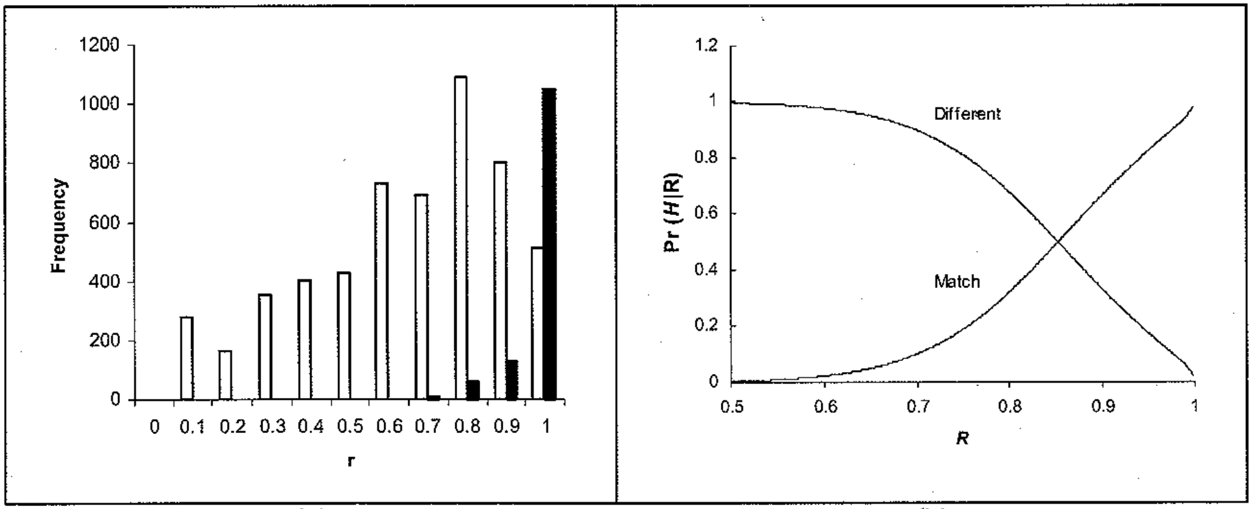
BAYESIAN PROBABILITIES FOR MATCHING SPECTRA

D Brynn Hibbert, Jianfeng Li
School Of Chemistry, University Of New South Wales, Sydney, NSW 2052 Australia
b.hibbert@unsw.edu.au

Methods (described elsewhere) to match spectra using a point-to-point correlation of digitised spectra give statistics that must further be interpreted. Typical use by the New South Wales Environment Protection Authority uses descriptors such as “definite match”, “possible match” and “no match” based on the experience of the analyst and due consideration of aspects of the incident. The use of Bayes' theorem allows inference of a probability of a given hypothesis among a number of hypotheses. For example, an oil spill may be hypothesised to have come from one of two possible sources (H_1, H_2) or neither (H_0). The evidence is the matching statistic between the spectrum of the spill and that of each suspect source (r_1, r_2). What is needed is the distribution of the statistic of all matched spectra (S(R)), and of all different spectra (D(R)). Figure 1(a) shows such distributions for fluorescence spectra of diesels. (These may be derived from historical data held by the EPA).



(a) (b)

Figure: (a) Histogram of correlations of FTIR spectra. Filled bars – replicates of same oils, Open bars – different oils. (b) Bayesian probabilities corresponding to the distributions in (a) for a given matching statistic R.

Bayes' theorem, applied to this problem, states that the probability of an hypothesis H_i given evidence E is related to the likelihood of the hypothesis $Pr(E|H_i)$ and its prior probability $Pr(H_i)$ by $Pr(H_i|E) = Pr(E|H_i) \times Pr(H_i) / [\square Pr(E|H_i) \times Pr(H_i)]$ where the sum is over all hypotheses (which must be exclusive and cover all possible outcomes). We show that this theorem leads to a straightforward assignment of a probability (Figure 1(b)) to matching statistics from spectra.