

This paper is dedicated to the memory of Peter G. Hall. It concerns a deceptively simple question: if one observe variables corrupted with measurement error of possibly very complex form, can one recreate asymptotically the clusters that would have been found had there been no measurement error? We show that the answer is yes, and that the solution is surprisingly simple and general. The method itself is to simulate, by computer, realizations with the same distribution as that of the true variables, and then to apply clustering to these realizations. Technically, we show that if one uses K-means clustering or any other risk minimizing clustering, and a multivariate deconvolution device with certain smoothness and convergence properties, then, in the limit, the cluster means based on our method converge to the same cluster means as if there is no measurement error. Along with the method and its technical justification, we analyze two important nutrition data sets, finding patterns that make sense nutritionally.