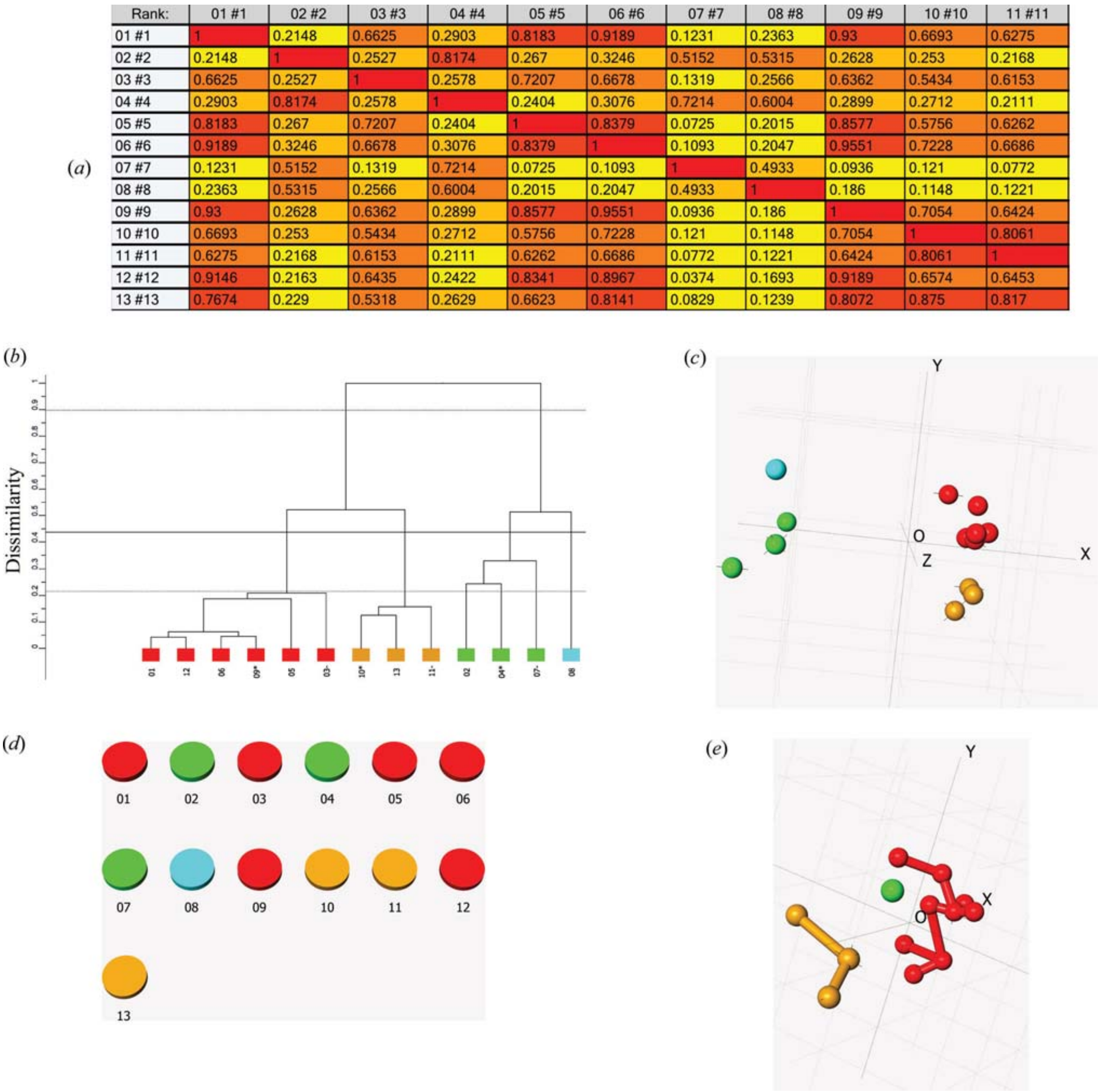
shows the default minimum spanning tree with 12 links. In Fig. 3.8.9(*f*) the scree plot indicates that three clusters will account for more than 95% of the data variability. The steep initial slope is a clear indication of good cluster estimation. The silhouettes are shown in Fig. 3.8.9(*g*–*i*). These were discussed in Section 3.8.5.1. In Fig. 3.8.9(*j*) the default parallel-coordinates plot for the same data is shown, and in Fig. 3.8.9(*k*) there is another view taken from the grand tour. These two plots validate the clustering and also indicate that there is no significant error introduced into the MMDS plot by truncating it into three dimensions.

3.8.6.1.1. *Aspirin data with amorphous samples included*

As a demonstration of the handling of data from amorphous samples, five patterns for amorphous samples were included in the aspirin data and the clustering calculation was repeated. The results are shown in Fig. 3.8.10. Fig. 3.8.10(*a*) shows the dendrogram. It can be seen that the amorphous samples are positioned as isolated clusters on the right-hand end. They also appear as an isolated cluster in the MMDS plot and the parallel-coordinates plots, as shown in Figs. 3.8.10(*b*) and (*c*). It could be argued that these samples should be treated as a single, five-membered cluster rather than five individuals, but we have found that this confuses the clustering algorithms, and it is clearer to the user if the data from amorphous samples are presented as separate classes.

3.8.6.2. *Phase transitions in ammonium nitrate*

Ammonium nitrate exhibits temperature-induced phase transformations. Between 256 and 305 K it crystallizes in the orthorhombic space group *Pmmm* with $a = 5.745$, $b = 5.438$, $c = 4.942$ Å and $Z = 2$; from 305 to 357 K it crystallizes in *Pbnm* with



Rank:	01 #1	02 #2	03 #3	04 #4	05 #5	06 #6	07 #7	08 #8	09 #9	10 #10	11 #11
01 #1	1	0.2148	0.6625	0.2903	0.8183	0.9189	0.1231	0.2363	0.93	0.6693	0.6275
02 #2	0.2148	1	0.2527	0.8174	0.267	0.3246	0.5152	0.5315	0.2628	0.253	0.2168
03 #3	0.6625	0.2527	1	0.2578	0.7207	0.6678	0.1319	0.2566	0.6362	0.5434	0.6153
04 #4	0.2903	0.8174	0.2578	1	0.2404	0.3076	0.7214	0.6004	0.2899	0.2712	0.2111
05 #5	0.8183	0.267	0.7207	0.2404	1	0.8379	0.0725	0.2015	0.8577	0.5756	0.6262
06 #6	0.9189	0.3246	0.6678	0.3076	0.8379	1	0.1093	0.2047	0.9551	0.7228	0.6686
07 #7	0.1231	0.5152	0.1319	0.7214	0.0725	0.1093	1	0.4933	0.0936	0.121	0.0772
08 #8	0.2363	0.5315	0.2566	0.6004	0.2015	0.2047	0.4933	1	0.186	0.1148	0.1221
09 #9	0.93	0.2628	0.6362	0.2899	0.8577	0.9551	0.0936	0.186	1	0.7054	0.6424
10 #10	0.6693	0.253	0.5434	0.2712	0.5756	0.7228	0.121	0.1148	0.7054	1	0.8061
11 #11	0.6275	0.2168	0.6153	0.2111	0.6262	0.6686	0.0772	0.1221	0.6424	0.8061	1
12 #12	0.9146	0.2163	0.6435	0.2422	0.8341	0.8967	0.0374	0.1693	0.9189	0.6574	0.6453
13 #13	0.7674	0.229	0.5318	0.2629	0.6623	0.8141	0.0829	0.1239	0.8072	0.875	0.817

Figure 3.8.9

The complete cluster analysis for the aspirin samples. (*a*) The correlation matrix, which is the source of all the clustering results. The entries are colour coded: the darker the shade, the higher the correlation. (*b*) The dendrogram. The colours assigned to the samples are used in all the visualization tools. (*c*) The corresponding MMDS plot. The clustering defined by the dendrogram is well defined. (*d*) The pie-chart view. (*e*) The minimum spanning tree.

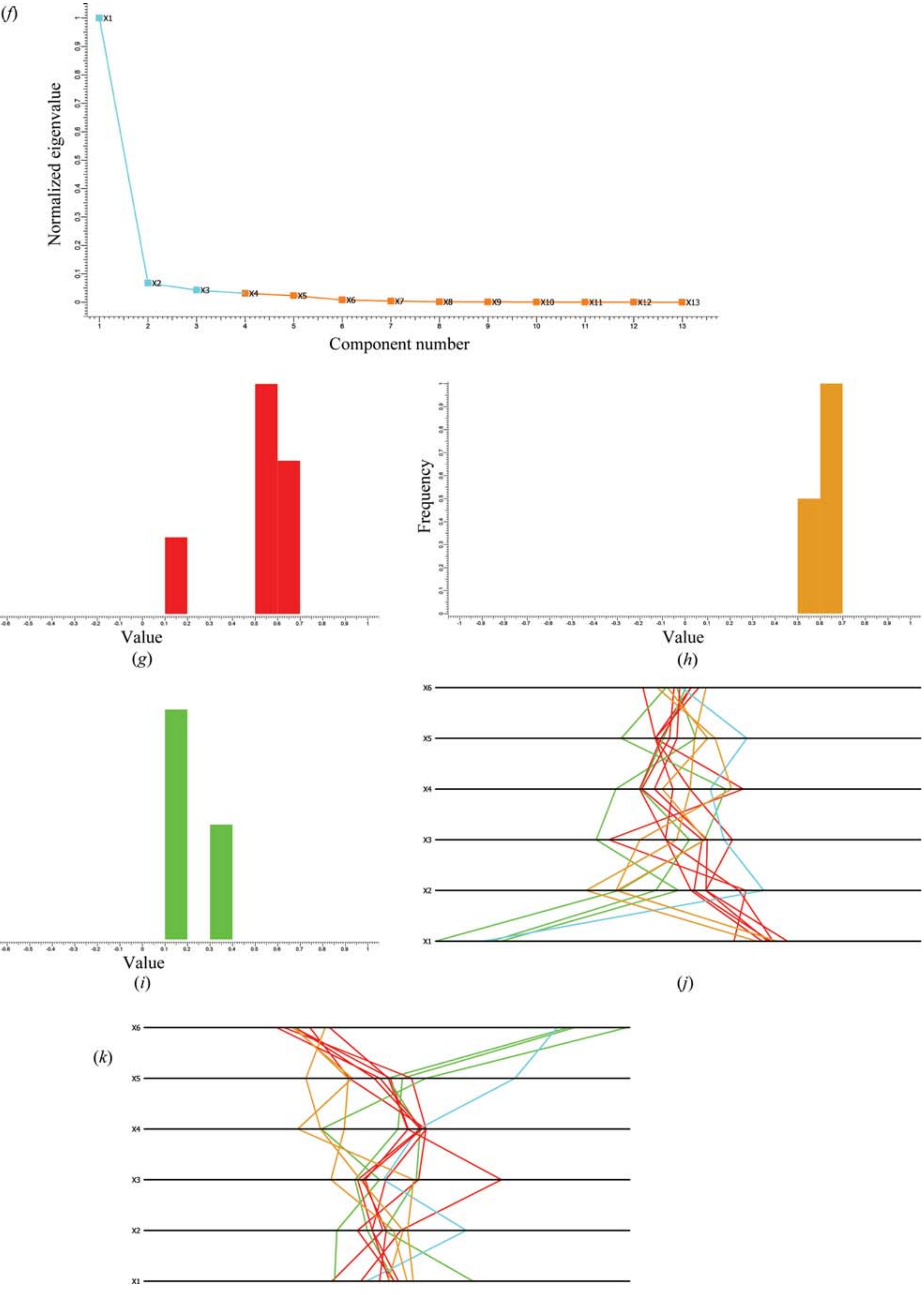


Figure 3.8.9 (continued)
The complete cluster analysis for the aspirin samples (continued). (*f*) The scree plot. It indicates that three clusters explain 95% of the variance of the distance matrix derived from (*a*). (*g*–*i*) The silhouettes for the red, the orange and the green clusters, respectively. These are discussed in detail in the caption to Fig. 3.8.8. (*j*) The default parallel-coordinates plot. The clusters are well maintained into the 4th, 5th and 6th dimensions. (*k*) Another view of the parallel coordinates using the grand tour. The clustering remains well maintained in higher dimensions.

references