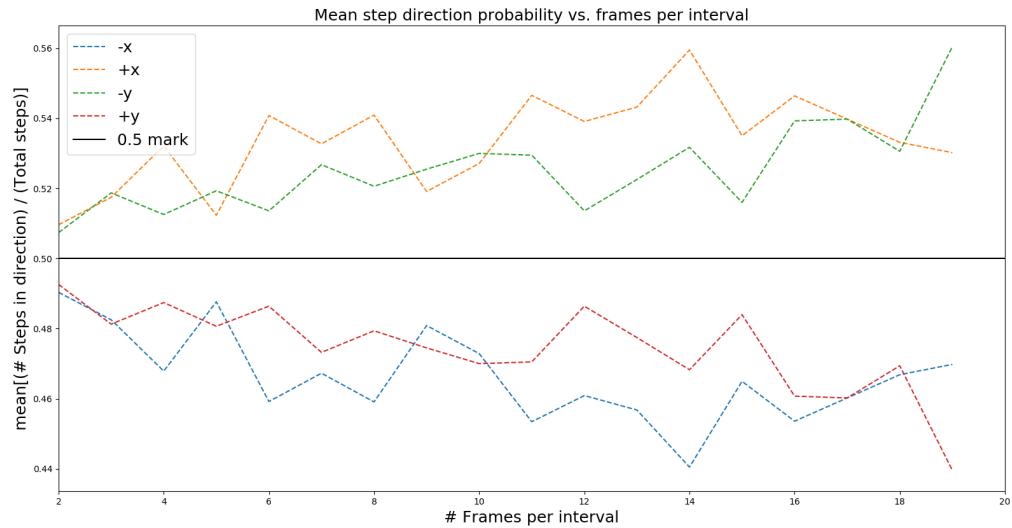


FIG. 16. Mean Step Displacement Probability versus Frames per Interval.

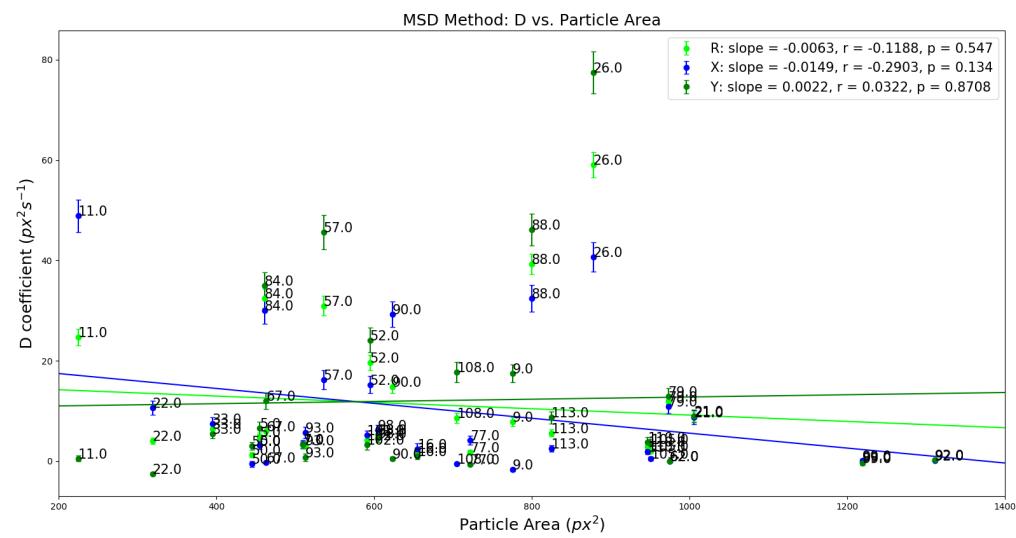


FIG. 17. MSD Method; Diffusion Constant versus Particle Area.

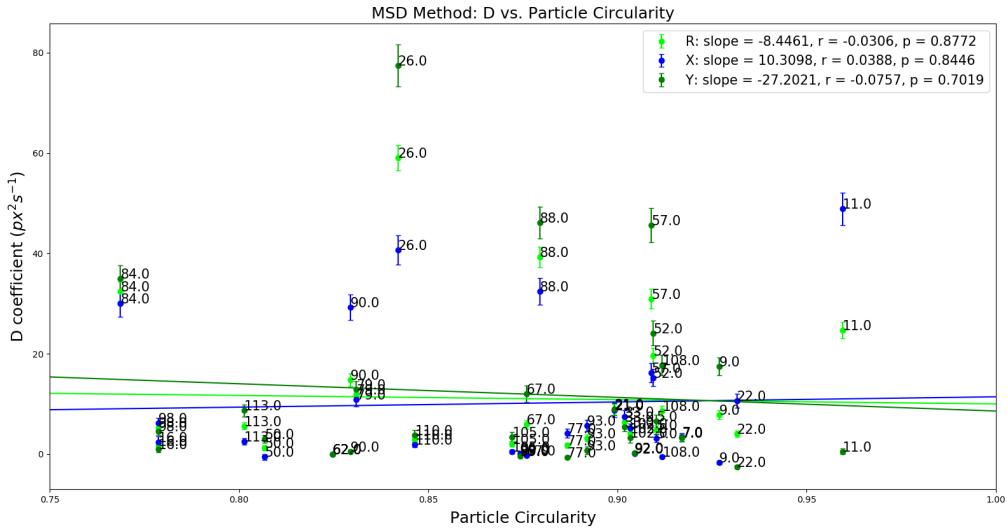


FIG. 18. MSD Method; Diffusion Constant versus Particle Circularuity.

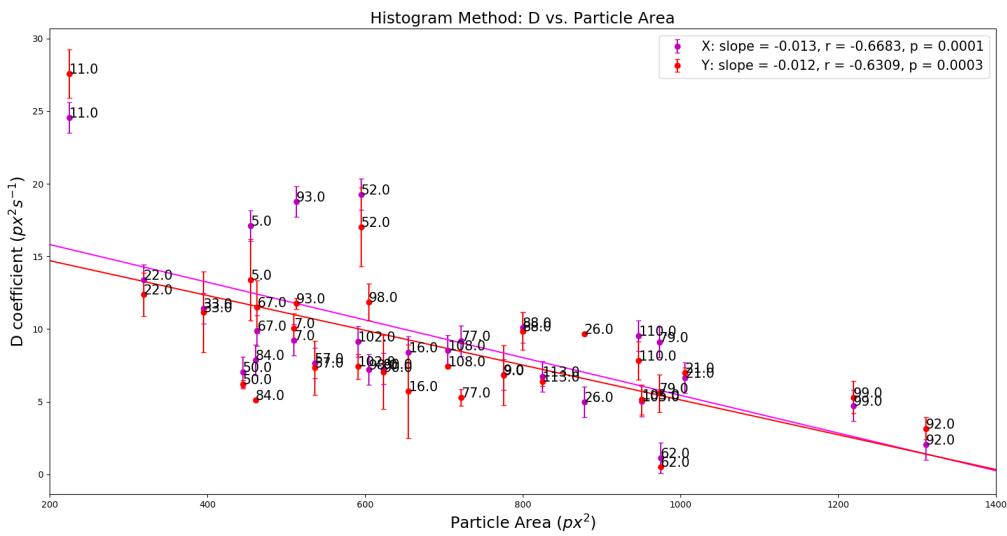


FIG. 19. Probability Step Distribution Histogram Method; Diffusion Constant versus Particle Area.

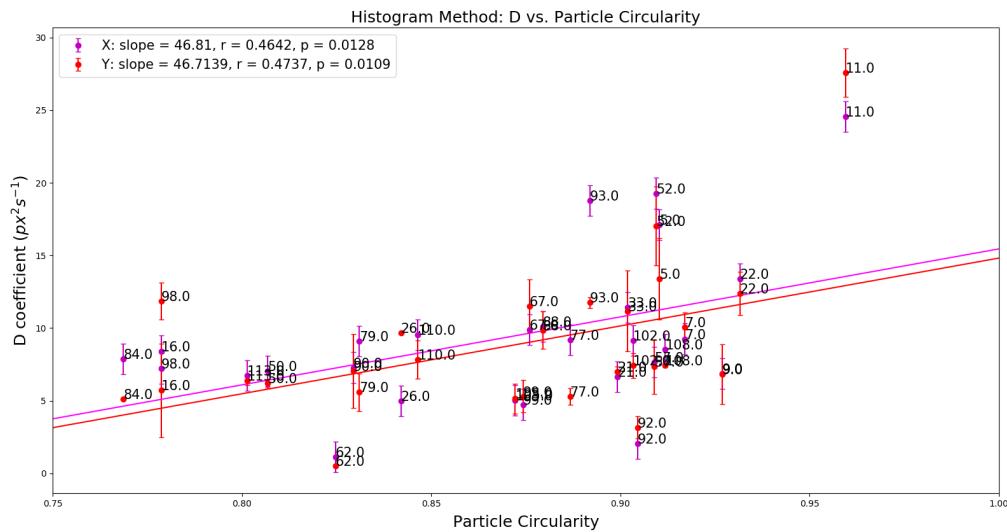


FIG. 20. Probability Step Distribution Histogram Method; Diffusion Constant versus Particle CircularitY.