Mathematical modelling of random clustering particles biology essay

Science, Biology

Submitted in partial fulfilment of the regulations for the Award of the Degree of Bachelor of Science with Honours in Biomedical Science DPPProject supervisor: Dr. W. P HaganChristopher VernonSchool of Biomedical SciencesFaculty of ScienceUniversity of UlsterColeraineNorthern IrelandThis project report has been prepared in accordance to the instructions to authors for the British Medical Journal.

ABSTRACT

The clumping process and growth of clusters is a statistically random process. By using automatic imaging techniques, we compare Roach's equations with simulated data, to produce an alternative equation, Y = A*X*EXP(-B*X). This was achieved using randomly generated particles and automated software which was able to identify particles and vary their radius to produce multiple sets of data. From this data a relationship between Roach's equation and randomly generated data could be determined. Results showed that although Roach's equation fitted well for single clumps and the total number of clumps, an improvement is necessary for clump sizes greater than one. The conclusions show that simulation using random plots is a valid clump generator and automated software was very efficient at carrying out the counting of clumps. The Roach equation was excellent for predicting all clumps and single clumps and poor at predicting clumps greater than one.

KEYWORDS

Clumping- a group of aggregated particles. Random- points made without a definite pattern. Lamina- a thin round disc. Particle- a discrete point. Two Dimensional- having dimensions of height and width only.

INTRODUCTION

Soap bubbles on a water surface, nano-particles, dust particles or red blood cells are conventionally thought of as separate discrete entities, but in reality they do not always exist as single units, they are often found as clusters or clumps. The forces of attraction that form these clusters include inter-atomic, van der Walls, electrostatic, magnetic and gravitational forces. The reasons for this depend on the nature of the particles and their surroundings. Additionally, these clumps or clusters may be either desirable, or alternatively confer on them properties which are markedly different from those of the single entity. It is important to know when particles may cluster as it can be an indication of concentration, not knowing enough about clustering leads to possible large underestimates in counting with consequential inaccuracies on results. These clusters are easily visible in the form of soap bubbles on a water surface, which are often found in clusters of variable size. One example of variable sized clusters is ice-cream, where the air is whipped into it, giving it a creamy texture and simultaneously prevents the growth of large ice crystals in storage. When ice-cream is taken out of the freezer it begins to thaw and the air bubbles decrease, so when it is re-frozen larger ice crystals form, this gives the ice cream a much chewier, less desirable, texture. Nano-particles often aggregate to form clumps or clusters

by Van Der Waals interactions between particles. The National Institute of Standards and Technology (NIST) are currently investigating the likelihood that nanoparticles will form clumps. The anticipated outcome is to create nano products that are nontoxic to human cells.[1] In cosmetics and other make-up these nanoparticles can clump and form larger particles when applied.[2] These bigger particles are more likely to attach to the mucus in the upper respiratory tract rather than travel into the lungs as single particles would, it is not certain if this exposure to the aggregated particles compared to single particles pose health risks but smaller particles appear to pose an increased health risk. Airborne aerosols such as dust, smoke and fumes can have major health implications especially in an industrial environment. This can range from numerous dust related respiratory conditions such as asthma and allergic alveolitis caused by mineral and organic dusts, infectious diseases being spread through airborne organisms or spores, heavy metal poisoning through metallic dusts such as lead, and lung cancers caused by dust particles of carcinogenic substances such as asbestos. Knowledge of how these particles cluster can be used to reduce hazard potential such as dust cloud formation.[3]Colloidal solutions are microscopically small particles in a solution that do not settle but remain suspended. It only in the liquid dispersed phase/ liquid continuous medium that clumping does not happen. Milk is the perfect example of a colloid as it never settles out, the milk particles are made up of fats which have the same charge; this means they repel each other, therefore, no clumping occurs. [4]Positively charged platelets, toxins and fats interact with red blood cells that are negatively charged causing them to aggregate. Clumping in red

blood cells is a visual indication that undigested protein is being absorbed into the cells. The accumulation of red blood cells sticking together, in a stack, means the blood tends to stay aggregated and cannot travel through the capillaries. Therefore, either in the long or short term tissues becomes oxygen deprived and are forced to endure their own metabolic waste which can be a cause for cancers.[5]The clumping process and growth of all these clusters is a statistically random process. The clumps may be held together by weak Van der Waals or other forces but initially the clumps have to form by chance collisions of particles. The aims of the project are to establish a robust theoretical model to predict the clumping of circular laminae randomly distributed in a 2D frame. In order to test Roachs theoretical model it will be necessary to generate a statistically significant number of randomly distributed laminae within this frame.[6] The number, N, of laminae randomly distributed within the frame will be allowed to vary in order to evaluate the effect of lamina density on the number and sizes of clumps. The expression clump will be used for a connected aggregate of laminae. Single non-overlapped particles will also be called clumps. Also an alternate equation was achieved quantitatively using LABFIT to fit the experimentally derived data to the theoretical models. Finally, the project will address how the theoretical models developed and tested are related to ‘ real life’ clumping situations such as environmental dust monitoring, nano-particles and health, bacterial counting and wound care after radiation therapy.

MATERIALS AND METHODS

Materials

The following software was installed according to the developer's instructions: ImageJ Version 1. 46 for Windows. ImageJ Graph plugin (2010/08/31)LABFIT Version 7. 2. 48Microsoft Excel 2010MINITAB 14

Methods

Using MINITAB 14 a model in which laminae of negligible thickness were located independently at random points on a two dimensional plane surface . Two columns of random integers were created with a minimum value of 0 and a maximum value of 2220. The rows of data created were, 50, 100, 150, 200, 250, 300, 400, 500, 600, 800, 1100, 1500 and 2220, these were then plotted in an X Y scatter graph. This replaced drawing the laminae by hand. Each scatter plot had laminae situated at random points, the number of these laminae was referred to as N, some of the laminae were isolated and others overlapped to form clumps. ImageJ is a computer program that can identify and label laminae, further to this, a plug-in was used to alter the radius of each lamina so they overlapped and produced a log of which laminae where overlapping. The simplest way to carry out this procedure was to create a macro. To work correctly the image that is being processed had to be in 8 bit and binary format, the scale set, the range of the particle size to be identified and the radius from the centroid determined. This was then saved as a log to an excel sheet.[Appendix 1]The log was saved to an Excel document; from this a Pivot Table was created. The pivot table allowed a large range of data to analysed and sorted. The table simplified the data

into amount of laminae in a clump verses the amount of clumps that were detected. This also produced the total number of clumps found and the actual number of laminae that the program recognised. This allowed specific data to be transferred to LABFIT for further analysis. LABFIT was used to fit the Roach equation to the experimental results and the ‘ Find Function’ was used to determine if it could detect a better equation for predicting random clumping.

RESULTS

C: UsersCHRIST~1AppDataLocalTempRar$DI00. 657500. tifFigure : Particle image that was created in MINITAB to be analysed in ImageJ. Figure : Scatter plots of clumps produced using MINITAB 14. X axis is the size of the clump, Y axis is the number of clumps. Range of laminae from 50-2220. Radius of laminae (left to right) 0. 1, 0. 2, 0. 3, 0. 4, 0. 5, 0. 6, 0. 7, 0. 8, 0. 9, 1. 0, 1. 1, 1. 2, 1. 3, 1. 4, 1. 5. Figure : Graph made in LABFIT showing Roach fit using the equation A*(4*pi*r^2*x^2)/(exp(4*pi*. r^2*x)-1) against the experimental scatter plots for laminae radius of 0. 1cm. Figure : Graph made in LABFIT showing Roach fit using the equation A*(4*pi*r^2*x^2)/(exp(4*pi*. r^2*x)-1) against the experimental scatter plots for laminae radius of 0. 4cm. Figure : Graph made in LABFIT showing Roach fit using the equation A*(4*pi*r^2*x^2)/(exp(4*pi*. r^2*x)-1) against the experimental scatter plots for laminae radius of 0. 8cmFigure : Graph showing Roach fit using the equation A*(4*pi*r^2*x^2)/(exp(4*pi*. r^2*x)-1) against the experimental scatter plots for laminae radius of 1. 5cm. radius/cmCorrelation Coefficient (R2)Value of A0. 10. 9991. 010. 20. 9990. 9970. 30. 9980. 9850. 40. 9970.

9650. 50. 9530. 9710. 60. 9550. 9250. 70. 9570. 9610. 80. 9720. 9980. 90. 9581. 0410. 8281. 221. 10. 9451. 051. 20. 9481. 081. 30. 9481. 11. 40. 9691. 141. 50. 951. 2Figure : Optimisations of radius (r) for fitting with Roach equation A*(4*pi*r^2*x^2)/(exp(4*pi*. r^2*x)-1)PARAMETERS: Mean UNCERTAINTIES: SD t P(t)A = 0. 12958531240E+01 SIGMAA = Calculation to proceed --- ----B = -0. 78112110971E-02 SIGMAB = Calculation to proceed --- ----Chi-Square: Deg. Freed.= 11 ChiSq.= 0. 511143E+03 Red. ChiSq.= 0. 464675E+02 => P(Red. ChiSq.)= 0. 000Analysis of Variance: F = (Sum Sq./Deg. Freed.)_reg / (Sum Sq./Deg. Freed.)_error => F = (0. 69947E+04/1) / (0. 51114E+03/ 11) = 0. 1505E+03 => P(F) = 0. 00Standard Deviation of the Fitting: 0. 681671E+01Correlation Coeficient: R^2yy(x) = 0. 9355535E+00 adjR^2yy(x) = 0. 9296947E+00Figure : Using the Find Function in LABFIT for 1. 2 (r= 1. 2/40=. 03) produced the equation1. 29*N*EXP(-. 0078*N) = A*N*EXP(0. 0078 *N). Figure : Graph showing LABFIT Find Functions Equation (Y = A*X*EXP(B*X) for laminae radius of 1. 2cm. Figure : Comparison of the number of isolated clumps with a radius of 0. 5 cm expected using Roach equation and the total number of clumps found by the experiment. Figure : Reproducibility for 9 runs (500 particles, radius = 0. 5cm, 20x20cm lattice)Figure : Graph showing simplified " degrees of freedom" measurement.

Discussion

Roach’s results table on clumps produced by the overlapping of circular laminae on a 1000 x 1000 lattice show that his observed and expected results were very similar.[7] The small variations may have happened for

numerous reasons including the fact the each circle was hand drawn from a point and possible human error could have occurred, along with points being covered by overlapping circles making them increasingly difficult if not impossible to count. This is more likely why; when a cluster was greater than eight he grouped them together. If Roach had defined the clusters to a higher result, his observed and expected results may have matched even better, an increase in data would allow the robustness of his equation to be fully tested. Moreover, the distribution was random therefore no results are going to be identical, this means that there will always be slight variations in results, but their significance needs to be measured. An automated image analysis procedure, ImageJ, was developed in order to quantify the numbers and sizes of clumps. In any case the unavoidable errors associated with manual counting are best avoided in curve fitting determinations. To achieve this, a range of images of lamina were produced[Fig 1], of varying N, randomly distributed within a frame of known dimension. Testing and developing of ImageJ to count the numbers of clumps within each frame and post processing these results in excel to utilise the ‘ pivot table’ facility ensured optimised accuracy. The ‘ fit’ for the total number of clumps and for clumps of known membership with theoretical models developed by Roach was also investigated. Using LABFIT also permitted fitting the amassed experimental data to a large library of standard functions to assess fit. An ‘ informed’ reassessment of the theoretical models was carried out by comparison with the ‘ best’ fitting mathematical functions. This was completed for $0 < N < 2220$ and all sizes of measured clumps. Overall the generation of random clumps in the Minitab simulation was very successful

because it trials the Roach equations under a very wide range of conditions. This was much wider and far more accurate than Roach could have achieved with manual counting. The random points that were produced in MINTAB , this shows the entire experimental scatter plots produced ranging from 50 laminae to 2220 laminae with an increasing radius of 0. 1cm. [Fig. 2] The LABFIT graphs were produced using Roaches equation to predict the total number of clumps from the randomly plotted laminae (N). Roaches equation was able to predict the number of plotted laminae that would form clumps. [Fig. 3]As the amount of lamina increased as seen in , the fit is still good but not optimised, the reason for this is that the fit line begins to curve and Roach's equation only seems to be able to fit well for straight lines.[Fig. 4]The graphs begins to level out again,[Fig. 5] and once again the prediction by Roach's equation becomes increasingly accurate as the number of laminae increase. [Fig. 6]The correlation coefficient also confirms that Roaches equation works most effectively on laminae with a smaller radius. The value of A is a pie exponential term, LABFIT needs a variable, in an ideal situation A is equal to one.[Fig 7] The reason why smaller laminae fit better is because the lamina covers less of the overall defined area. In Roach's experiment, the percentage of area occupied by one 0. 5cm disc is 0. 002%, in my fit the percentage of area occupied by one 0. 75 disc is 0. 004% meaning the percentage of overall area covered by a single disc has doubled and the disc size has not. Using the ' Find Function' in LABFIT, a new exponential straight line equation was developed. Y = 1. 29*N*EXP(-. 0078*N). In terms of π, r, B is equal to 0. 0078 is (πr)2, (πr)2 = 0. 0088 and this seems to the closest set of constants which fit the value 0. 0078. [Fig. 8,

9]Roach’s equation produces good predictions for all clumps and 1's, but overestimates for all larger clumps.[Fig. 19] The possible reason for this is that Roach obtained the total number of clumps of all sizes by considering a sequence of Bernoulli trials but this cannot be rigorously applied to the formation of clumps greater than size one. The Bernoulli trials process is the mathematical abstraction of coin tossing, but because of its wide applicability, it is usually stated in terms of a sequence that has two possible outcomes, success and failure. The trials are independent and the outcome of one trial has no influence over the outcome of another trial. This means that it asks the theoretical question that if a single lamina is put down one at a time in a defined area, will it be beside another lamina or not. The more laminae, the more likely it will be beside another such as a success failure fit, therefore larger clusters are produced when more laminae are introduced. On each trial, the probability of success is p and the probability of failure is 1-p, so a Bernoulli trial is very good at predicting if a disc landing in a defined area lands on its own or joins a clump but it is not good at predicting the size of the clumps it joins up with, only the overall number of clumps.[8]Where N is the number of measurements, xi is each individual measurement, and x-baris the mean of all measurements. The quantity (xi - x-bar) is called the " residual" or the " deviation from the mean" for each measurement. The quantity (N -1) is called the " degrees of freedom" for the measurement.[Fig. 11] Which can be reduced to S 1/. So there is not a lot to be gained by doing hundreds of measurements because Ö1/N which is proportional to the standard deviation falls off quite quickly then flattens out. So doing 100's of measurements is not the way to reduce standard deviation.[Fig. 12]The

imaging software can be used in a variety of other tasks such as environmental dust monitoring. Glass plates are used to collected settled dust particles which are later analysed. This analysis may happen after the damage has already been caused by over exposure. Using this imaging software an automated warning system can be developed to alert users in either a confined or large area that the environmental air is becoming contaminated.[9] Moreover this software can be used to quantify bacterial cells on a plate. This time consuming inaccurate process could be automated to remove human error providing increased accuracy and detecting cells that are overlapping which may have been overlooked.[10] Using the modelling method, predictions of cell morphology can be made,[these clumping models can also be used to predict better treatment strategies for preventing radiation damage accurate clumping data of the radioactive particles would allow them to be removed effectively directly after treatment via cleaning solutions compared to staying within the epidermis for several months causing desquamation.[11] Producing nano particle sprays that do not pose health risks and can be fully tested to form large enough particles so they will be filtered out by the airways before being introduced into the blood stream.[12]

Conclusion

Simulation using random plots is a valid generator, this allowed multiple data sets to be produced quickly and randomly. Using the ImageJ software combined with the macros that were created allowed very efficient analysis of these scatter plots, more so that any manual counting methods could.

ImageJ also allows an increased amount of variables to be analysed such as the number of laminae and their radius, further to this it was absolutely better than manual counting as it eliminated manual counting errors. Roach's equation offers excellent prediction for all clumps and for clumps of one, but for clumps greater than one the fit is poor. LABFIT offers a good alternative to roach which fits to the equation A*N*EXP(0. 0078 *N), this equation could be used for clumps that are greater than one and could produce better results for clumps with increase laminae. Further work should investigate measuring different size of lamina on the same lattice, investigation into the Monte Carlo non circular laminae approach and 3 Dimensional Poisson distributions.[13, 14]A more complete understanding of the prediction of the distribution of points, lamina and spheres in two dimensional and three dimensional spaces will contribute to a more complete predictive framework necessary for the understanding of the microscopic structure of particulate assemblies ranging from nano to micrometre sized components. This work has concentrated on same sized lamina on a 2 dimensional lattice, but current image analysis tools are able to reliably count variable size and non regular lamina in 3 dimensions.
The currently available theory for predicting variable size and non-regular lamina has been developed using Monte Carlo methodology. These theories await rigorous evaluation using the methodology which has evolved during the course of this work.

References

Figures