CHAPTER SIX

GRAPHICAL METHODS

6.1 INTRODUCTION

In some circumstances a visual display more easily yields insights than the inspection of tables of parameter estimates. This is often the case with three-mode models as they have many variables and complex relationships between components. In this chapter I present the use of graphical methods in modeling two sets of variables over a third mode. The methods are based on the three-mode models of Chapter Five. These methods offer the user graphical views of the relationships between variables, components and modes.

The chapter is organized as follows: in Section **6.2** I introduce background on biplots and joint plots. Biplots for canonical correlation analysis (CCA) were developed by Ter Braak (1990), who also suggested biplots for redundancy analysis (RA), which I develop in more detail. I also develop biplots for Procrustes rotation (PR). In Section **6.3** I extend these biplots to three-mode data with joint plots. In Section **6.4** I discuss plots of components scores for three-mode methods. Lastly, in Section **6.5** I present the use of residual plots in three-mode modeling. Note that all the programming code that generated the graphs is found in Appendix Four.

6.2 BIPLOTS

The biplot is a technique devised by Gabriel (1971) to represent a matrix approximately by a two-dimensional plot of two sets of two vectors. If the rank of the matrix is two, the representation is exact. A biplot is often, but not necessarily, derived from the two pairs of components corresponding to the largest singular values from the SVD. Eckart and Young (1936) proved that such a two-dimensional approximation to a matrix is optimal in a least squares sense. Generally, any rank-two approximation to a matrix **X** can be decomposed as $\mathbf{AB}'$:

$$\mathbf{X} \cong \mathbf{AB}' = \left[\mathbf{u}_1, \mathbf{u}_2\right]\begin{bmatrix}\lambda_1 & 0 \\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}\mathbf{v}_1' \\ \mathbf{v}_2'\end{bmatrix}.$$

There are three common equivalent ways of displaying this two-dimensional approximation to **X**.

$$\text{1. } \mathbf{A} = \left[\mathbf{u}_1\sqrt{\lambda_1} \quad \mathbf{u}_2\sqrt{\lambda_2}\right] \text{ and } \mathbf{B} = \left[\mathbf{v}_1\sqrt{\lambda_1} \quad \mathbf{v}_2\sqrt{\lambda_2}\right] \tag{6.1}$$

$$\text{2. } \mathbf{A} = \left[\mathbf{u}_1\lambda_1 \quad \mathbf{u}_2\lambda_2\right] \text{ and } \mathbf{B} = \left[\mathbf{v}_1 \quad \mathbf{v}_2\right]$$

$$\text{3. } \mathbf{A} = \left[\mathbf{u}_1 \quad \mathbf{u}_2\right] \text{ and } \mathbf{B} = \left[\mathbf{v}_1\lambda_1 \quad \mathbf{v}_2\lambda_2\right].$$

Which of the three displays one chooses depends on whether one prefers to emphasize the rows or the columns.

If one thinks of the rows as corresponding to subjects and the columns as corresponding to variables, then **A** plots the subjects and **B** the variables. The first axis on the two-dimensional graph represents both the first component for **A** and the first component for **B**, while the second axis represents both the second component of **A** and the second component of **B**. Thus a biplot is two graphs superimposed upon each other: a graph of subject scores and a graph of variable scores. The approximate value of any variable for any subject can be determined by the inner product of the subject and variable vectors. Further, the inner product between vectors of two variables or two subjects is an approximate measure of the closeness or similarity of those variables or subjects. For example, in the third type of biplot the covariance between any two variables is approximated by their inner product on the biplot.

6.2.1 Biplots for Canonical Correlation Analysis

In this section and in the subsequent ones I outline biplots for CCA, RA and PR. These plots have in common that they provide an approximation to $\mathbf{S}_{\mathrm{XY}}$, the matrix of covariances between the X-variables and Y-variables. They will differ in the invariance properties that they invoke. The resulting biplots will be based on vectors whose values are directly related to the variate weights of the associated method.

As a preliminary I shall review the interpretation of $\mathbf{S}_{\mathrm{XY}}$. If the X-variables and Y-variables are standardized to unit length, then $\mathbf{S}_{\mathrm{XY}}[\mathrm{i},\mathrm{j}]$ is the correlation between the i^{th} X-variable, x_{i}, and the j^{th} Y-variable, y_{j}. If only the X-variables are standardized to unit length, then $\mathbf{S}_{\mathrm{XY}}[\mathrm{i},\mathrm{j}]$ is the square root of the regression variance of y_{j} explained by x_{i} in a simple linear regression of y_{j} on x_{i} (by the regression variance I refer to the sums of squares regression divided by the $n-1$, which is also sometimes known as the mean square regression). If neither the X-variables nor the Y-variables are standardized, then $\mathbf{S}_{\mathrm{XY}}[\mathrm{i},\mathrm{j}]$ is the covariance between x_{i} and y_{j}.

In this section I describe biplots of structure coefficients associated with CCA. These biplots will also have the property that they approximate $\mathbf{S}_{\mathrm{XY}}$ with invariance to linear transformations of both the X-variables and the Y-variables. Plots of structure coefficients have been proposed as a means to graphically interpret canonical correlation analysis (Caillez & Pagès, 1976; Israëls, 1987; van der Geer, 1986). Structure coefficients are defined as the correlations between the variables and the canonical variates. Ter Braak (1990) puts plots of structure coefficients in the framework of biplots. He shows their optimality property and discusses their interpretation. The following development of biplots for CCA is from Ter Braak (1990).

Let the matrix $\mathbf{AB}'$ denote a rank-two weighted least squares approximation to $\mathbf{S}_{\mathrm{XY}}$ where the variables are standardized, thus $\mathbf{S}_{\mathrm{XY}}$ is a correlation matrix. The ways to factor $\mathbf{AB}'$ are of the form:

$$\mathbf{A}=\left[\mathbf{S}_{\mathrm{XX}}^{1/2}\mathbf{W}^{*}\mathbf{E}^{\alpha-1}\right]_2 \qquad \mathbf{B}=\left[\mathbf{S}_{\mathrm{YY}}^{1/2}\mathbf{V}^{*}\mathbf{E}^{\alpha}\right]_2, \tag{6.2}$$

where $[\mathbf{Z}]_2$ indicates the first two columns of a matrix $\mathbf{Z}$; $\mathbf{V}^{*}$ is a $p\times r$ orthonormal matrix such that $\mathbf{V}^{*}=\mathbf{S}_{\mathrm{YY}}^{1/2}\mathbf{V}$, where $\mathbf{V}$ is the $p\times r$ matrix of Y-canonical variates; $\mathbf{W}^{*}$ is an $m\times r$ orthonormal matrix such that $\mathbf{W}^{*}=\mathbf{S}_{\mathrm{XX}}^{1/2}\mathbf{W}$, where $\mathbf{W}$ is the $m\times r$ matrix of X-canonical variates; and $m,p\le r$, where r is the rank of $\mathbf{R}_{\mathrm{XY}}$. Here $\mathbf{E}$ is an $r\times r$ diagonal matrix with diagonal entries that are the canonical correlations $\mu_1\ge\mu_2\ge\ldots\ge\mu_r\ge 0$. The most common ways to factor $\mathbf{AB}'$ are to set $\alpha=1$, $1/2$ or 0.

The vectors on the biplot have the following interpretations. $\mathbf{S}_{\mathrm{YY}}^{1/2}\mathbf{V}^{*}$ or equivalently, $\mathbf{S}_{\mathrm{YY}}\mathbf{V}$, is the matrix of structure coefficients (correlations) of the Y-variables with the Y-canonical variates. Similarly, $\mathbf{S}_{\mathrm{XX}}^{1/2}\mathbf{W}^{*}$ or $\mathbf{S}_{\mathrm{XX}}\mathbf{W}$ is the matrix of structure coefficients of the X-variables on the X-canonical variates. If $\alpha=1$, then $\mathbf{B}=\mathbf{S}_{\mathrm{YY}}^{1/2}\mathbf{V}^{*}\mathbf{E}$ is the matrix of correlations between the Y-variables and the X-canonical variates. In the biplot display, the inner product of the vector corresponding to the i^{th} X-variable with the vector corresponding to the j^{th} Y-variable is a rank-two approximation of the $(\mathrm{i},\mathrm{j})^{\mathrm{th}}$ element of $\mathbf{S}_{\mathrm{XY}}$. This is because $\mathbf{S}_{\mathrm{XX}}^{1/2}\mathbf{W}^{*}\mathbf{E}^{1-\alpha}\mathbf{E}^{1}\mathbf{V}^{*}\mathbf{S}_{\mathrm{YY}}^{1/2}=\mathbf{S}_{\mathrm{XX}}^{1/2}\mathbf{S}_{\mathrm{XX}}^{-1/2}\mathbf{S}_{\mathrm{XY}}\mathbf{S}_{\mathrm{YY}}^{-1/2}\mathbf{S}_{\mathrm{YY}}^{1/2}=\mathbf{S}_{\mathrm{XY}}$.

These biplots are optimal in the sense that one obtains an optimal rank-two approximation to $\mathbf{S}_{XY}$ which is invariant to linear transformations of **X** and **Y**. To see this, define the problem as finding **A** and **B**, each of rank-two, such that the expression below is minimized:

$$\left\| \mathbf{S}_{XX}^{-1/2}(\mathbf{S}_{XY} - \mathbf{AB}')\mathbf{S}_{YY}^{-1/2} \right\|^2 , \tag{6.3}$$

where for a matrix **Z**, $\|\mathbf{Z}\|^2 = \text{trace}(\mathbf{Z}'\mathbf{Z})$. The solution to (6.3) is provided by the singular value decomposition of $\mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$,

$$\mathbf{W}^*\mathbf{E}\mathbf{V}^{*\prime} = \mathbf{S}_{XX}^{-1/2}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2} ,$$

where **A** and **B** are defined as in (6.2).

Ter Braak (1990) also gives an alternative version of the biplot for CCA. Instead of plotting **A** and **B** of (6.2), one plots

$$\mathbf{A} = \left[\mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{E}^{\alpha-1}\right]_2 \quad \mathbf{B} = \left[\mathbf{S}_e^{1/2}\mathbf{V}^*\mathbf{E}^{\alpha}\right]_2 \tag{6.4}$$

where $\mathbf{S}_e$ is the residual sum of squares when the products matrix of **Y** with respect to **X** is subtracted from the total variation for **Y**. That is:

$$\mathbf{S}_e = \mathbf{S}_{YY} - \mathbf{S}_{XY}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY} . \tag{6.5}$$

Ter Braak justifies this biplot in terms of approximating the matrix of regression coefficients in a multivariate regression of **Y** on **X**. However, it can be justified in the same way that (6.2) is justified, by finding an optimal approximation to $\mathbf{S}_{XY}$ that is invariant to non-singular transformations of the X-variables and the Y-variables. That is, one finds **A** and **B** such that

$$\left\| \mathbf{S}_{XX}^{-1/2\,\prime}(\mathbf{S}_{XY} - \mathbf{AB}')\mathbf{S}_e^{-1/2} \right\|^2 , \tag{6.6}$$

is minimized.

When the X-variables are thought to cause the Y-variables in a regression sense then $\mathbf{S}_e$ is a more natural choice than $\mathbf{S}_{YY}$. For example, if the X-variables are group indicators then $\mathbf{S}_e$ becomes the within-groups covariance matrix. The biplot then indicates which groups are well discriminated, and which responses contribute to the discrimination.

6.2.2 Biplots for Redundancy Analysis

For an analogous biplot for redundancy analysis Ter Braak suggests finding **A** and **B** such that one has an optimal approximation to $\mathbf{S}_{XY}$ which is invariant only to non-singular transformations of the X-variables. This definition leads to **A** and **B** that are functions of the redundancy variables. Find **A** and **B** such that

$$\left\| \mathbf{S}_{XX}^{-1/2}(\mathbf{S}_{XY} - \mathbf{AB}') \right\|^2$$

is minimized. Which leads to

$$\mathbf{A} = \left[\mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{E}^{\alpha-1}\right]_2 \qquad \mathbf{B} = \left[\mathbf{V}\mathbf{E}^{\alpha}\right]_2 ,$$

where $\mathbf{W}^*$ is an $m \times r$ orthonormal matrix such that $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}\mathbf{W}$, where **W** is the $m \times r$ matrix of canonical coefficients for the X-variables; **V** is a $p \times r$ orthonormal matrix of redundancy variates; and **E** is an $r \times r$ diagonal matrix whose j^{th} element is $(n-1)^{-1}$ times the root of the variation of the j^{th} Y-variate explained by the j^{th} X-variate, where *n* is the sample size.

The matrices of biplot vectors, **A** and **B**, are functions of the redundancy weights. However, Ter Braak's interpretation of these weights is incorrect. He states (1990) correctly that with $\alpha = 1$ **A** is the matrix of structure coefficients. However, it is not correct that "The elements of **B** are not only correlations but also canonical coefficients". The correct interpretation is that $\mathbf{B} = [\mathbf{VE}]_2$ is a matrix whose $(i, j)^{th}$ element is $(n-1)^{-1}$ times the root of the variation of y_i explained by a simple linear regression on the j^{th} redundancy variate, $\mathbf{w}_j'\mathbf{x}$. To see this, notice that if one regresses y_i on $\mathbf{w}_j'\mathbf{x}$, the regression variance is $\mathbf{y}_i'\mathbf{X}\mathbf{w}_j(\mathbf{w}_j'\mathbf{X}'\mathbf{X}\mathbf{w}_j)^{-1}\mathbf{w}_j'\mathbf{X}'\mathbf{y}_i = \left(\mathbf{y}_i'\mathbf{X}\mathbf{w}_j\right)^2$, since $\mathbf{w}_j'\mathbf{X}'\mathbf{X}\mathbf{w}_j = 1$. Hence $\mathbf{y}_i'\mathbf{X}\mathbf{w}_j$ is the root of the regression variance. The matrix of root variances of the y_i regressed against the $\mathbf{w}_j'\mathbf{x}$ is $\mathbf{Y}'\mathbf{X}\mathbf{W}$, and $\mathbf{Y}'\mathbf{X}\mathbf{W} = (n-1)\mathbf{VEW}'\mathbf{W} = (n-1)\mathbf{VE}$.

An alternative factorization is worth mentioning: $\mathbf{A} = [\mathbf{S}_{XX}^{1/2}\mathbf{WE}]_2$ and $\mathbf{B} = [\mathbf{V}]_2$. Here **B** is just the matrix of weights of the redundancy variates for the Y-variables. **A** is a matrix whose elements are $(n-1)^{-1}$ times the root of the variation explained by each X-variable of each Y-variate.

6.2.3 Biplots for Procrustes Rotation

Biplots for PR which are analogous to those for CCA and RA can also be developed. One wishes to find **A** and **B** that minimize $\|(\mathbf{S}_{XY} - \mathbf{AB}')\|^2$. This method yields a biplot approximation to $\mathbf{S}_{XY}$ that is optimal in a least squares sense. Consistent with PR, this plot is invariant neither to non-singular transformations of the X-variables nor the Y-variables. As with the biplots for CCA and RA, the matrices of vectors **A** and **B** are interpretable in terms of a PR analysis. $\mathbf{A} = [\mathbf{W}]_2$ is a matrix of coefficients for the X-variates. $\mathbf{B} = [\mathbf{VE}]_2$ is a matrix showing the covariance of each Y-variable with the Procrustes rotation variates, as $\mathbf{W}'\mathbf{S}_{XY}\mathbf{I} = \mathbf{W}'\mathbf{WEV} = \mathbf{EV}$.

The biplot for PR can be interpreted as showing which X-variables covary with which Y-variables. It also can be interpreted as showing which variables are fit well in the Procrustes rotation. That is, it shows which X-variables match the pattern of the Y-variables when rotated.

In summary, there are biplots for displaying CCA, RA and PR that are optimal in a weighted least squares sense and which yield markers related to the estimated model parameters.

6.3 JOINT PLOTS

Chapter Five extended the CCA, RA and PR models to three-mode data, that is, multiple occasions and multiple datasets. Analogously, this section extends biplots for CCA, RA and PR to three-mode data. The biplots of Section **6.2** were based on singular value decompositions of a matrix. Joint plots (Kroonenberg 1983) are based on the Tucker2 or PARAFAC (orth.) decomposition of a three-mode array.

Denote an $m \times p \times g$ three-mode array as $\underline{\mathbf{C}}$. To examine the relationship between the component weights of two modes one can make joint plots. One possibility is to create a series of joint plots, one joint plot for each component of the non-examined mode; i.e., one for each occasion. Another possibility is to make one averaged joint plot. Each joint plot is analogous to a biplot. If one wants to examine the relationship between the subject and variable modes, then what is plotted is based on either $\mathbf{G}\mathbf{C}_{\mathrm{k}}\mathbf{H}'$, $\mathrm{k} = 1,\ldots,g$, or $\mathbf{G}\overline{\mathbf{C}}\mathbf{H}'$, where $\mathbf{G}$ is the matrix of components for the subject mode, $\mathbf{H}$ is the matrix of components for the variables mode, $\mathbf{C}_{\mathrm{k}}$ is the k^{th} slice of the core box, that is the matrix $\underline{\mathbf{C}}[,,\mathrm{k}]$, and $\overline{\mathbf{C}}$ is the average of the $\mathbf{C}_{\mathrm{k}}$. As $\mathbf{C}_{\mathrm{k}}$ and $\overline{\mathbf{C}}$ are generally not diagonal, a SVD is performed on $\mathbf{C}_{\mathrm{k}}$ to factor it. So what is plotted is $\mathbf{A}_{\mathrm{k}}$ and $\mathbf{B}_{\mathrm{k}}$ where:

$$\mathbf{A}_{\mathrm{k}}\mathbf{B}_{\mathrm{k}}' = \mathbf{G}\mathbf{C}_{\mathrm{k}}\mathbf{H}' = \mathbf{G}(\mathbf{U}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}\mathbf{V}_{\mathrm{k}}')\mathbf{H}' = \left(\tfrac{m}{p}\right)^{1/4}(\mathbf{G}\mathbf{U}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{\alpha})(\boldsymbol{\Lambda}_{\mathrm{k}}^{1-\alpha}\mathbf{V}_{\mathrm{k}}'\mathbf{H}')\left(\tfrac{p}{m}\right)^{1/4},$$

or $\mathbf{A}_{\mathrm{k}} = \left(\tfrac{m}{p}\right)^{1/4}(\mathbf{G}\mathbf{U}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{\alpha})$ and $\mathbf{B}_{\mathrm{k}} = (\boldsymbol{\Lambda}_{\mathrm{k}}^{1-\alpha}\mathbf{V}_{\mathrm{k}}'\mathbf{H}')\left(\tfrac{p}{m}\right)^{1/4}$, where m is the number of measurements in the subjects mode and p the number measurements in the variables mode, $\mathbf{G}$ and $\mathbf{H}$ are defined to be the two specific pairs of components, thus $\mathbf{G}$ is an $m \times 2$ matrix and $\mathbf{H}$ is a $p \times 2$ matrix; and $\mathbf{C}_{\mathrm{k}}$ and $\overline{\mathbf{C}}$ are the 2×2 submatrices of the core matrices corresponding to the selected components. One typically chooses the two components for $\mathbf{G}$ and $\mathbf{H}$ with the largest associated core elements. However, one can make a joint plot between any two components.

Further details: α is chosen typically to be 1/2. The weightings $\left(\tfrac{m}{p}\right)^{1/4}$ and $\left(\tfrac{p}{m}\right)^{1/4}$ put the distances from the origin of the subject vectors on the same scale as those of the variable vectors (Kroonenberg 1983). This makes the plots easier to view, though one may choose other weightings (see Section **6.2**, in particular (6.1), for a discussion on the weighting of biplots).

6.3.1 Joint Plots for Canonical Correlation Analysis

In this section I extend the biplots of structure coefficients for CCA at one occasion to joint plots of structure coefficients for multiple occasions. Here one models $\mathbf{S}_{\mathrm{XY}}$ over the third mode. The assumptions necessary for this particular biplot are the same as for CCA/third, that $\mathbf{S}_{\mathrm{XX}}$ is constant over the third mode, and that either $\mathbf{S}_{\mathrm{YY}}$ or $\mathbf{S}_{\mathrm{e}}$ (6.5) is also constant over the third mode (see Section **5.5.3** for a discussion of these assumptions). The subsequent developments apply to both $\mathbf{S}_{\mathrm{e}}$ and $\mathbf{S}_{\mathrm{YY}}$, although I use $\mathbf{S}_{\mathrm{YY}}$.

Plot $\mathbf{A}_{\mathrm{k}}$ and $\mathbf{B}_{\mathrm{k}}$ for each occasion k, where $\mathbf{A}_{\mathrm{k}}$ and $\mathbf{B}_{\mathrm{k}}$:

$$\mathbf{A}_k = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}_k\boldsymbol{\Lambda}_k^{1/2}, \quad \mathbf{B}_k = \mathbf{S}_{YY}^{1/2}\mathbf{V}^*\mathbf{Z}_k\boldsymbol{\Lambda}_k^{1/2},$$

where $\mathbf{V}^*$ is a $p \times 2$ orthonormal matrix such that $\mathbf{V}^* = \mathbf{S}_{YY}^{1/2}[\mathbf{V}]_2$, where $\mathbf{V}$ is the $p \times r$ matrix of Y-canonical variates from the CCA/third model; $\mathbf{W}^*$ is a $m \times 2$ orthonormal matrix such that $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}[\mathbf{W}]_2$, where $\mathbf{W}$ is the $m \times q$ matrix of X-canonical variates from the CCA/third model; and $\mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{Z}'_k = \mathbf{C}_k$ is the singular value decomposition of $\mathbf{C}_k$, the k^{th} core matrix from the CCA/third model. An alternative is to plot $\mathbf{A} = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}\boldsymbol{\Lambda}^{1/2}$ and $\mathbf{B} = \mathbf{S}_{YY}^{1/2}\mathbf{V}^*\mathbf{Z}\boldsymbol{\Lambda}^{1/2}$, where $\mathbf{U}\boldsymbol{\Lambda}\mathbf{Z}' = \overline{\mathbf{C}}$.

Note that the $\mathbf{S}_{XX}^{1/2}\mathbf{W}^*$ and the $\mathbf{S}_{YY}^{1/2}\mathbf{V}^*$ are matrices of structure coefficients.

Next I discuss the sense in which these joint plots are optimal. The function one minimizes is the least squares lack of fit of a rank-two approximation to the matrix of covariances between the X-variables and the Y-variables, $\frac{1}{n_k - 1}\mathbf{X}'_k\mathbf{Y}_k$, $k = 1, \ldots, g$. The loss function should be invariant to non-singular transformations of the X-variables and the Y-variables, as is CCA. Further, $\mathbf{A}_k$ should span the same space for $k = 1, \ldots, g$, as should $\mathbf{B}_k$. Thus the problem reduces to finding $\mathbf{A}_k$ and $\mathbf{B}_k$ of rank-two such that

$$\sum_{k=1}^{g}\left\| \mathbf{S}_{XX}^{-1/2}\left(\frac{1}{n_k - 1}\mathbf{X}'_k\mathbf{Y}_k - \mathbf{A}_k\mathbf{B}'_k\right)\mathbf{S}_{YY}^{-1/2}\right\|^2$$

is minimized. Clearly the optimal solution is given by a two-component Tucker2 decomposition of $\frac{1}{n_k - 1}\mathbf{S}_{XX}^{-1/2}\mathbf{X}'_k\mathbf{Y}_k\mathbf{S}_{YY}^{-1/2}$, $k = 1, \ldots, g$. That is

$$\mathbf{W}^*\mathbf{C}_k\mathbf{V}^{*\prime} = \frac{1}{n_k - 1}\mathbf{S}_{XX}^{-1/2}\mathbf{X}'_k\mathbf{Y}_k\mathbf{S}_{YY}^{-1/2},$$

for $k = 1, \ldots, g$. Thus $\mathbf{W}^*\mathbf{C}_k\mathbf{V}^{*\prime} = \mathbf{S}_{XX}^{-1/2}\mathbf{A}_k\mathbf{B}'_k\mathbf{S}_{YY}^{-1/2}$ and

$$\mathbf{A}_k = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}_k\boldsymbol{\Lambda}_k^{\alpha}, \quad \mathbf{B}_k = \mathbf{S}_{YY}^{1/2}\mathbf{V}^*\mathbf{Z}_k\boldsymbol{\Lambda}_k^{1-\alpha}, \tag{6.7}$$

where one performs a singular value decomposition on $\mathbf{C}_k$ to get $\mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{Z}'_k = \mathbf{C}_k$.

If one restricts the columns of $\mathbf{A}_k$ to be proportional over $k = 1, \ldots, g$, i.e., $\mathbf{A}_c[, i] = f\mathbf{A}_d[, i]$, for $c \neq d$, and likewise makes the same restriction for $\mathbf{B}_k$, then one can use an argument similar to the one above to show that the optimal joint plots are based on the PARAFAC (orth.) solution. Such joint plots are easier to interpret as the axes in the joint plots correspond to the two pairs of components. This is so because the core matrices are diagonal, and thus the matrices of structure coefficients are not rotated (by $\mathbf{U}_k$ or $\mathbf{Z}_k$ in (6.7)).

6.3.2 Joint Plots for Redundancy Analysis

The biplot for RA, which plotted structure coefficients and redundancy variates, can be extended to joint plots for multimode data in a manner analogous to how the biplot for CCA was extended to joint plots for CCA. The multimode extension of the biplot for RA is

$$\mathbf{A}_k = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}_k\mathbf{\Lambda}_k^{1/2}, \quad \mathbf{B}_k = \mathbf{V}^*\mathbf{Z}_k\mathbf{\Lambda}_k^{1/2}, \tag{6.8}$$

where $\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{Z}_k = \mathbf{C}_k$ is the singular value decomposition of the kth core matrix $\mathbf{C}_k$. $\mathbf{W}^*$ is an $m \times 2$ orthonormal matrix such that, $\mathbf{W}^* = \mathbf{S}_{XX}^{1/2}[\mathbf{W}]_2$, where $\mathbf{W}$ is the $m \times q$ matrix of canonical coefficients for the X-variables; and $\mathbf{V}^* = [\mathbf{V}]_2$, where $\mathbf{V}$ is a $p \times r$ orthonormal matrix of redundancy variates. One can also plot $\mathbf{A} = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}\mathbf{\Lambda}^{1/2}$ and $\mathbf{B} = \mathbf{V}^*\mathbf{Z}\mathbf{\Lambda}^{1/2}$, where $\mathbf{U}\mathbf{\Lambda}\mathbf{Z}' = \overline{\mathbf{C}}$.

Notice that $\mathbf{S}_{XX}^{1/2}\mathbf{W}^*$ is the matrix of structure coefficients and $\mathbf{V}$ is the matrix of redundancy coefficients.

These biplots are optimal in that they yield a display approximating the matrices $\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k$, $k = 1,\dots,g$, based on a loss function that is invariant to non-singular linear transformations of the X-variables, as is RA. Further, $\mathbf{A}_k$ should span the same space for $k = 1,\dots,g$, as should $\mathbf{B}_k$. The problem reduces to finding $\mathbf{A}_k$ and $\mathbf{B}_k$ of rank-two such that

$$\sum_{k=1}^{g}\left\|\mathbf{S}_{XX}^{-1/2}\left(\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k - \mathbf{A}_k\mathbf{B}_k'\right)\right\|^2$$

is minimized. Clearly the optimal solution is given by a two-component Tucker2 decomposition of $\frac{1}{n_k - 1}\mathbf{S}_{XX}^{-1/2}\mathbf{X}_k'\mathbf{Y}_k$, $k = 1,\dots,g$. That is $\mathbf{W}^*\mathbf{C}_k\mathbf{V}^{*\prime} = \frac{1}{n_k - 1}\mathbf{S}_{XX}^{-1/2}\mathbf{X}_k'\mathbf{Y}_k$. Thus $\mathbf{W}^*\mathbf{C}_k\mathbf{V}^{*\prime} = \mathbf{S}_{XX}^{-1/2}\mathbf{A}_k'\mathbf{B}_k$ and $\mathbf{A}_k = \mathbf{S}_{XX}^{1/2}\mathbf{W}^*\mathbf{U}_k\mathbf{\Lambda}_k^{\alpha}$, $\mathbf{B}_k = \mathbf{V}^*\mathbf{Z}_k\mathbf{\Lambda}_k^{1-\alpha}$, where $\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{Z}_k' = \mathbf{C}_k$.

If one restricts the columns of $\mathbf{A}_k$ to be proportional over $k = 1,\dots,g$, i.e. $\mathbf{A}_c[,i] = f\mathbf{A}_d[,i]$, for $c \neq d$, and likewise makes the same restriction for $\mathbf{B}_k$, then one can use an argument similar to the one above to show that the optimal joint plots are based on the PARAFAC (orth.) solution. Such joint plots are easier to interpret as the axes in the joint plots correspond to the two pairs of components. This is so because the core matrices are diagonal, and thus the matrices of structure coefficients are not rotated (by $\mathbf{U}_k$ or $\mathbf{Z}_k$ in (6.8)).

Example 6.1

Figure 6.1 shows the joint plot for the Shenendoah data from Section **5.4**. The plot is based on the RA/third solution for the first two components of the PARAFAC (orth.) model. Since for this example the plots are roughly similar across time, instead of showing the joint plots for each occasion, the plot based on the sum of the core matrices is shown. This gives a general picture of how the two modes, geological variables and streamwater variables, relate.

The vectors corresponding to the nine geological variables are numbered one through nine. Recall that the first five geological variables refer to bedrock types, ab2400 refers to altitude, DD refers to drainage density, E/W refers to east or west slope, and Dev. refers to the presence of development. The vectors corresponding to the fourteen streamwater variables are lettered from a to p, skipping l and o; they are more self-explanatory. The origin is labeled with a zero. The key to the numbering and lettering is given in **Figure 6.2**.

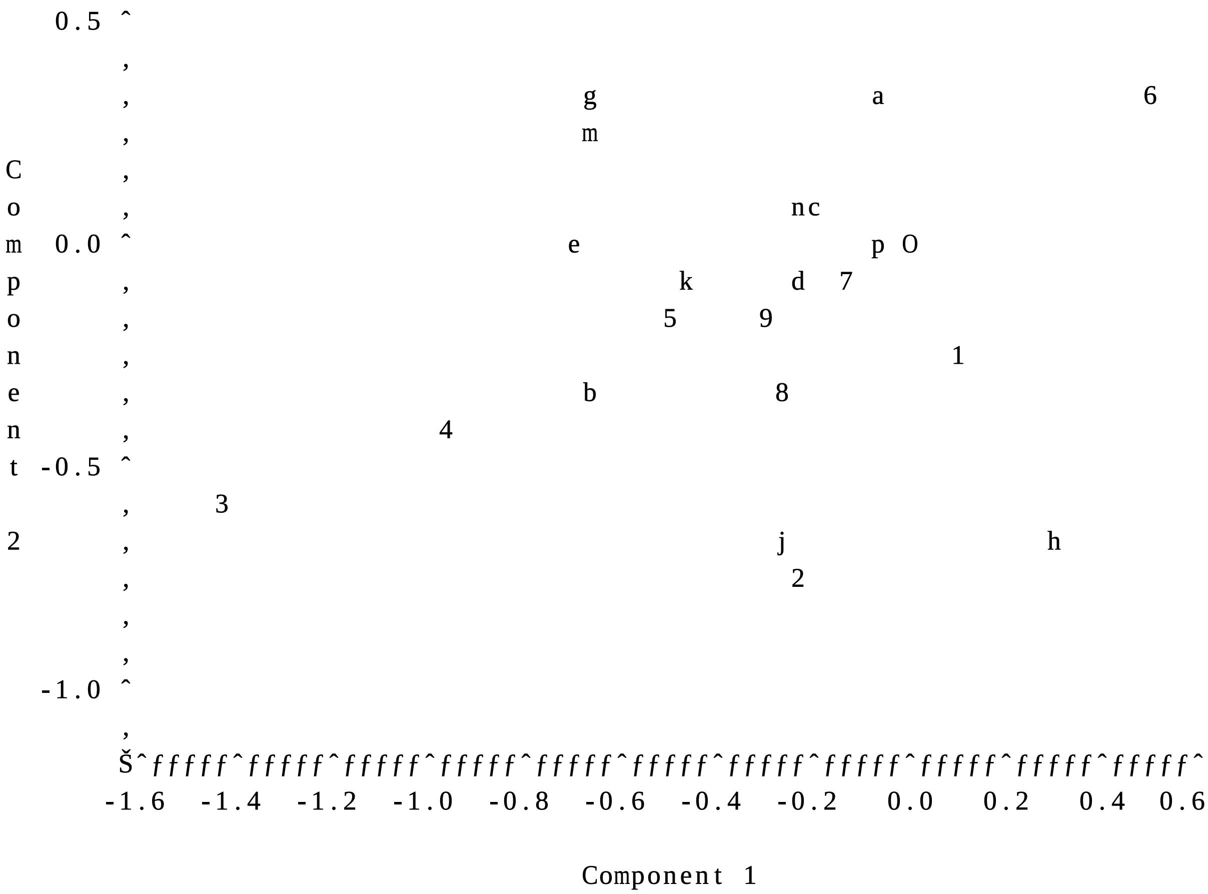

Figure 6.1 Joint Plot for the Sum of Core Matrices for PARAFAC (orth.)

a	discharge	h	K^{+}	1	Antietam
b	conductivity	i	alkalinity	2	Hampton
c	pH	j	$SO_4^{=}$	3	Catoctin
d	temperature	k	Cl^{-}	4	Pedlar
e	Ca^{++}	m	$SiO_4^{=}$	5	Old Rag
f	Mg^{++}	n	NO_3^{-}	6	ab2400
g	Na^{+}	p	NH_4^{+}	7	DD
				8	E/W
				9	Dev.

Figure 6.2 Key to Symbols

The interpretation of the joint plot is aided by consideration of what the components are and how the vectors relate. The first axis corresponds to the first geological and streamwater variables component pair, which relates to carbonic acid weathering. The second axis corresponds to the second component pair, which relates to acidification. Next, consider that since the geological variables are standardized, the $(\mathrm{i},\mathrm{j})^{\mathrm{th}}$ element of $\mathbf{S}_{\mathrm{XY}}$ for a given occasion shows $(n-1)^{-1}$ times the root of the variance of the j^{th} Y-variable explained in a simple linear regression on the i^{th} X-variable. Consequently, the inner product between a geological variable's vector and a streamwater variable's vector yields an approximation to $(n-1)^{-1}$ times the root of the regression variance of the streamwater variable predicted by the geological variable. This is in an averaged sense since the joint plot is based on an averaged core matrix. Thus if a geological and streamwater variable are near each other on the plot, such as Hampton bedrock (2) and $\mathrm{SO}_4^=$ (j), then the geological variable is a strong predictor of the streamwater variable in a simple linear regression.

The advantage of the joint plot is that it lets one view all of the variables in relation to each other. Geological variables which are located near one another are similar in the variation of the streamwater variables they explain. For example, Catoctin (3) and Pedlar (4) bedrock types explain the streamwater variables similarly, though Catoctin (3) has a stronger effect since it is further from the origin. Likewise, streamwater variables which are near to each other are similar in that they are explained by the same geological variables. For example, the sodium (g), silica (m) and calcium (e) ions are near each other. They result from the same processes, and thus are predicted by the same geological variates. Important also is the distance from the origin. Geological variables that are near the origin, such as drainage density (7), are not strong predictors in the rank-two RA/time model upon which the joint plot is based. Similarly streamwater variables near the origin, such as ammonium (p), are not well explained by the rank-two RA/time model.

A limitation of these joint plots is that they only display a two-components solution. Thus the effects of possible lower order components are not seen. For example, ammonium and drainage density are both related to the fourth pair of components in the analysis in Section **5.4**.

6.3.3 Joint Plots for Procrustes Rotation

The biplot for PR can be extended to joint plots for multimode data in a manner analogous to how the biplots for CCA and RA were extended to joint plots. The multimode extension of the biplot for PR is

$$\mathbf{A}_{\mathrm{k}} = \mathbf{W}\mathbf{U}_{\mathrm{k}}\mathbf{\Lambda}_{\mathrm{k}}^{1/2}, \quad \mathbf{B}_{\mathrm{k}} = \mathbf{V}\mathbf{Z}_{\mathrm{k}}\mathbf{\Lambda}_{\mathrm{k}}^{1/2}, \tag{6.9}$$

where $\mathbf{W}^* = [\mathbf{W}]_2$, where $\mathbf{W}$ is the $m \times q$ matrix of PR coefficients for the X-variables; $\mathbf{V}^* = [\mathbf{V}]_2$, where $\mathbf{V}$ is a $p \times r$ orthonormal matrix of PR variates; and where $\mathbf{U}_{\mathrm{k}}\mathbf{\Lambda}_{\mathrm{k}}\mathbf{Z}_{\mathrm{k}} = \mathbf{C}_{\mathrm{k}}$ is the singular value decomposition of the k^{th} core matrix $\mathbf{C}_{\mathrm{k}}$. One can also plot $\mathbf{A} = \mathbf{W}\mathbf{U}\mathbf{\Lambda}^{1/2}$ and $\mathbf{B} = \mathbf{V}\mathbf{Z}\mathbf{\Lambda}^{1/2}$, where $\mathbf{U}\mathbf{\Lambda}\mathbf{Z}' = \overline{\mathbf{C}}$.

Note that **W** is the $m \times r$ matrix of PR coefficients for the X-variables and **V** is the $p \times r$ matrix of PR coefficients for the Y-variables.

These biplots are optimal in that they yield a display approximating the matrix $\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k$. Consistent with PR, the solution is not invariant linear transformations of the variables. $\mathbf{A}_k$ should span the same space, as should the $\mathbf{B}_k$. The problem reduces to finding $\mathbf{A}_k$ and $\mathbf{B}_k$ of rank-two such that

$$\sum_{k=1}^{g}\left\|\left(\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k - \mathbf{A}_k\mathbf{B}_k'\right)\right\|^2$$

is minimized. Clearly the optimal solution is given by a Tucker2 decomposition of $\frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k$, $k = 1,\ldots,g$. That is $\mathbf{W}\mathbf{C}_k\mathbf{V}' = \frac{1}{n_k - 1}\mathbf{X}_k'\mathbf{Y}_k$. Thus $\mathbf{W}\mathbf{C}_k\mathbf{V}' = \mathbf{A}_k'\mathbf{B}_k$ and $\mathbf{A}_k = \mathbf{W}\mathbf{U}_k\mathbf{\Lambda}_k^{\alpha}$, $\mathbf{B}_k = \mathbf{V}\mathbf{Z}_k\mathbf{\Lambda}_k^{1-\alpha}$, where $\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{Z}_k' = \mathbf{C}_k$.

If one restricts the columns of $\mathbf{A}_k$ to be proportional over $k = 1,\ldots,g$, i.e. $\mathbf{A}_c[,i] = f\mathbf{A}_d[,i]$, for $c \neq d$, and likewise makes the same restriction for $\mathbf{B}_k$, then one can use an argument similar to the one above to show that the optimal joint plots are based on the PARAFAC (orth.) solution. Such joint plots are easier to interpret as the axes in the joint plots correspond to the two pairs of components. This is so because the core matrices are diagonal, and thus the matrices of structure coefficients are not rotated (by $\mathbf{U}_k$ or $\mathbf{Z}_k$ in (6.9)).

6.4 PLOTS OF THE COMPONENT SCORES

The score on a component for a given subject is generally defined as a weighted sum of the variables, the weights being component weights. Define the $m \times 1$ vector of scores for subjects on the b^{th} variable component at the k^{th} occasion, denoted by $\mathbf{Q}_k[,b]$, as

$$\mathbf{Q}_k[,b] = \mathbf{D}_k\mathbf{H}[,b].$$

An equivalent form is $\mathbf{Q}_k[,b] = \mathbf{G}\mathbf{C}_k[,b]$. One could also define scores for the variables on the subject components.

In some applications it may be useful to inspect the scores of all combinations of the elements of two modes on the components of the third mode. For instance, for longitudinal data the scores of each subject-time combination on the variable components can be used to inspect the development of a subject's score on the variable components over time (Kroonenberg 1983). Component scores serve as an intermediate level of condensation between the raw data and the three-mode model.

Example 6.2

In **Figures 6.3**, **6.4**, **6.5** and **6.6** I present the scores of the geological variables at the six occasions on each of the four streamwater components. These scores are based on the Tucker2 solution with four geological and four streamwater variates, as given in Section **5.4**.

The scores give a picture of how the geological variables relate to the streamwater components. I start by making two general observations on the plots. First, one notices that in the plots for the first three components (**Figures 6.3**, **6.4** and **6.5**) is that the relative positions of the geological variables are steady, even roughly proportional, over time. This is a reflection of the fact that the off-diagonal elements of the core matrices are near zero. Indeed, if one plotted scores based on the PARAFAC (orth.) model the positions of the scores would be exactly proportional over time. Only for the fourth variate does this not hold true. If one looks at the core matrix in **Table 5.3** for the March 1982 one sees a large off-diagonal element (1.25) between the fourth streamwater variate and first geological variate. Second, note that the range of the scores narrows for the subsequent components. This is because each component accounts for less of the variation than the previous.

I will just point out a few notable details about the score plots to provide a sense for how one interprets them. First, recall from Section **5.4** that the first streamwater component is interpreted to be related to the results of the weathering due to carbonic acid, with heavy weightings for, silica, alkalinity and the alkaline ions. **Figure 6.3** shows where the geological variables measure on this component over time. For example one sees that altitude (6) has a generally low score, but is particularly low at the first occasion, August 1981. Also, note in **Figure 6.5** that drainage density (7) has a high score on the third component at the fourth occasion (March 1982). This component was related in Section **5.4** to high silica and sodium, the results of plagioclastic weathering. Drainage density explains a relatively good deal of the variation of this component and of the associated streamwater variables at this occasion.

The key to the numbering is shown in Figure **6.2**. Note that some numbers may be obscured by others.

```
         ,
     1.0 ^             3
         , 3
         ,
         ,                                                 3
         , 8           9           3           3                       3
         , 9           4
     0.5 ^ 4                                               9
         , 7           8           8                       4           9
         ,                         4                       7           4
         ,                                     4           8           8
         ,                         7
         ,                                     5
     0.0 ^                         5                       5           5
s        ,
c        , 5           5                       7
o        ,
r        ,
e        ,
s   -0.5 ^
         , 2                                   6
         ,
         ,                         2
         ,                                                             6
         , 1                       1           2           6
    -1.0 ^             2           6           1
         ,
         ,                                                             1
         ,             1                                   1
         ,
         ,
    -1.5 ^ 6
         ,
         Šf^fffffffffff^fffffffffff^fffffffffff^fffffffffff^fffffffffff^ff
         Aug. 81   Sept. 81     Jan. 82     Mar. 82      May 82    June 82

                                       occasion
```

Figure 6.3 Scores on the First Streamwater Variate

```
          ,
     1.0  ˆ
          ,
          ,
          ,                                                   2
          ,  2         2         2         7         2
          ,            2
     0.5  ˆ                                2
          ,                                                   1
          ,  1         3         1         9         1        8
s         ,  3         1         3         4         3
c         ,  4         4         4         1         4        4
o         ,  7                   7         3                  3
r    0.0  ˆ            7
e         ,
s         ,  5         5         5         5         5        5
          ,
          ,  6                   6
          ,            6                             6
    -0.5  ˆ                                                   6
          ,
          ,                                6
          ,
          ,
          ,
    -1.0  ˆ
          Šƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒ
           Aug. 81   Sept. 81    Jan. 82   Mar. 82    May 82   June 82

                                  occasion
```

Figure 6.4 Scores on the Second Streamwater Variate

```
              ,
      1.0     ˆ
              ,                                   7
              ,
              ,
              ,
              ,
      0.5     ˆ
              ,
              ,
              ,  5          5
s             ,                      5                         5           5
c             ,  7                               4             7           8
o     0.0     ˆ  1          1        7           5
r             ,             2        1           1             1           1
e             ,  2          4        2           8             2           9
s             ,  9          9        6           2
              ,  6                   3                         6
              ,             6                    6             3           3
     -0.5     ˆ                                  3
              ,
              ,  3
              ,             3
              ,
              ,
     -1.0     ˆ
              ,
              Šƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆƒƒ
              Aug. 81     Sept. 81     Jan. 82     Mar. 82      May 82     June 82

                                          occasion
```

Figure 6.5 Scores on the Third Streamwater Variate

```
      1.0 ˆ
          ,
          ,
          ,
          ,
          ,                                  2
      0.5 ˆ                                  1
          ,
          ,                                  5
          ,           7
          , 7                    1           6          7          7
          , 8         8          7                      8          2
      0.0 ˆ 2         2          5                      2          1
s         , 1         1          4                      1
c         ,           3          6                      3          3
o         ,                      3
r         , 6         6                                 6          6
e         ,                                  4
s    -0.5 ˆ
          ,                                  3
          ,
          ,
          ,
          ,
     -1.0 ˆ
          ,                                  7
          ,
          ,
          ,
          ,
     -1.5 ˆ
          Šƒˆƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒˆƒƒ
          Aug. 81   Sept. 81    Jan. 82    Mar. 82     May 82    June 82

                                    occasion
```

Figure 6.6 Scores on the Fourth Streamwater Variate

6.5 RESIDUAL PLOTS

Kroonenberg (1983) recommends sums of squares plots to assess the quality of fit of the elements of a mode. For each variable (or subject) one plots its sums of squares fit (sums of squares explained) against its sums of squares residuals (sums of squares lack of fit).

Example 6.3

Below in **Figure 6.7** is the residual plot for the Shenendoah data based on the estimates for the four component PARAFAC (orth.) model from Section **5.5**. Plotted are the sums of squares residuals versus sums of squares fit for the fourteen variables. Note that the total sums of squares for each variable is the total of the sums of squares explained by a separate multiple regression at each occasion.

In this plot one sees the relationship variance of variables temperature (d), discharge (a) and nitrate (n) are modeled relatively poorly by the three-mode model. If one recalls the analysis of Section **5.5**, these were variables that did not participate in any of the processes attributed to the four variate pairs. On the other hand, some streamwater variables are modeled well by the three-mode models, such as alkalinity.

The key relating the numbers to the variables is given in **Figure 6.8**.

```
L       ‚
a       ‚
c  2.0 ˆ
k       ‚
        ‚          d
o  1.5 ˆ          a
f       ‚
        ‚        n
f  1.0 ˆ
i       ‚                             j
t       ‚                        k
   0.5 ˆ            c                          h            e bi    f
        ‚          p                                                g
        ‚
   0.0 ˆ
        ‚
        Šffffˆfffffˆfffffˆfffffˆfffffˆfffffˆfffffˆfffffˆfffffˆfffffˆfffffˆffff
            0.5   1.0   1.5   2.0   2.5   3.0   3.5   4.0   4.5   5.0

                         Sums of Squares Explained
```

Figure 6.7 Residual Plot for the Sums of Squares Explainable of the Streamwater Variables

a	discharge	h	K^+
b	conductivity	i	alkalinity
c	pH	j	$SO_4^=$
d	temperature	k	Cl^-
e	Ca^{++}	m	$SiO_4^=$
f	Mg^{++}	n	NO_3^-
g	Na^+	p	NH_4^+

Figure 6.8 Key to Symbols

6.6 SUMMARY

In this chapter I showed how to use graphical displays to aid in the analysis of the relationship between two sets of variables over time. These displays were related to the CCA, RA, or PR parameters of the three-mode models of Chapter Six. The joint plots showed how the variables from the two sets related to each other. The component plots were useful for showing the interactions between variables and components over time. In all, these plots allow the researcher to get a quick, visual appreciation of the relationships between many variables at different occasions.