Supplementary Info. In house developed MATLAB function for misclassification error rate

```
function y=estmisclrate(T,nnn)

%function to determine misclassification error rate when the true number of
clusters is known and the numbers of cluster members are equal
%nnn- vector determining the boundaries of the clusters in the simulated data
% example of nnn: [0 20 40 60 80 100] - represents situation of 5 clusters
% with 20 members in each cluster
%T- vector representing cluster membership dtermined by some clustering
%method
%example of T: [1 1 1 1 3 3 2 3 1.....4 4 4 3 4 1]
% the length of T should be equal to last number in nnn, here length(T)=100
% note that the exact values of labels T(i) are of no importance,
%i.e. it is OK to label all members of cluster #1 as cluster #5
%as soon as all members of cluster #5 are labeled as members of cluster #1
%same is right for any permutations of cluster labels
Kn=length(nnn)-1; %number of clusters
% count the number labels (e.g.i=1:5) within the boundaries of the true
clusters
%e.g. b(1,1) is the number of '1' in the the first 20 numbers of T(k),
%b(1,2) is the the number of '2' in the first 20 numbers of T(k),
%b(2,1) is the number of '1' in the second 20 numbers of T(k)
 for k=1:Kn
    for i=1:Kn
        a= find(T(nnn(k)+1:nnn(k+1))==i);
        b(k,i)=length(a);
        clear a
    end
end
%determine all the possible permutations of cluster labels
x=[1:Kn];
xp=perms(x);
Lxp=size(xp,1);
%determine the number of labels within the boundaries of the true clusters
%for all possible permutations of the labels
% determine the permutation for which the sum of non-diagonal elements is
% the smallest
%since the 'names of the clusters do not matter' this provides the value of
%missclassification error rate 'y'
for n=1:Lxp
    for k=1:Kn
        for i=1:Kn
            B(k,i)=b(k,xp(n,i));
        end
    end
   dB=diag(B);
   ssB=sum(sum(B));
   miscl(n)=1-sum(dB)/ssB;
end
y=min(miscl);
end
```