Appendix F Linear discriminant analysis on S-wave amplitudes

In Chapter 4, we applied linear discriminant analysis to the problem of finding the ratio of ground motions that would be optimal indicators of magnitude. We applied the method on P-wave amplitudes. In this Appendix, we apply the same method on S-wave amplitudes. We address the following question: what ratio of S-wave acceleration to displacement is optimally indicative of magnitude? (We examine the acceleration to displacement ratio since, as we found in Chapter 4, the velocity dependence is not as statistically significant as that of acceleration and displacement.)

Recall that linear discriminant analysis involves solving the following eigenvalue problem:

$$\sum_w^{-1} \cdot \sum_a u = \lambda \cdot u \tag{F.1}$$

where $\sum_a$ and $\sum_w$ are the *among group* and *within group* covariance matrices, which are in turn, defined as a function of the data (or observation) matrix X. (See Chapter 4: A short note on linear discriminant analysis, Eqns 4.20 through 4.26).

Let the data (or observation) matrix $X_{S,2}$ be a 3373×2 matrix, with *S-wave* $log(acc)$ and $log(disp)$ as its two columns. The 5 groups are defined according to magnitude as they were in Chapter 4: $M < 3$, $3 \leq M < 4$, $4 \leq M < 5$, $5 \leq M < 6$, $M \geq 6$. The eigenvalues and eigenvectors of $\sum_w^{-1} \cdot \sum_a$ are:

$$\lambda_{1,S} = 3113.7 \quad , \qquad u_{1,S}^T = \begin{bmatrix} 0.36 & -0.93 \end{bmatrix} \tag{F.2}$$

$$\lambda_{2,S} = 36.7 \quad , \qquad u_{2,S}^T = \begin{bmatrix} -0.84 & 0.54 \end{bmatrix} \tag{F.3}$$

(Note: I've added the subscript S to distinguish the S-wave results in Eqns. F.2 and F.3 from those of those of the P-wave analysis in Chapter 4.) As discussed in Chapter 4, the eigenvector corresponding to the largest eigenvalue yields the linear combination that optimally separates the different groups, $Z_S = X_{S,2} \cdot u_{1,S} = 0.36log(acc) - 0.93log(disp)$. Thus, given a new set of S-wave observations $X_{new,S}$, we can project this new data vector onto $u_{1,S}$. The earthquake generating the new data is classified according to $Z_{new,S} = X_{new,S} \cdot u_{1,S}$. If $Z_{new,S} < 1.31$, then the new earthquake is in group 5, $M \geq 6$. If $1.31 < Z_{new,S} \leq 2.16$, then the event is in group 4, $5 \leq M < 6$. If $2.16 < Z_{new,S} \leq 2.82$, the event is in group 3, $4 \leq M < 5$. If $2.82 < Z_{new,S} \leq 3.35$, the event is in group 2, $3 \leq M < 4$. If $Z_{new,S} > 3.35$, then the event is in group 1, $M < 3$.

If we regress magnitude on Z_S, we find that

$$\begin{aligned} \hat{M}_{S,LDA2} &= -1.4599 Z_S + 8.05 \\ &= -0.526 log(acc) + 1.358 log(disp) + 8.05 \end{aligned} \tag{F.4}$$

It is interesting that eigenvectors from the S-wave analysis are nearly identical to those from the P-wave analysis, which are shown below

$$\lambda_{1,P} = 2663.1 \quad , \qquad u_{1,P}^T = \begin{bmatrix} 0.36 & -0.93 \end{bmatrix} \tag{F.5}$$

$$\lambda_{2,P} = 46 \quad , \qquad u_{2,P}^T = \begin{bmatrix} -0.83 & 0.56 \end{bmatrix} \tag{F.6}$$

The linear discriminant analysis on S-wave amplitudes is summarized by Figure F.1. Figure F.1 shows the (CISN) magnitude of the records in our database plotted against $Z_S = X_{S,2} \cdot u_{1,s} = 0.36log(acc) - 0.93log(disp)$. The curves drawn on the x axis are the best-fit normal curves to the Z_S coordinates of the data within each group. The vertical dashed lines are the decision boundaries described above. A solid black line shows the best-fit linear relationship between Z_S and magnitude. For comparison, the best-fit linear relationship between the optimal ratio of P-wave acceleration and displacement, Z_P, is plotted as a dashed black line. It appears that the magnitude estimates based on the S-wave ground motion ratios have a larger degree of saturation than the magnitude estimates based on the P-wave ratios. We can see this in Figure F.1, with the dashed line (representing magnitude to P-wave ra-

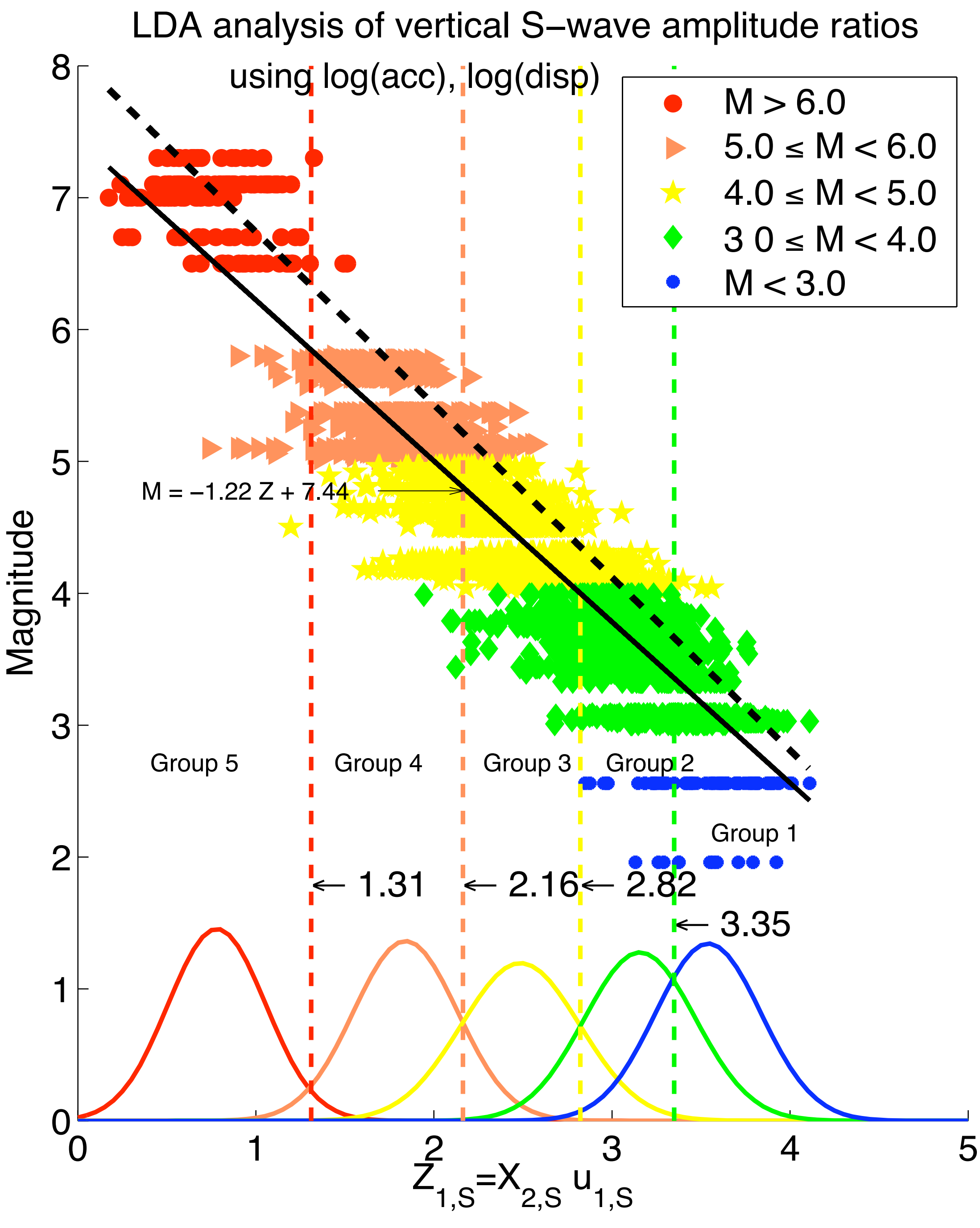

Figure F.1: Linear discriminant analysis of vertical S-wave $log(acc)$ and $log(disp)$.

tio relationship) consistently higher than the solid line (representing the relationship between S-wave ground motion ratio and magnitude).

How important is it to know whether the observed amplitudes are from P- or S-waves? In using the envelope amplitude attenuation relationships, it is very important to know whether the observations are from P- or S-waves. In Chapters 2 and 3, we saw that P- and S-waves are sufficiently different in their dependence on magnitude and epicentral distance that mistaking one for the other will make a difference. In using linear discriminant analysis results, it is not as important to be able to distinguish between P- and S-waves. The weight vectors that describe the linear combinations that are optimally indicative of magnitude are practically identical for P- and S-waves. For both body waves, $Z = 0.36log(acc) - 0.93log(disp)$. In Figure F, the solid lines mark the P-wave decision boundaries; the dashed lines denote those for the S-wave analysis. The unshaded regions are intervals of Z where both P- and S-wave classification agree. For example, the interval $Z < 1.31$ is in Group 5 ($M > 6$) regardless of whether we use the P- or S-wave decision boundaries. The shaded regions are intervals of Z where there are discrepancies between P- and S-wave analyses. Within these shaded regions, the P-wave decision boundaries place the earthquake in a higher magnitude (larger event) group than the S-wave analysis. For example, shaded region $1.31 < Z < 1.72$, is in Group 5 ($M > 6$) based on the P-wave results, and in Group 4 ($5 < M \leq 6$) based on the S-wave decision boundaries.

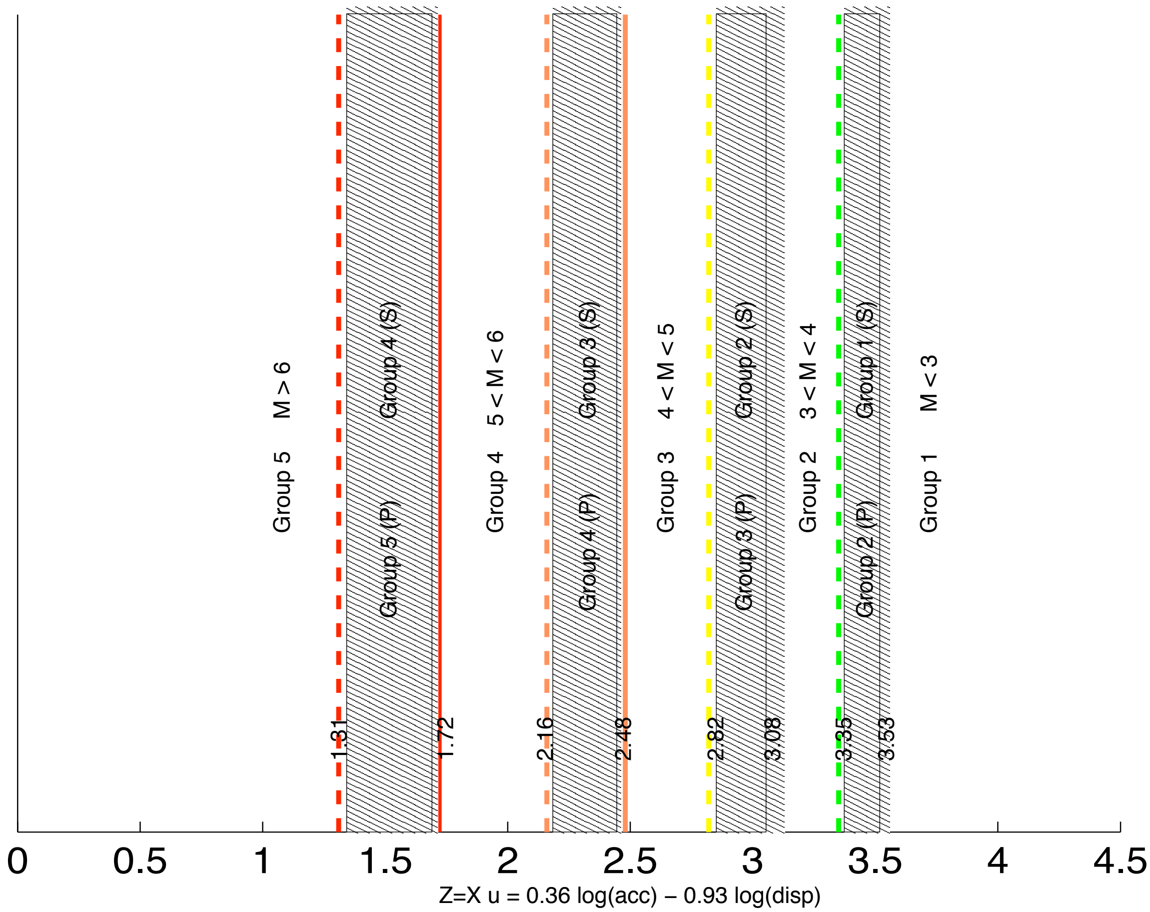

Figure F.2: Decision boundaries for LDA analysis on P- and S-waves. The solid lines mark the decision boundaries from the P-wave analysis. The dashed lines mark the S-wave analysis decision boundaries. In the unshaded regions, there is agreement between the P- and S-wave classification. In the shaded intervals, the P-wave decision boundaries place the earthquake in a higher magnitude group than does the S-wave analysis.

Confusion matrix for LDA of S-wave amplitudes as magnitude indicators
using acceleration and displacement
Row: actual group; column: classification based on LDA

	Group 1	Group 2	Group 3	Group 4	Group 5	Total obs.
Group 1	71% (52)	20% (20)	0% (0)	0% (0)	0% (0)	=72
Group 2	27% (292)	60% (659)	12% (134)	$< 1\%$ (4)	0% (0)	=1094
Group 3	$< 1\%$ (3)	15% (220)	68% (976)	17% (239)	$< 1\%$ (1)	=1439
Group 4	0% (0)	0% (0)	13% (82)	83% (516)	4% (25)	=623
Group 5	0% (0)	0% (0)	0% (0)	3% (4)	97% (141)	=145

Table F.1: Confusion matrix for linear discriminant analysis of vertical S-wave amplitudes as magnitude indicators (using acceleration and displacement). The confusion matrix provides an idea of how often the decision boundaries from the linear discriminant analysis correctly classify (or misclassify) the observations. Recall that group 1 is $M < 3$, group 2 is $3 \leq M < 4$, group 3 is $4 \leq M < 5$, group 4 is $5 \leq M < 6$, and group 5 is $M \geq 6$.