

# Principal Component Analysis of Dipole Moment Derivative Signs of Chloroform

Elisabete Suto, M.M.C. Ferreira, and Roy E. Bruns\*

Instituto de Quimica, Universidade Estadual de Campinas, CP 6154, 13081 Campinas, SP, Brazil

Received 20 September 1990; accepted 11 March 1991

The use of principal components as a basis for a graphical procedure to analyze polar tensor data is proposed. Molecular orbital and experimental polar tensor data for all possible sign combinations of the  $\partial\mathbf{p}/\partial Q_j$  of  $\text{CHCl}_3$  and  $\text{CDCl}_3$  are represented graphically as principal component scores facilitating sign selection for the  $\partial\mathbf{p}/\partial Q_j$ . The graphs are particularly useful in analyzing an apparent conflict in  $\partial\mathbf{p}/\partial Q_j$  sign choices based on the isotopic invariance criterion and molecular orbital results for the  $A_1$  symmetry species of these molecules. The numerical impacts of individual sign ambiguities for the  $\partial\mathbf{p}/\partial Q_j$  on the polar tensor data are measured by the variances associated with the principal components. Assuming the  $\partial\mathbf{p}/\partial Q_j$  sign sets with indeterminate signs provide replicated results for the polar tensor elements, their errors are estimated and compared with errors obtained previously by propagating intensity uncertainties through the polar tensor equations.

## INTRODUCTION

The reduction of gas phase vibrational intensity data into molecular parameters is hindered by sign ambiguities in the dipole moment derivatives with respect to the normal coordinates, the  $\partial\mathbf{p}/\partial Q_j$ . Sign determinations are normally attempted comparing polar tensor elements or dipole moment derivatives with respect to symmetry coordinates,  $\partial\mathbf{p}/\partial S_k$ , for isotopically related molecules.<sup>1</sup> In special cases other experimental information, such as the signs of the Coriolis interaction constants,<sup>2</sup> has been used to eliminate sign ambiguity; however complete sign attribution based on experimental data has always resorted to a comparison of derivative values for hydrogen/deuterium substituted molecules. Unambiguous sign attributions are usually not possible due to experimental errors in the vibrational intensities and normal coordinates. These errors are not always easy to estimate especially when the molecular spectrum contains overlapping bands or there is some doubt about the force field approximations used in the normal coordinate calculations. For some hydrogen containing molecules and their deuterium analogues several pairs of sign sets of the  $\partial\mathbf{p}/\partial Q_j$  have polar tensor element or  $\partial\mathbf{p}/\partial S_k$  values identical within the estimated experimental errors. Each pair contains one sign set of the  $\partial\mathbf{p}/\partial Q_j$  for the hydrogen

analogue and another for the deuterium one. Quantum chemically calculated values have been shown to be very useful in choosing among such experimentally acceptable sign sets<sup>3</sup> although wave function limitations often lessen our confidence in sign choices based on the theoretical estimates.

Sign selection would be much easier if a reliable two- or three-dimensional graphical scheme could be devised representing the values of the polar tensor elements as a function of the sign alternatives of the  $\partial\mathbf{p}/\partial Q_j$ . This graphical procedure could incorporate molecular orbital estimates as well as experimental polar tensor data permitting a simultaneous visual comparison of theoretical values with all possible experimental polar tensor values. Graphical representations<sup>4</sup> of the  $G$ -sum rule, reported some time ago, are only partially successful for sign determination analysis due to statistical losses occurring for two dimensional representations of the higher order polar tensor space.

In this work we propose the use of principal components<sup>5</sup> as the basis of this graphical procedure. An application using the experimental vibrational intensity data of Kim and King<sup>6</sup> and Tanabe and Saeki<sup>7</sup> for  $\text{CHCl}_3/\text{CDCl}_3$  and *ab initio* molecular orbital values is reported. Signs of the  $\partial\mathbf{p}/\partial Q_j$  are chosen graphically by simultaneously comparing principal component scores of experimental and theoretical polar tensor elements. The statistical importance of the fundamental vibrational intensities in sign determinations is shown to be a natural consequence of this procedure.

\*Author to whom all correspondence should be addressed.

## PRINCIPAL COMPONENT REPRESENTATION OF POLAR TENSOR ELEMENTS OF CHLOROFORM

The polar tensor of chloroform is defined as a juxtaposition of atomic polar tensors

$$P_X = \{P_X^{(C)} : P_X^{(H)} : P_X^{(Cl_1)} : P_X^{(Cl_2)} : P_X^{(Cl_3)}\}.$$

Each atomic polar tensor,  $P_X^{(\alpha)}$ , is given by

$$P_X^{(\alpha)} = \begin{pmatrix} \partial p_x / \partial x_\alpha & \partial p_x / \partial y_\alpha & \partial p_x / \partial z_\alpha \\ \partial p_y / \partial x_\alpha & \partial p_y / \partial y_\alpha & \partial p_y / \partial z_\alpha \\ \partial p_z / \partial x_\alpha & \partial p_z / \partial y_\alpha & \partial p_z / \partial z_\alpha \end{pmatrix} = \begin{pmatrix} p_{xx}^{(\alpha)} & p_{xy}^{(\alpha)} & p_{xz}^{(\alpha)} \\ p_{yx}^{(\alpha)} & p_{yy}^{(\alpha)} & p_{yz}^{(\alpha)} \\ p_{zx}^{(\alpha)} & p_{zy}^{(\alpha)} & p_{zz}^{(\alpha)} \end{pmatrix}$$

where  $\alpha$  represents the specific atom. Using a molecular orientation with the H, C, and one of the Cl atoms in the  $xz$ -plane and with the H atom along the  $Z$ -axis simplifies the atomic polar tensors since several elements of the atomic polar tensors are required to be zero by symmetry. Also for convenience the polar tensor elements of each symmetry species of  $\text{CHCl}_3$ ,  $A_1$ , and  $E$ , are treated separately. Four non-zero polar tensor elements,  $p_{zz}^{(C)}$ ,  $p_{zz}^{(Cl)}$ ,  $p_{zz}^{(Cl)}$ , and  $p_{zz}^{(H/D)}$ , exist for the  $A_1$  species whereas five elements,  $p_{xx/yy}^{(C)}$ ,  $p_{xx/yy}^{(H/D)}$ ,  $p_{xx}^{(Cl)}$ ,  $p_{zz}^{(Cl)}$ , and  $p_{yy}^{(Cl)}$ , having different numerical values, are not zero by symmetry for the  $E$  species. The dependence of the element values as a function of the derivative signs can be studied in four- and five-dimensional spaces for these symmetry species where each polar tensor element is represented as a coordinate axis and points in these spaces stand for the polar tensor values of the different sign combinations. Two- or three-dimensional projections of these higher order spaces obtained by plotting one tensor element against another cannot be used for determining isotopically invariant sign choices because of the lack of relevant statistical information about the polar tensor elements not included in the projections. On the other hand linear transformations of polar tensor element variables corresponding to rotations of the original coordinate axes can provide projections with more statistical information. Indeed principal component rotations provide bi and tridimensional projections containing a maximum of statistical information for all possible linear transformations of the original polar tensor elements. The first principal component corresponds to the direction in higher order space explaining a maximum of the data variance. The second component is perpendicular to the first one and explains a maximum of the residual variance. Hence a graph using as ordinates the first two principal components is the bidimensional projection containing the largest amount of statistical information for all linear transformations of the polar tensor elements

and can be useful in studying the dependence of the polar tensor element values as a function of the signs of the  $\partial \mathbf{p} / \partial Q_j$ . If desirable, a third principal component, perpendicular to the first two, can be used in the graphical representations.

The principal component equation<sup>8</sup> applied to polar tensor elements can be expressed as

$$p_{i,\sigma\nu}^{(\alpha)} = \bar{p}_{\sigma\nu}^{(\alpha)} + \sum_{a=1}^A t_{ia} b_{a,\sigma\nu}^{(\alpha)} + e_{i,\sigma\nu}^{(\alpha)} \quad (1)$$

where  $\sigma, \nu = x, y, z, i$  represents the  $i$ th set of signs of the  $\partial \mathbf{p} / \partial Q_j$  and  $\bar{p}_{\sigma\nu}^{(\alpha)}$  is the average value of the  $\sigma\nu$ th polar tensor element of the  $\alpha$ th atom for all possible  $\partial \mathbf{p} / \partial Q_j$  sign choices. The  $b_{a,\sigma\nu}^{(\alpha)}$  elements are called loadings and are the direction cosines relating the rotated coordinate system to the original one. The  $t_{ia}$  values are the scores giving the new coordinate values of the  $i$ th set of signs for the  $a$ th principal component. These values are used to construct the low dimensional representations of the higher order space. The  $e_{i,\sigma\nu}^{(\alpha)}$  are residual values expressing the difference between the experimental value of  $p_{i,\sigma\nu}^{(\alpha)}$  and the one predicted by the principal component model. These residuals contain both experimental and modeling error. If the  $e_{i,\sigma\nu}^{(\alpha)}$  values are larger than the experimental errors and a single bidimensional projection is not sufficient to give an accurate representation of the sign dependence of the polar tensor element values additional projections involving the third, fourth, etc. principal components can be investigated.

Principal components for the polar tensor elements are easily calculated by first constructing a data matrix,  $\mathbf{X}$ , for which each possible sign combination corresponds to a row and each column contains the values of a specific nonzero polar tensor element. Hence, including both  $\text{CHCl}_3$  and  $\text{CDCl}_3$  the data matrices for the  $A_1$  and  $E$  symmetry species are  $16 \times 4$  and  $16 \times 5$ , respectively. The principal components are the eigenvectors obtained by diagonalizing the covariance matrix,  $\mathbf{X}'\mathbf{X}$ , for each symmetry species. Each eigenvalue gives the quantity of data variance explained by its associated principal component.

For the  $E$  species the  $p_{xx}^{(C)}$  and  $p_{yy}^{(C)}$  polar tensor element values are identical by symmetry. This is also true for these two elements of the hydrogen polar tensor. Since no additional statistical information is obtained by including both the  $xx$  and  $yy$  elements in the principal component analysis the  $yy$  tensor element values have not been included in the data matrices.

Principal component scores for molecular orbital polar tensor values are simply obtained substituting these values into eq. (1) using the loadings obtained from the polar tensor experimental values. The principal component model is assumed accurate in this calculation and the  $e_{i,\sigma\nu}^{(\alpha)}$  values are zero.

## CALCULATIONS

Experimental fundamental intensity values for  $\text{CHCl}_3$  and  $\text{CDCl}_3$  were taken from Table II of reference 6. Normal coordinates were calculated using the force field reported by Rouff and Burger<sup>9</sup> and are essentially identical with the normal coordinates reported in Table III of reference 6 except for an apparent typographical error there for the  $L_{33}$  element of  $\text{CDCl}_3$ . Our value of  $-0.0115 \text{ u}^{-1/2}$  is 10 times smaller in absolute magnitude than the value given in reference 6. Also, as expected, its value is very similar to the value of  $L_{33}$  for  $\text{CHCl}_3$ ,  $-0.009074 \text{ u}^{-1/2}$  in reference 6. The bond lengths, valence angles, and dipole moment used in the polar tensor calculations are the same as those used by Kim and King.<sup>6</sup> The Cartesian coordinate system and chloroform orientation used in these calculations are identical to those of reference 6. One Cl atom and the H and C atoms are located in the  $xz$  plane with the C atom at the origin and the H atom along the positive  $Z$  axis.

Molecular orbital values of the  $\text{CHCl}_3$  polar tensor elements were calculated using the Gaussian 82 computer program.<sup>10</sup> Experimental polar tensor values were calculated from the fundamental intensities, normal coordinates, dipole moment, and molecular geometry data using the TPOLAR program.<sup>11</sup> Principal components for the polar tensor data were calculated using a microcomputer version of the ARTHUR/75 program.<sup>12</sup> Other general statistic packages, such as SAS, Statgraph, and SPSS could just as well have been used. The  $X$  matrix is formed as follows. Each row corresponds to polar tensor values for one of the  $\partial \mathbf{p}/\partial Q_j$  sign combinations of either  $\text{CHCl}_3$  or  $\text{CDCl}_3$ . The four columns of this matrix for the  $A_1$  symmetry species have values of the  $p_{zz}^{(C)}$ ,  $p_{zz}^{(H)}$ ,  $p_{zz}^{(D)}$  and  $p_{zz}^{(H/D)}$  polar tensor elements. This matrix for the  $E$  symmetry species has five columns for the  $p_{xx}^{(C)}$ ,  $p_{xx}^{(H)}$ ,  $p_{xx}^{(D)}$ ,  $p_{xx}^{(H/D)}$ ,  $p_{yy}^{(C)}$ ,  $p_{yy}^{(H)}$ ,  $p_{yy}^{(D)}$ ,  $p_{yy}^{(H/D)}$  tensor values. Molecular orbital values of the tensor elements are *not* used to calculate the principal components. Rather,

principal component transformations determined from the polar tensor values obtained from the experimental intensities are applied to the theoretical polar tensor values. The transformed values can then be compared with the principal component scores derived from the experimental data.

## RESULTS AND DISCUSSION

The principal component equations calculated for the  $A_1$  and  $E$  symmetry species are presented in Table I. Three principal components completely describe the variance in the  $A_1$  species since the four  $A_1$  tensor elements contain a redundancy relating the  $p_{zz}$  elements. The scores of the first and second principal components for the  $\partial \mathbf{p}/\partial Q_j$  sign alternatives are plotted in Figure 1. The first principal component ( $\text{PC}_1$ ), corresponding to the largest eigenvalue of the  $X'X$  matrix, accounts for 90.9% of the total polar tensor data variance. In other words this transformed coordinate statistically explains almost 91% of the polar tensor data variation provoked by the sign ambiguities in the  $\partial \mathbf{p}/\partial Q_j$ . It discriminates between the sign choices for  $\partial \mathbf{p}/\partial Q_2$  which is related to the  $A_2$  intensity values representing 84.3% and 93.1% of the  $A_1$  symmetry species intensity sums of  $\text{CHCl}_3$  and  $\text{CDCl}_3$ , respectively. A positive sign for this derivative, independent of the signs of the other derivatives, corresponds to positive  $\text{PC}_1$  scores whereas sign choices with  $\partial \mathbf{p}/\partial Q_2 < 0$  have negative  $\text{PC}_1$  scores. These scores are predominantly determined by the values of  $p_{zz}^{(C)}$ ,  $p_{zz}^{(H)}$ , and  $p_{zz}^{(D)}$  which have the largest absolute loadings on  $\text{PC}_1$  (Table I).

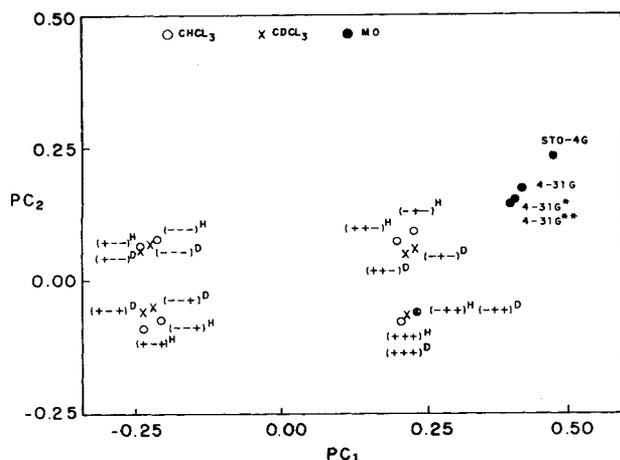
The second principal component,  $\text{PC}_2$ , accounts for 8.4% of the total variance and discriminates between the sign choices for  $\partial \mathbf{p}/\partial Q_3$ . The squares of the values of this derivative are proportional to the  $A_3$  intensities of  $\text{CHCl}_3$  and  $\text{CDCl}_3$  accounting for 9.6% and 5.7% of the total  $A_1$  intensity sums of these molecules. Positive  $\text{PC}_2$  scores correspond to a negative sign choice for this derivative whereas sign alter-

**Table I.** Principal component score equations for the  $A_1$  and  $E$  symmetry species of chloroform.<sup>a</sup>

$A_1$ symmetry species	
$\text{PC}_1 = 0.81 p_{zz}^{(C)} + 0.12 p_{zz}^{(H)} + 0.48 p_{zz}^{(D)} - 0.31 p_{zz}^{(H/D)}$	(90.9) <sup>b</sup>
$\text{PC}_2 = 0.45 p_{zz}^{(C)} + 0.02 p_{zz}^{(H)} - 0.88 p_{zz}^{(D)} - 0.16 p_{zz}^{(H/D)}$	(8.4)
$\text{PC}_3 = -0.21 p_{zz}^{(C)} + 0.94 p_{zz}^{(H)} - 0.04 p_{zz}^{(D)} - 0.24 p_{zz}^{(H/D)}$	(0.7)
$E$ symmetry species	
$\text{PC}_1 = 0.84 p_{xx}^{(C)} + 0.06 p_{xx}^{(H)} - 0.47 p_{xx}^{(D)} + 0.22 p_{xx}^{(H/D)} - 0.13 p_{yy}^{(C)}$	(92.8)
$\text{PC}_2 = 0.28 p_{xx}^{(C)} - 0.80 p_{xx}^{(H)} + 0.12 p_{xx}^{(D)} - 0.46 p_{xx}^{(H/D)} + 0.23 p_{yy}^{(C)}$	(7.1)
$\text{PC}_3 = -0.23 p_{xx}^{(C)} + 0.05 p_{xx}^{(H)} - 0.62 p_{xx}^{(D)} - 0.02 p_{xx}^{(H/D)} + 0.74 p_{yy}^{(C)}$	(0.1)

<sup>a</sup>The symbol H stands for both the hydrogen and the deuterium atoms.

<sup>b</sup>Percentage variance explained by the principal component.



**Figure 1.** Score graph of the first two principal components of polar tensor data for the  $A_1$  symmetry species of  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . Within parenthesis signs of the  $\partial\mathbf{p}/\partial Q_i$  in the order  $i = 1, 2, 3$  for the H and D substituted analogues are presented.

natives with  $\partial\mathbf{p}/\partial Q_3 > 0$  have negative  $\text{PC}_2$  scores. The sign of this derivative is strongly influenced by values of  $p_{zz}^{(C)}$  and  $p_{zx}^{(Cl)}$  which have large absolute loadings on this principal component.

The first two components explain 99.3% of the variance in the  $A_1$  species polar tensor data for  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . The third component accounting for only 0.7% of the variance partially discriminates between positive and negative values of  $\partial\mathbf{p}/\partial Q_1$ . The intensities of the first fundamental bands of  $\text{CHCl}_3$  and  $\text{CDCl}_3$  are 6.1 and 1.1% of the total intensity sums. Although these intensity values are not that much smaller than the  $A_3$  values, the principal component results clearly show that the effect of sign uncertainty for  $\partial\mathbf{p}/\partial Q_1$  results in less than 1% of the total variation in the  $A_1$  symmetry species polar tensor data. On the other hand the sign ambiguity for  $\partial\mathbf{p}/\partial Q_3$  accounts for almost 10% of this variance. Hence, it is not surprising that the sign of  $\partial\mathbf{p}/\partial Q_1$  can not be determined comparing polar tensor results for the  $\text{CHCl}_3$  and  $\text{CDCl}_3$  experimental intensities.

The principal component graph in Figure 1 provides an approximate representation, containing 99.3% of the total variance, of the four dimensional space spanned by the  $A_1$  symmetry polar tensor values. The isotopic invariance criterion favors sign sets for which points representing the  $\text{CHCl}_3$  and  $\text{CDCl}_3$  molecules are in close proximity. For the sign alternatives in Figure 1 the  $(- + +)$  and  $(+ + +)$  sets best satisfy this criterion.

Errors in the principal component scores can be propagated from the errors in the polar tensor elements. Assuming no error in the principal component transformation the variance in the principal component scores can be estimated as

$$\hat{V}(\text{PC}_a) = \hat{V} \left( \sum_{\sigma\nu} b_{a,\sigma\nu}^{(\alpha)} p_{i,\sigma\nu}^{(\alpha)} \right) = \sum_{\sigma\nu} b_{a,\sigma\nu}^{(\alpha)^2} \hat{V}(p_{i,\sigma\nu}^{(\alpha)}) \quad (2)$$

where  $\hat{V}$  represents the variances of the quantities in parentheses. Standard errors of the principal component scores are simply the square roots of their variances. Applying the principal component loadings in Table I and the 99% confidence standard errors of reference 6 error uncertainties of the size of the symbols in Figure 1 (or somewhat smaller) are obtained for the principal component scores. This analysis indicates preferences for the  $(- + +)$  and  $(+ + +)$  signs of the  $\partial\mathbf{p}/\partial Q_j$  of  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . However Kim and King<sup>6</sup> clearly state that their error estimates are lower limiting values. Not all sources of error<sup>13</sup> were included in their analysis. Especially important are error contributions to the polar tensor elements due to the  $\text{CHCl}_3$  and  $\text{CDCl}_3$  normal coordinate uncertainties. Systematic errors from this and other sources could increase the experimental polar tensor element uncertainties.

Principal component scores for polar tensor values calculated using STO-4G, 4-31G, 4-31G\*, and 4-31G\*\* basis sets are also represented in Figure 1. The 4-31G results are somewhat better than the STO-4G ones since they are closer to the experimentally derived points. The MO values, transformed in principal component scores, clearly favor the  $(+ + -)$  and  $(- + -)$  sign choices. Recalling that uncertainty in the sign of  $\partial\mathbf{p}/\partial Q_1$  has a negligible effect on the polar tensor values the four sign alternatives positioned in the upper right quadrant can be considered to be derived from fourfold replicate experiments with  $\partial\mathbf{p}/\partial Q_j$  signs represented by  $(\pm + -)$ . Average and standard error values for these sign choices are given in Table II. Note that the standard errors of the replicates are about 0.02  $e$  or less and not that much larger than the errors reported for the polar tensor values in reference 6 so that the interpretation of these values using electronic configuration concepts is not seriously hampered using our more conservative error estimates. Averages and error uncertainties for the group of sign alternatives in the lower right quadrant  $(\pm + +)$  are also presented in Table II. These sets of polar tensor values considered as replicate results produce the same error as those for the  $(\pm + -)$  sign alternatives for  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . Admitting errors of these sizes in the experimental polar tensor values, the  $(\pm + -)$  sign alternatives can be preferred based on the proximities of their polar tensor values to those obtained from the MO calculations. The molecular orbital estimates of the polar tensor elements have also been included in Table II for comparison with the experimental values. Note that the largest changes upon reversing the  $\partial\mathbf{p}/\partial Q_3$  sign occurs for  $p_{zz}^{(C)}$  and  $p_{zx}^{(Cl)}$  which have the largest absolute loadings for  $\text{PC}_2$ . Our preferred sign set has polar tensor values in close agreement with those obtained in reference 6, where molecular orbital results were also used.

It should be mentioned that the  $(\pm - -)$  and  $(\pm - +)$  sign alternatives might also be considered

**Table II.** Preferred polar tensor element values and error estimates for chloroform ( $e$ ).<sup>a</sup>

$A_1$	$p_{zz}^{(C)}$	$p_{zz}^{(H)}$	$p_{zz}^{(Cl)}$	$p_{zz}^{(Cl)}$
( $\pm + -$ ) <sup>b</sup>	0.213 $\pm$ 0.019	0.031 $\pm$ 0.020	0.047 $\pm$ 0.016	-0.081 $\pm$ 0.002
( $\pm + +$ )	0.156 $\pm$ 0.019	0.024 $\pm$ 0.019	0.169 $\pm$ 0.002	-0.060 $\pm$ 0.004
Ref. 6a	0.198 $\pm$ 0.009	0.045 $\pm$ 0.004	0.046 $\pm$ 0.012	0.081 $\pm$ 0.002
4-31G**	0.417	-0.020	0.067	-0.132
4-31G*	0.413	-0.021	0.069	-0.127
4-31G	0.432	0.007	0.052	-0.148
STO-4G	0.480	0.117	0.020	-0.199

$E$	$p_{xx/yy}^{(C)}$	$p_{xx/yy}^{(H)}$	$p_{xx}^{(Cl)}$	$p_{zz}^{(Cl)}$	$p_{yy}^{(Cl)}$
( $+ - \pm$ )	1.128 $\pm$ 0.004	-0.055 $\pm$ 0.000	-0.590 $\pm$ 0.018	0.285 $\pm$ 0.001	-0.125 $\pm$ 0.021
Ref. 6a	1.147 $\pm$ 0.008	-0.050 $\pm$ 0.002	-0.581 $\pm$ 0.010	0.127 $\pm$ 0.002	-0.151 $\pm$ 0.009
4-31G**	1.397	-0.069	-0.694	0.141	-0.192
4-31G*	1.402	-0.074	-0.699	0.138	-0.191
4-31G	1.286	-0.044	-0.618	0.110	-0.213
STO-4G	1.242	-0.057	-0.533	0.093	-0.257

<sup>a</sup>Units of electrons,  $e$ .  $1 e = 4.803 \text{ D } \text{\AA}^{-1}$ .

<sup>b</sup>Preferred sets of polar tensor elements. See text for discussion.

as isotopically invariant sets of polar tensor values. These alternatives contain polar tensor values with signs opposite to those given in Table II and were eliminated in our analysis because of inconsistencies with the MO calculated values.

A graph of the first and second principal components for the  $E$  symmetry species polar tensor data is shown in Figure 2. As indicated in Table I these components span 99.9% of the original five ordered space. The first component which accounts for 92.8% of the total polar tensor variance discriminates between signs of the  $\partial\mathbf{p}/\partial Q_5$  derivatives, positive signs of this derivative having negative scores whereas negative signs have positive scores. The  $A_5$  intensity values provide contributions of 87.8% and 59.3% to the total  $\text{CHCl}_3$  and  $\text{CDCl}_3$   $E$  fundamental intensity sums. The second component describing 7.1% of the

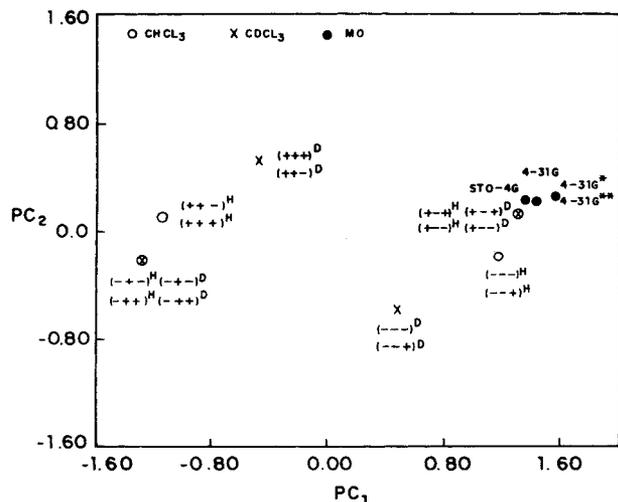
data variance discriminates between sign sets with different signs for  $\partial\mathbf{p}/\partial Q_4$ . The  $A_4$  intensity values of 30.2 and 100 km/mol contribute 12.1% and 40.7% to the  $E$  symmetry species intensity sums for  $\text{CHCl}_3$  and  $\text{CDCl}_3$  respectively. However uncertainty in the sign of  $\partial\mathbf{p}/\partial Q_4$  has a much smaller impact on the polar tensor values than might have been expected if only the sizes of the  $A_4$  intensity values are considered.

The third component explains only 0.1% of the total variance yet discriminates between the signs of the  $\partial\mathbf{p}/\partial Q_6$  derivative. However since this component spans such a small variance, a secure sign selection for  $\partial\mathbf{p}/\partial Q_6$  based on the  $\text{PC}_3$  scores can not be made. This is not surprising since the  $A_6$  values for both  $\text{CHCl}_3$  and  $\text{CDCl}_3$  are almost zero.

Inspection of Figure 2 shows that only two pairs of sign sets ( $+ - \pm$ ) and ( $- + \pm$ ) are isotopically invariant for  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . Of these the ( $+ - \pm$ ) alternatives are in excellent agreement with the STO-4G and 4-31G results. These alternatives for the  $E$  symmetry species polar tensor elements of  $\text{CHCl}_3$  and  $\text{CDCl}_3$  can be considered as fourfold replicate results. Average and standard errors for these polar tensor elements are included in Table II. These results are also in excellent agreement with the values originally determined by Kim and King.<sup>6</sup>

## CONCLUSIONS

The polar tensor formalism has resulted in dramatic improvements in the application and interpretation of fundamental vibrational intensities. The concept of dipole moment derivatives as atomic properties in molecules rather than as chemical bond properties has allowed successful intensity predictions based on transference procedures<sup>14</sup> and the formulation of



**Figure 2.** Score graph of the first two principal components of polar tensor data for the  $E$  symmetry species of  $\text{CHCl}_3$  and  $\text{CDCl}_3$ . Within parenthesis signs of the  $\partial\mathbf{p}/\partial Q_5$

methane molecules. These same models also predict that the intensity sums of rotational isomers are the same and explain the surprising regularity found for derivatives of the  $X_2CY$  ( $X = F, Cl, Br$  and  $Y = O, S$ ) molecules. In addition the polar tensor formalism naturally allows for the treatment of the rotational contributions which encumbered interpretations based on values of dipole moment derivatives with respect to symmetry coordinates,  $\partial \mathbf{p} / \partial S_j$ .

However the polar tensor formalism does suffer from one inconvenience which is less aggravating for treatments using symmetry coordinates. Analyses performed in symmetry coordinate space are usually lower dimensional than those in atomic Cartesian coordinate space. In our applications for the  $A_1$  and  $E$  symmetry species of trichloromethane symmetry coordinate spaces are three dimensional whereas the corresponding atomic Cartesian spaces are four and five dimensional, respectively, after elimination of trivial redundancies imposed by symmetry. The conservation of charge relation,  $\sum \mathbf{P}_X = \mathbf{0}$ , is embedded in the polar tensor element values complicating the intensity analysis. Chemometric techniques, such as principal component analysis, are especially efficient at automatically eliminating redundancies, permitting reductions to lower ordered spaces and facilitating graphical representations of vibrational intensity results.

As shown here, principal component analysis appears to be a very useful tool for analyzing the sign ambiguities of dipole moment derivatives with respect to normal coordinates. Bidimensional graphical representations approximate the spatial orientations of all possible sign combinations for all the isotopically related molecules as well as their relative positions with respect to MO calculated values. Apparent conflicts for sign attributions based on different criteria, for the chloroform application reported here isotopic invariance and MO calculated results, can be analyzed graphically permitting more secure decisions about the signs of the  $\partial \mathbf{p} / \partial Q_j$ . Also the impact of specific derivative sign ambiguities on the polar tensor data is measured by the variances associated with principal components discriminating between these derivatives. For derivatives calculated using weak intensities, signs can not often be determined. However advantage of this situation can be taken by considering all indeterminate sign set alternatives as sources of estimates of the true polar tensor values. Error estimates obtained in this

way can be compared with error estimates in the polar tensor elements propagated from the intensity measurement errors, normal coordinate uncertainties, and other error sources.

Finally it should be mentioned that the values of the effective charges of our preferred sign sets are in excellent agreement with the values reported by Kim and King.<sup>6</sup> The latter values were used recently in the investigations proposing the electronegativity models for vibrational intensities of substituted methanes.<sup>15</sup>

The authors gratefully acknowledge partial financial support from FAPESP, CNPq and FINEP. E.S. is supported by a CNPq graduate student fellowship.

## References

1. M. Gussoni, in *Vibrational Intensities in Infrared and Raman Spectroscopy*, W.B. Person and G. Zerbi, Eds., Elsevier, Amsterdam, 1982, p. 96–120.
2. S. Kondo, T. Nakanaga, and S. Saeki, *J. Chem. Phys.*, **73**, 5409 (1980).
3. R.E. Bruns, in *Vibrational Intensities in Infrared and Raman Spectroscopy*, W.B. Person and G. Zerbi, Eds., Elsevier, Amsterdam, 1982, p. 143–157.
4. B.B. Neto and R.E. Bruns, *J. Chem. Phys.*, **71**, 5042 (1979).
5. K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, Academic Press, New York, 1979, p. 213–254.
6. K. Kim and W.T. King, *J. Chem. Phys.*, **80**, 978 (1984).
7. K. Tanabe and S. Saeki, *Spectrochim Acta Part A*, **26**, 1469 (1970).
8. S. Wold, *Pattern Recognition*, **8**, 127 (1976).
9. A. Ruoff and H. Burger, *Spectrochim. Acta Part A*, **26**, 989 (1970).
10. J.S. Binkley, M.J. Frisch, D.J. DeFrees, K. Raghavahari, R.A. Whiteside, H.B. Schelgel, E.M. Fluder and J.A. Pople, Gaussian 82, Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburg, PA, 1984.
11. A.B.M.S. Bassi, Doctoral Dissertation, Universidade Estadual de Campinas, (1975).
12. I.S. Scarminio and R.E. Bruns, *Trends Anal. Chem.*, **8**, 326 (1989).
13. J.W. Russell, C.D. Needham, and J. Overend, *J. Chem. Phys.*, **45**, 3383 (1966).
14. B.J. Krohn, W.B. Person, and J. Overend, *J. Chem. Phys.*, **65**, 969 (1976); B.J. Krohn, W.B. Person, and J. Overend, *J. Chem. Phys.*, **67**, 5091 (1977). B.B. Neto, M.N. Ramos and R.E. Bruns, *J. Chem. Phys.*, **85**, 4515 (1986).
15. B.B. Neto, I.S. Scarminio, and R.E. Bruns, *J. Chem. Phys.*, **89**, 1887 (1988). B.B. Neto and R.E. Bruns, *J. Phy. Chem.*, **94**, 1764 (1990).