Supplementary Material

The original data matrix including all single index values is shown in Table S1 of the supplementary material. The list of 30 substances along with their structures and cluster numbers can be found in Table S2. Chapter S 3.3.1 describes the minimal mathematical example we constructed and calculated to first establish our process.

Table S1. Original Data Matrix Including All Single Index Values Subsequently Subjected to the Combined Cluster Analysis and Principal Component Analysis. This Table was Imported into MATLAB® according to Chapter 3.3.

	ZI	WI	RI	BI	IIAC	IIMB	IIEC	TII	MM	WS (g/L)	VP (hPa)
Acetone	1,2955	0,4529	1,5710	8,6095	4,8000	11,1000	0,4155	0,4231	5,908	10000,0000	246,00
Cadaverine	1,2060	0,0000	2,8685	9,6166	5,3333	37,3333	0,4232	0,2900	4,866	1021,8000	0,90
Dimethyldisulfide	1,3710	0,0000	1,3710	16,3241	4,6000	12,7000	0,4207	0,3563	9,420	2,5000	38,00
Acetic Acid	1,5000	0,5177	2,4056	11,5733	4,2500	7,8750	0,4352	0,4529	7,506	602,9000	16,00
2 -- Heptanone	1,1433	0,2636	3,6635	9,9016	5,4545	40,7727	0,4161	0,2781	5,190	4,3000	4,50
n -- Heptane	0,8865	0,0000	2,8057	7,9391	5,5652	42,9565	0,4130	0,2770	4,357	0,0500	47,00
Piperidine	1,1661	0,0000	2,7933	9,1015	5,8824	23,0588	0,4209	0,2096	5,009	10000,0000	33,00
Propionic Acid	1,4354	0,4264	2,8454	11,3557	4,7273	13,0909	0,4301	0,4044	6,735	20,2500	4,00
Pyrrolidine	1,1981	0,0000	2,3788	8,7377	5,8571	16,7143	0,4218	0,2418	5,080	10000,0000	170,00
Skatol	1,2448	1,0526	3,9977	13,0239	5,5789	30,6316	0,4519	0,1275	6,904	0,4500	1,33
Toluene	0,9968	0,9710	3,0566	10,2249	5,2000	21,2000	0,4450	0,1979	6,143	0,4700	29,00
Triethylamine	1,0907	0,0000	2,0254	8,9828	5,4545	38,4545	0,4144	0,2970	4,600	133,0000	69,00
Water	0,9183	0,0000	0,9183	4,7354	2,0000	1,3333	0,4714	0,5443	6,007	10000,0000	23,40
Benzene	1,0000	1,0000	1,0000	10,5061	5,0000	14,5000	0,4553	0,2300	6,509	1,8000	100,00
4 -- Isopropenyl -- 1 -- Methylcyclohexene	0,9612	0,3912	4,0270	9,8942	5,6923	49,0000	0,4266	0,1500	5,240	0,0300	1,90
Ethanol	1,2244	0,0000	2,4194	7,3672	4,6667	9,1111	0,4234	0,4605	5,119	10000,0000	58,00
Pyridine	1,3486	0,9940	2,7322	12,5100	4,9091	12,7273	0,4579	0,2387	7,191	10000,0000	20,50
2, 5 - Dimethylpyrazine	1,4056	0,9544	2,4056	13,6475	5,2500	25,5000	0,4395	0,1743	6,759	32,0000	4,00
Hexanal	1,1674	0,2933	3,5766	9,8437	5,3684	31,2632	0,4181	0,3115	5,272	5,0000	12,00
1, 3 -- Pentadiene	0,9612	0,6000	3,1808	8,3981	4,6154	18,7692	0,4366	0,3336	5,240	0,6900	530,00
n -- Hexane	0,8813	0,0000	2,4464	7,4547	5,5000	33,8000	0,4125	0,3041	4,309	0,1600	162,00
Methylcyclopentane	0,9183	0,0000	2,7947	7,7379	6,0000	24,9444	0,4167	0,2084	4,676	0,0420	147,00
Heptanal	1,1433	0,2636	3,7888	10,0792	5,4545	40,0909	0,4179	0,2834	5,190	1,2500	0,86
Benzaldehyde	1,2958	1,2958	3,2359	13,6649	5,0000	19,1429	0,4553	0,2044	7,580	3,3000	1,30
n -- Decane	0,8960	0,0000	3,2028	9,0423	5,6875	76,4375	0,4141	0,2186	4,446	0,00005	1,00
3 -- Octanone	1,1239	0,2399	3,8635	10,1408	5,4167	49,3333	0,4194	0,2495	5,128	4,5000	2,00
1 -- Octen -- 3 -- ol	1,1239	0,2399	4,0537	10,4005	5,3600	50,1600	0,4266	0,2588	5,128	0,0000	1,00
Octanal	1,1239	0,2399	3,9737	10,2971	5,5200	49,9200	0,4177	0,2600	5,128	0,5600	1,00
Nonanal	1,1078	0,2204	4,1375	10,4996	5,5714	60,7500	0,4176	0,2401	5,080	0,0960	0,35
2 -- Nonanone	1,1078	0,2204	4,0391	10,3601	5,5714	61,5000	0,4162	0,2370	5,080	0,3700	0,83

Table S2. Molecular Structure of the Substances Analyzed in the Combined Cluster Analysis and Subsequent Two-step Principal Component Analysis

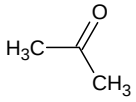
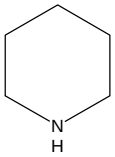
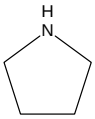
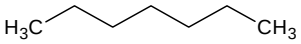
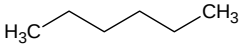
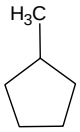
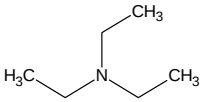
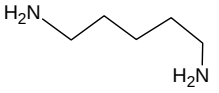
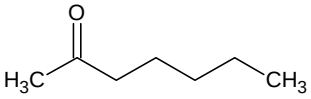
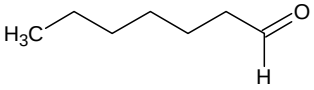
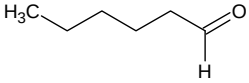
Name	Structure
Cluster 1	
Acetone	
Name	**Structure**
Ethanol	CH_3CH_2OH
Piperidine	
Pyrrolidine	
Cluster 2	
n-heptane	
n-hexane	
Methylcyclopentane	
Triethylamine	
1,3-pentadien	
Cluster 3	
Cadaverine	
2-heptanon	
Heptanal	
Hexanal	

Table S2. Contd…..

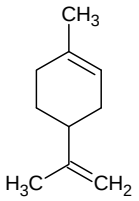
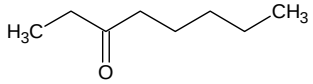
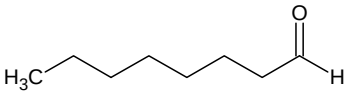
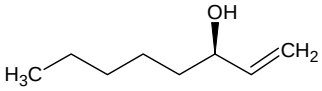
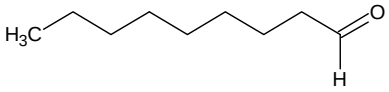
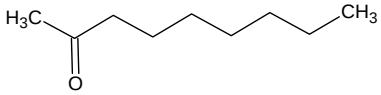
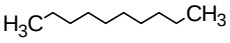
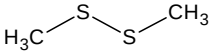
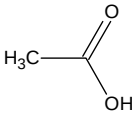
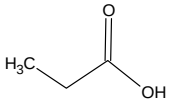
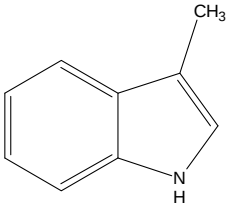
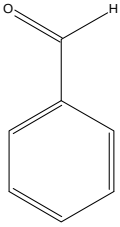
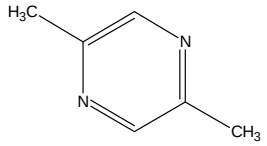
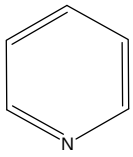
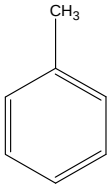
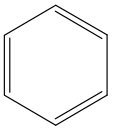
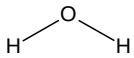
4-isopropenyl-1-methylcyclohexene	
3-octanone	
Octanal	
1-octen-3-ol	
Nonanal	
2-nonanon	
n-decane	
Cluster 4	
Dimethyldisulfide	
Cluster 5	
Acetic acid	
Propionic acid	
Cluster 6	
Skatole	
Benzaldehyde	

Table S2. Contd…..

2,5-dimethylpirazine	
Pyridine	
Toluene	
Benzol	
Cluster 7	
Water	

S 3.3.1. Descriptive Minimal Mathematical Example

Here we present our procedure first on a constructed data set. The set is minimal as we need at least three indices for being able to apply our two-step procedure without getting trivial intermediate results for our Principal Components. Furthermore, we need at least 3 different clusters which should ideally be represented by only two Principal Components. Of course, those clusters should not all contain only one substance. Thus, we set up with a minimum of four substances spread over three clusters. The three index values for each substance s (index number indicates cluster) can be mathematically expressed *via* the following vectors:

$$s_{1a} = \begin{pmatrix} 15 \\ -9 \\ 21 \end{pmatrix};; s_{1b} = \begin{pmatrix} 21 \\ -9 \\ -3 \end{pmatrix};; \; s_2 = \begin{pmatrix} 21 \\ -6 \\ 12 \end{pmatrix} and \; s_3 = \begin{pmatrix} 9 \\ -18 \\ 0 \end{pmatrix}$$

The three main foci of the clusters result in:

$$c_1 = \begin{pmatrix} 18 \\ -9 \\ 9 \end{pmatrix};; \; c_2 = \begin{pmatrix} 21 \\ -6 \\ 12 \end{pmatrix} and \; c_3 = \begin{pmatrix} 9 \\ -18 \\ 0 \end{pmatrix}$$

In order to achieve an equal weight and subsequently an equal contribution to the further data processing, a ``z-score" transformation is applied. First, this includes a shift among the arithmetic mean

$$\mu = \begin{pmatrix} 16 \\ -11 \\ 7 \end{pmatrix}$$

and as a second step, data are scaled regarding the standard deviation in each component. As the standard deviation is already equal in each component of our minimal example, we passed on without the scaling step as it would not influence our result at all but facilitate further calculations. The calculated index values for the transformed data resulted in the shifted vectors for the substances:

$$\tilde{s}_{1a} = \begin{pmatrix} -1 \\ 2 \\ 14 \end{pmatrix};; \tilde{s}_{1b} = \begin{pmatrix} 5 \\ 2 \\ -10 \end{pmatrix};; \tilde{s}_2 = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} \text{ and } \tilde{s}_3 = \begin{pmatrix} -7 \\ -7 \\ -7 \end{pmatrix}$$

and the main foci:

$$\tilde{c}_1 = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix};; \tilde{c}_2 = \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix} \text{ and } \tilde{c}_3 = \begin{pmatrix} -7 \\ -7 \\ -7 \end{pmatrix}$$

The correlation matrix of the main foci is given by

$$\begin{pmatrix} 78 & 78 & 78 \\ 78 & 78 & 78 \\ 78 & 78 & 78 \end{pmatrix}$$

and the corresponding eigenpairs are calculated to the following eigenvalues and eigenvectors:

$$\lambda_1 = 234, v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix};; \lambda_{2,3} = 0, v_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \text{and } v_3 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Each eigenvector is a Principal Component with its eigenvalue representing its weight in terms of original information it contains. So the first Principal Component in our minimal descriptive example is as follows:

$$pc_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

With the first Principal Component all transformed data of the substances can be expressed as:

$$\left(\tilde{s}_{1a}\ \tilde{s}_{1b}\ \tilde{s}_2\ \tilde{s}_3\right) = \left(\begin{array}{ccc} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 1 & -1 & -1 \end{array}\right)\left(\begin{array}{cccc} 5 & -1 & 5 & -7 \\ -7.5 & 7.5 & 0 & 0 \\ -1.5 & 1.5 & 0 & 0 \end{array}\right)$$

It is clear, that $\tilde{s}_{1a}$ and $\tilde{s}_2$ are not discriminable as they both have the same value of ``5'' for their first index entry. The matrix decomposition did not lead to the identification of any further Principal Components. Hence, it is sensible to repeat the whole process on the remaining data basis taking into account the complete variance of the data and not only the main foci, as it was carried out in the first round of PCA. The remaining data basis is given as

$$\begin{pmatrix} -7.5 & 7.5 & 0 & 0 \\ -1.5 & 1.5 & 0 & 0 \end{pmatrix}$$

with its correlation matrix

$$\begin{pmatrix} 225/2 & 45/2 \\ 45/2 & 9/9 \end{pmatrix}$$

and the corresponding eigenpairs (containing again an eigenvector and its eigenvalue):

$$\nu_1 = 117, u_1 = \begin{pmatrix} 5 \\ 1 \end{pmatrix};; \nu_2 = 0, u_2 = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$$

From the first eigenpair of this second round of PCA the next Principal Component (second in the total process) can be calculated as follows:

$$pc_2 = \begin{pmatrix} 1 & -1 \\ 0 & 2 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ -6 \end{pmatrix} \propto \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$$

The transformed data are now applied to this new basis given by the Principal Components (supplemented with the orthogonal complement) and calculated as follows:

$$\left(\tilde{s}_{1a}\ \tilde{s}_{1b}\ \tilde{s}_2\ \tilde{s}_3\right) = \left(\begin{array}{ccc|c} 1 & 2 & 4 & \\ 1 & 1 & -5 & \\ 1 & -3 & 1 & \end{array}\right) \left(\begin{array}{cccc} 5 & -1 & 5 & -7 \\ -3 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

As in this minimal descriptive example the last row of the matrix only shows entries of ``0", the four substances can be correctly displayed by the two Principal Components, as they already cover 100 % of the original information.