How to Effectively Visualize & Quantify a 3-D Image (or How Long Do We Mix to Get Homogeneous Product)?

Tom Filloon (Statistics), Dave Dunlop (Beauty) Procter & Gamble - Cincinnati, Ohio

Abstract

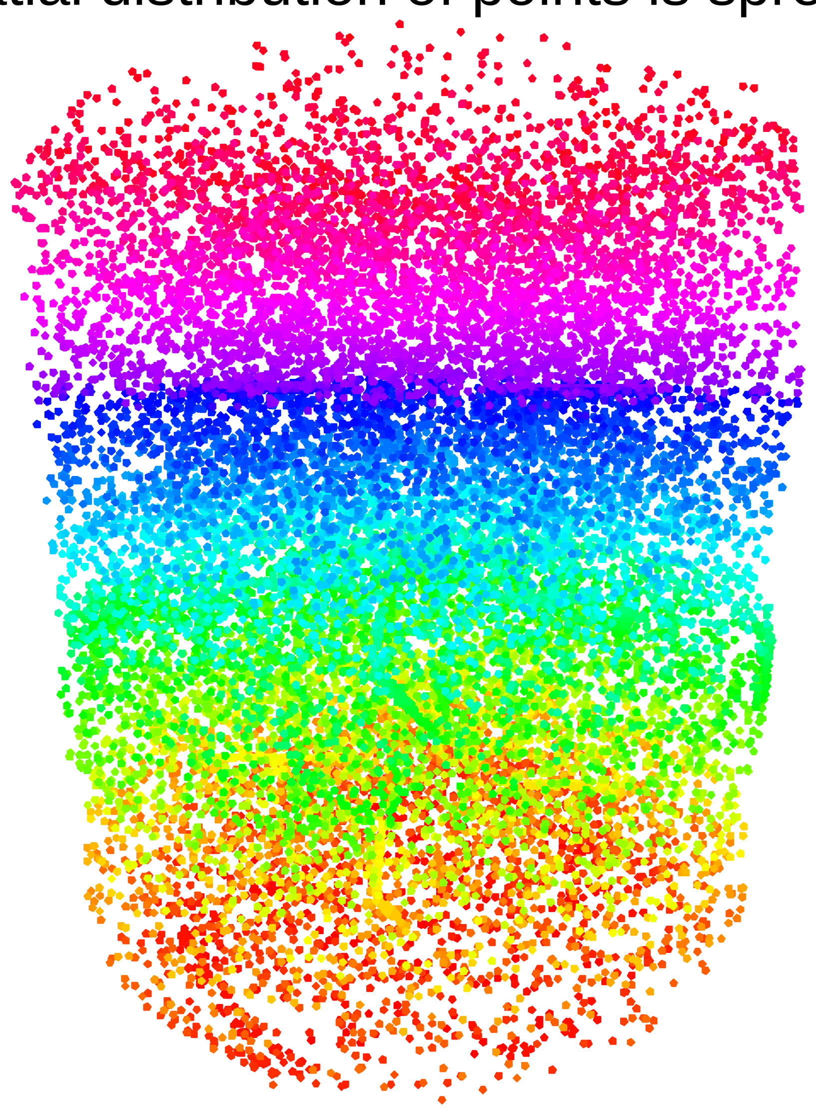
To determine if we have a homogeneous product after mixing, we need to understand the spatial (3D) distribution of particles and how to evaluate whether particles are randomly dispersed or not. It will be shown how nearest neighbor distances can be used for evaluating product homogeneity.

References

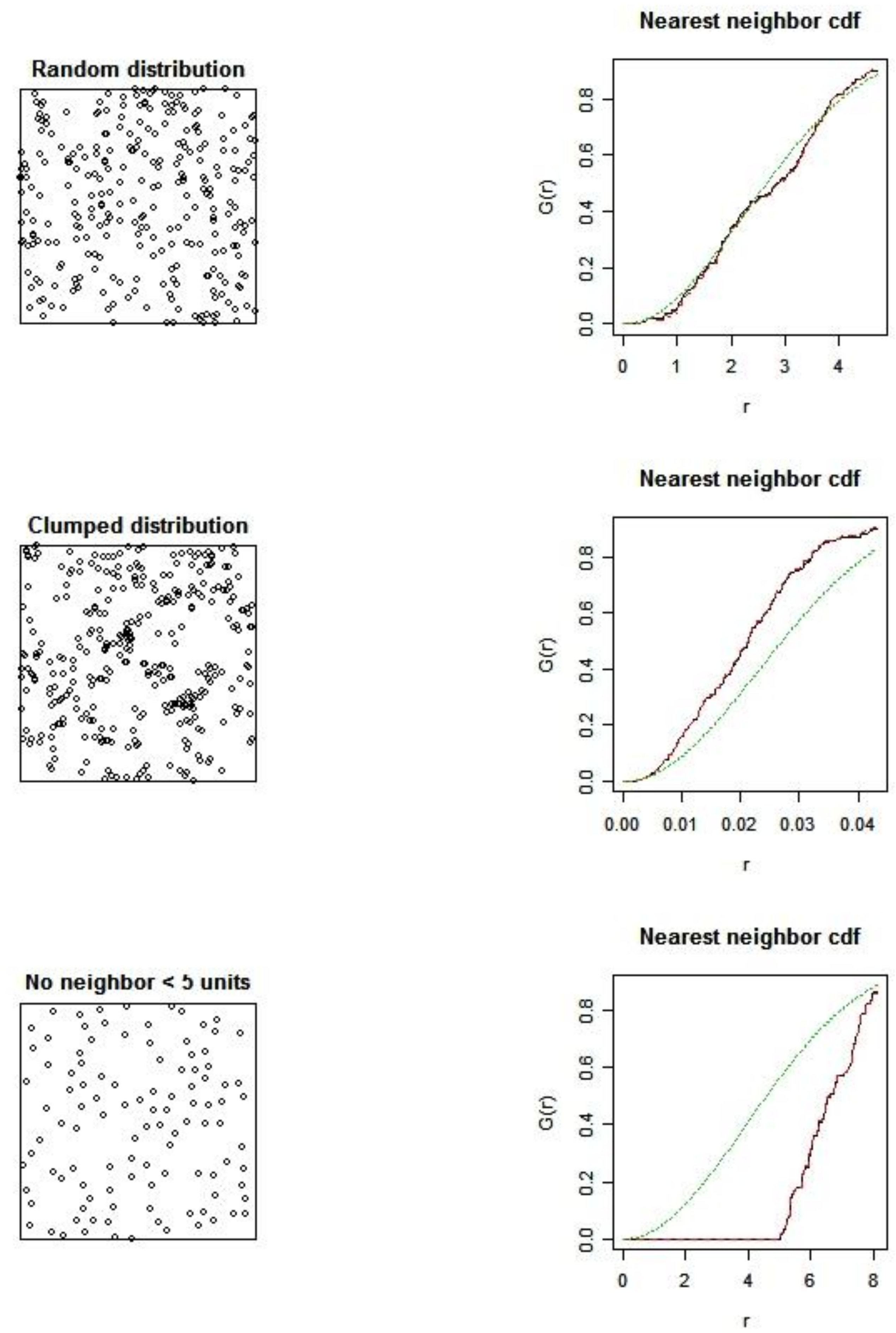
1. Ripley, B **Spatial Statistics**.
2. Cressie, N **Statistics for Spatial Data**
3. TG Filloon SLR April 2007

Introduction

How long does one need to mix a large mixing tank before the resulting product is homogeneous? Computer simulations are run to perform virtual mixing experiments. A large number of virtual particles (~20,000) are tracked over time while a computer simulates how they would move in a large cylindrical mixing tank in operation [*see picture below*]. Then after a given amount of time, one can visualize the spatial distribution of points by knowing their exact locations. The 'statistical' task then is to somehow quantify this 3-dimensional data into a summary measure as to whether the spatial distribution of points is spread out enough.



Spatial Statistics – Point Processes

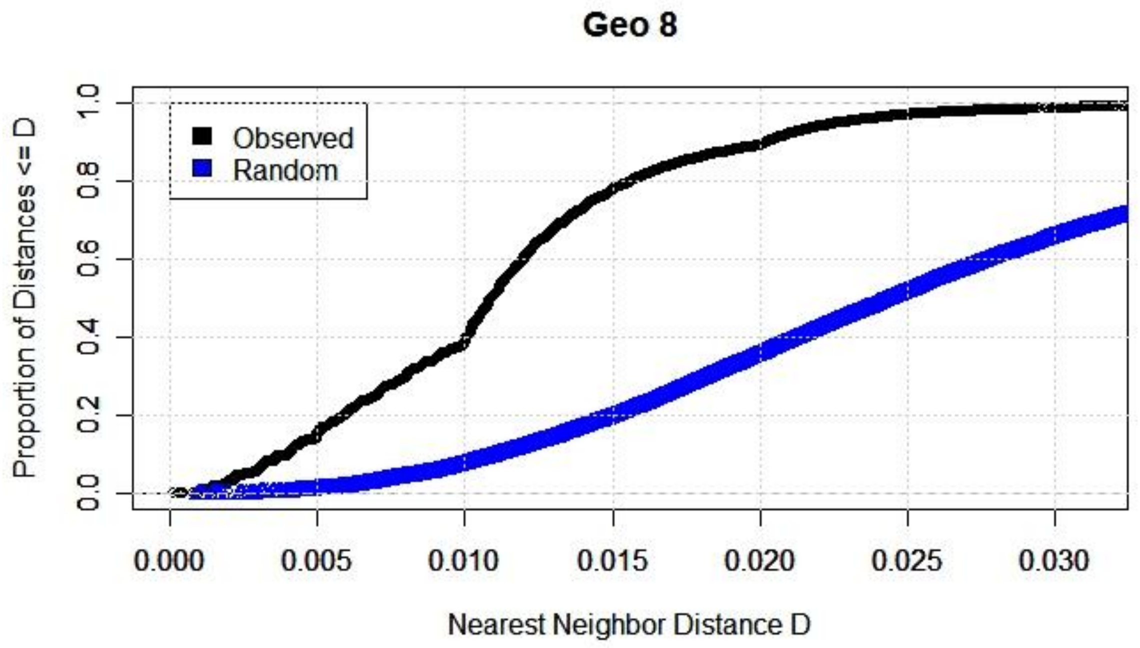


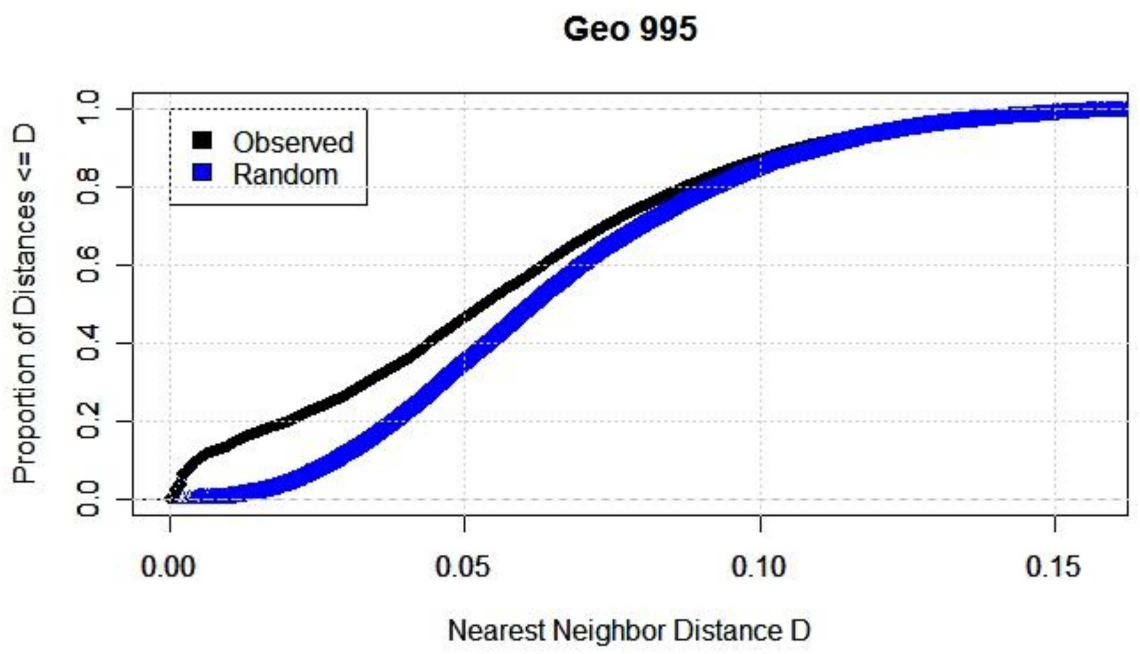
One approach to analyzing such data is to determine the average number of points within radius r of a random point, which yields a function K(r). If the K function is too large for small values of r (relatively to random Poisson process), this would indicate clustering /clumping of data.

Another more intuitive measure (**that will be used here**) is to determine, for each data point, the distance to its 'nearest neighbor' and then to compare this distribution of nearest neighbor distances to what would be expected if the data were from a random Poisson process. We will construct the empirical cumulative distribution function of these nearest neighbor distances and define it as G(r). Hence, G(r) represents the proportion of data points whose nearest neighbor is less than or equal to r units away, and will take the classical stair-step form.

3-D Nearest Neighbor Distributions

2 Simulations
– notice how Geo 995 simulation has been better mixed (closer to random distribution)





Computational Approach

Fortunately, via R-news (*thanks go to NHH Roger Bivand, for making me aware of UMD David Mount's work*) , I was able to get an off-CRAN R package (ann) that does inter-point distances calculations efficiently (C code) and hence, this implementation is done via the R software. An additional R package (rgl) allows for interactive 3-D plotting of the points to enable visualization of data and this statistical summary all in the same software. Furthermore, since the R software is free, I have loaded all of this capability onto the collaborator's desktop for his ease of use as this project moves forward.

Summary

A user-friendly, fast approach for interactive visualization & understanding of a spatial distribution of points was freely available within the R software (libraries *rgl, ann*). This approach is simple and generalizable to any 3-D (or 2-D!) problems.