Effect of ommitting an element on the configuration of factors

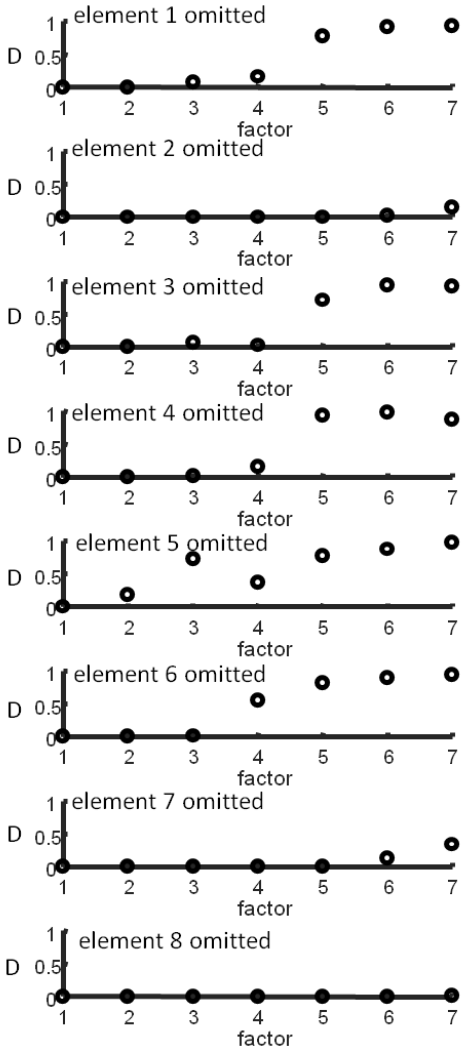
by Bill Menke, December 21, 2017; after a conversation with Blair Goodridge

Consider an $N \times M$ sample matrix $\mathbf{S}$ with N rows representing samples and M columns representing concentrations of elements. We use singular value decomposition to write $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}$ and then interpret the result as a product of factor loadings $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}$ and factors $\mathbf{F} = \mathbf{V}^{\mathrm{T}}$. Each of the M rows of $\mathbf{F}$ is a factor $\mathbf{f}^{(\mathrm{i})}$ and each of the M columns gives the concentration of elements in that factor. The factors are mutually perpendicular and orderd by decreasing singular value $\lambda_{\mathrm{i}} = [\mathbf{\Lambda}]_{\mathrm{ii}}$, so that the lower-order factors capture most of the variance of the samples.

Now suppose that element j is ommitted from the analysis, leading to a "reduced" $(M-1) \times (M-1)$ factor matrix $\mathbf{F}_{\mathrm{r}}$ containing $(M-1)$ factors $\mathbf{f}_{\mathrm{r}}^{(\mathrm{i})}$ each described by $(M-1)$ elements. How much does $\mathbf{f}_{\mathrm{r}}^{(\mathrm{i})}$ differ from an $\mathbf{f}^{(\mathrm{i})}$? Since the $\mathbf{f}_{\mathrm{r}}$'s omit element j, we must project $\mathbf{f}^{(\mathrm{i})}$onto the subspace that omits element j, leadig to a projected vector $\mathbf{f}_{\mathrm{p}}^{(\mathrm{i})}$, before performing the comparison. Let's call Furthermore, we can compare only the first $(M-1)$ factors, since $\mathbf{f}_{\mathrm{r}}^{(\mathrm{M})}$does not exist.

Since the $\mathbf{f}_{\mathrm{r}}$**'s** are unit vectors, a reasonable measure of the deviation between the two factors is $D = 1 - |c|$ with $c = \mathbf{f}_{\mathrm{r}}^{(\mathrm{i})} \cdot \mathbf{f}_{\mathrm{p}}^{(\mathrm{i})} / \left|\mathbf{f}_{\mathrm{p}}^{(\mathrm{i})}\right|$. Here c is the cosine of the angle between the two vectors; the absolute value is introduced into the deviation so that both parallel and anti-parallel vectors have deviation $D = 0$, whereas perpendicular vectors have $D = 1$.

I have applied these ideas to the Atlantic rock dataset provided by Menke and Menke (2016), which gives concentrations for $M = 8$ major oxides. In this dataset, the first 4 factors account for most of the variance of the samples. The results (see plot) inicate that factor 1 is insensitive to the ommission of any element (which is reasonable, since factor 1 is usually near the mean sample and omitting an element does not change values of the other components of a multidimensional mean). Generally speaking, the lower order factors have less deviation than the higher order factors (probably because the latter are more influences by noise in the data, which changes radically when one omits an element). Omitting element 8 leads to no deviation of any factor (which may mean that element 8 is so uncorrelated with any other element that it is its own factor).



Reference. Menke, W. and J. Menke, Environmental Data Analysis with MATLAB, Second Edition (textbook), Academic Press (Elsevier), 342pp, 2016.

Figure. Each graph shows the deviation D of factors caused by removing one element from the sample matrix.