A Chemometric Analysis of *Ab Initio* Vibrational Frequencies and Infrared Intensities of Methyl Fluoride

ANA LUIZA M. S. DE AZEVEDO and BENÍCIO B. NETO
Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50739, Recife, PE, Brazil

IEDA S. SCARMINIO
Departamento de Química, Universidade Estadual de Londrina, Londrina, PR, Brazil

ANSELMO E. DE OLIVEIRA and ROY E. BRUNS*
Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13081-970 Campinas, SP, Brazil

Received 28 November 1994; accepted 14 April 1995

ABSTRACT

Factorial design and principal component analyses are applied to CH_3F infrared frequencies and intensities calculated from *ab initio* wave functions. In the factorial analysis, the quantitative effects of changing from a 6-31G to a 6-311G basis, of including polarization and diffuse orbitals, and of correcting for electron correlation using the second-order Møller-Plesset procedure are determined for all frequencies and intensities. The most significant main effect observed for the frequencies corresponds to the shift from Hartree-Fock to MP2 calculations, which tends to lower all frequency values by approximately 100 cm^{-1}. For the intensities, the main effects are larger for the CF stretching and the CH_3 asymmetric stretching modes. Interaction effects between two or more of the four factors are found to be of minor importance, except for the interaction between correlation and polarization. The principal component analysis indicates that wave functions with polarization and diffuse orbitals at the second-order Møller-Plesset level provide the best estimates for the harmonic frequencies, but not for the intensities. For the frequencies, the first principal component distinguishes between MP2 and Hartree-Fock calculations, while the second component separates the wave functions with polarization orbitals from those without these orbitals. For the intensities, the separation is similar but less well

*Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 17, No. 2, 167–177 (1996)

CCC 0192-8651 / 96 / 020167-11

defined. This analysis also shows that wave function optimization to calculate accurate intensities is more difficult than an optimization for frequencies.

Introduction

The selection of a wave function capable of providing accurate estimates of molecular vibrational frequencies and infrared intensities is a difficult task. A wave function providing accurate frequency values does not always yield good intensity values. Furthermore, wave functions resulting in accurate frequency and intensity values for some characteristic group vibrations do not necessarily provide good estimates for other characteristic groups. Wave function modifications that improve the agreement between the calculated and experimental values for some frequencies and intensities often result in poorer estimates for others.

The basis set dependence of calculated spectral parameters has been studied for a wide variety of small molecules. Stanton et al.[1] reported Hartree-Fock (HF) and many-body perturbation results for the infrared intensities of methane, hydrogen cyanide, formaldehyde, ammonia, water, and hydrogen fluoride, noting significant dependencies on basis set characteristics and electron correlation treatment. In a study of 13 molecules including those just mentioned, Yamaguchi et al.[2] emphasize the importance of using split valence polarized basis sets for accurate theoretical intensity predictions. Simandirus et al.[3] demonstrate the efficiency of correcting for electron correlation effects using the MP2 procedure. Miller et al.[4] state that polarization and diffuse functions on heavy atoms are essential for obtaining accurate infrared intensities and suggest the use of the MP2/6-31 + G(d) wave function as a reasonable compromise between computational expense and reliability.

Taken in its most general form, wave function optimization can be regarded as a multivariate statistical problem. The effects of several wave function characteristics, such as the type of basis set employed, the presence of polarization, and/or diffuse orbitals and the treatment of electron correlation, must be examined for all the infrared frequencies and intensities. Multivariate statistical optimization procedures, which are becoming increasingly important in experimental investigations, could provide the foundations of a strategy for choosing wave function characteristics resulting in generally more accurate frequency and intensity values. The use of these procedures furnishes a systematic scheme for identifying the main difficulties involved in attempting to obtain accurate values for all the vibrational parameters of a given molecule.

In this work two multivariate statistical techniques, factorial design[5] and principal component analysis,[6] are applied to the wave function optimization problem. Two-level factorial designs are used first to examine how wave function modifications affect the calculated infrared frequencies and intensities. Within the factorial scheme, the effect of each type of wave function modification on the calculated values of all the frequencies and intensities can be quantitatively assessed. As a result, those modifications that are important for the accurate calculation of the frequencies and intensities of a given structural group can be identified. The other multivariate technique, principal component analysis (PCA), provides a statistical criterion for choosing, out of a group of trial functions, the one that best reproduces all the infrared frequency and intensity values. The PCA results are conveniently represented in two-dimensional graphs that can be used to evaluate the overall ability of each wave function in reproducing the experimental results. The information obtained from the two kinds of statistical analysis complement each other, as we shall see.

Our main objective is to investigate how multivariate statistical methods can be used to provide more accurate wave functions for chemical problems dealing with a relatively large number of molecular properties. Although the present work is only concerned with spectral data, the methods employed here are generally applicable to all the properties normally obtained from molecular wave functions. To illustrate the use of these multivariate techniques, we selected the vibrational frequencies and infrared intensities of methyl fluoride. These data have been chosen for two reasons. First, they are well known and have been subjected to extensive experimental[7-11] and theoretical[12-14] investigations. Second, the fluorine atomic polar tensor in methyl fluoride is important for empirical intensity calculations for a variety of fluorine-containing molecules.[15-17]

Sosa and Schlegel[12] have calculated all the fundamental vibrational frequencies and infrared intensities of methyl fluoride using some 30 different wave functions, in an attempt to evaluate the importance of correcting for electron correlation in the calculation of these quantities, as well as to estimate the effects of including multiple sets of polarization and diffuse functions in the atomic basis sets. Their work showed that treatment of electron correlation at the second-order Møller-Plesset (MP2) level results in overall improvement in the agreement between the calculated vibrational frequencies and the harmonic frequencies determined from the observed vibrational bands. Infrared intensity values were shown to be more sensitive than the frequencies to the size and the nature of the basis set employed.

Calculations

Ab initio calculations of the CH_3F fundamental vibrational frequencies and infrared intensities were performed as prescribed by the 2^4 factorial design shown in Table I. Four wave function characteristics (or factors, in statistical terminology) were investigated at two levels: (1) the use of a 6-31G or a 6-311G wave function, (2) the presence or absence of polarization orbitals in the basis set, (3) the presence or absence of diffuse orbitals, and (4) the use (or not) of second-order Møller-Plesset perturbation corrections to the Hartree-Fock level calculations. The 16 wave functions obtained by all combinations of the two levels of these four factors were used to calculate the CH_3F vibrational frequencies and infrared intensities.

The main effect of a given factor on the calculated parameters is defined by

$$(\varepsilon f)_i = (\bar{R}_+)_i - (\bar{R}_-)_i \qquad (1)$$

where $(\varepsilon f)_i$ is the effect of the i^{th} factor and $(\bar{R}_+)_i$ and $(\bar{R}_-)_i$ are the average results (i.e., calculated frequencies or intensities for a given vibrational mode) at the high (+) and low (−) levels of this factor. Interaction effects of two or more factors are calculated using the same equation, except that the (+) and (−) levels are determined by multiplying the signs in the columns of the factors involved in the interaction.[5]

This simple equation can be used to calculate main and interaction effects because the factorial design in Table I is orthogonal. Each level of each factor is included in half of the 16 calculations performed. The quantities $\bar{R}_+$ and $\bar{R}_-$ are averages of eight values in all cases. For example, the effect, on a given frequency or intensity, of including polarization orbitals in the basis set is given by the difference between the average frequency or intensity values calculated from the eight wave functions containing polarization orbitals in their basis sets and the average of the values calculated without these orbitals. Since all possible combinations of the high and low levels of the other three factors are present in both $\bar{R}_+$ and $\bar{R}_-$, this difference represents the average change on the calculated frequency or intensity produced by the inclusion of polarization orbitals. Overall, four main effects (one for each factor) plus six second-order, four third-order, and one fourth-order interaction effects can be calculated for each of the CH_3F spectral parameters, using the 16 MO calculations listed in Table I.

Principal components, which were calculated separately for the vibrational frequencies and intensities, are the eigenvectors of the matrix product $\mathbf{X}^t\mathbf{X}$ where $\mathbf{X}^t$ is the transpose of the $\mathbf{X}$ matrix. The $\mathbf{X}$ matrix has 16 rows, corresponding to the 16 wave functions in the factorial design of Table I,

TABLE I.
A 2^4 Factorial Design for the Calculation of the CH_3F Vibrational Frequencies and Intensities.

Factors	Levels −	Levels +
1. Basis set	6-31G	6-311G
2. Polarization functions	absent	present
3. Diffuse functions	absent	present
4. Electron correlation	Hartree-Fock	Møller-Plesset 2

	Wave Function	Factorial designation			
1	HF/6-31G	−	−	−	−
2	HF/6-311G	+	−	−	−
3	HF/6-31G**	−	+	−	−
4	HF/6-311G**	+	+	−	−
5	HF/6-31 + + G	−	−	+	−
6	HF/6-311 + + G	+	−	+	−
7	HF/6-31 + + G**	−	+	+	−
8	HF/6-311 + + G**	+	+	+	−
9	MP2/6-31G	−	−	−	+
10	MP2/6-311G	+	−	−	+
11	MP2/6-31G**	−	+	−	+
12	MP2/6-311G**	+	+	−	+
13	MP2/6-31 + + G	−	−	+	+
14	MP2/6-311 + + G	+	−	+	+
15	MP2/6-3 1 + + G**	−	+	+	+
16	MP2/6-3 1 1 + + G**	+	+	+	+

and six columns, one for each of the six fundamental frequencies or infrared intensities. The original data matrices were preprocessed before the eigenvector calculations. The intensity values were centered on their respective means, while the frequency values were autoscaled.[7] The resulting matrices can be represented in a multidimensional space, as shown in Figure 1. Each coordinate axis represents one of the preprocessed frequencies (or intensities), and the results of the molecular orbital calculations correspond to points in this space. The doubly degenerate *E* vibrational modes are represented by only three columns, because the three additional columns needed to express this degeneracy would be identical to three columns already present in the **X** matrix.

The first eigenvalue of the matrix product $\mathbf{X}^t\mathbf{X}$ is equal to the amount of statistical variance explained by the first eigenvector.[6] This eigenvector, which defines the first principal component axis, points in the direction of maximum statistical variance, as indicated in Figure 1. The second eigenvector (the second principal component) is perpendicular to the first one and explains a maximum amount of the residual variance in the data–that is, variance not explained by the first eigenvector. If the first two eigenvectors explain a significant amount of the total variance, a principal component score plot in which they are the coordinate axes provides a faithful two-dimensional projection of the six-dimensional frequency or intensity space. In such situations, two-dimensional plots can be used to assess the quality of wave functions for calculating accurate frequency and intensity values.

The factorial design and principal component calculations were carried out using computer programs developed in our laboratories.[18] The *ab initio* molecular orbital calculations were performed with the Gaussian 92 computer package[19] on an IBM RISC 6000 workstation. The frequencies and intensities were calculated using optimized equilibrium geometries for each wave function.

Factorial Effects on the Infrared Frequencies and Intensities

The CH_3F infrared frequency and intensity values obtained from the 16 molecular orbital calculations specified by the 2^4 factorial design are presented in Tables II and III, respectively. The experimental infrared intensities[8,9] and the estimated harmonic frequencies[11] are included for comparison. The most important main and interaction effects of the wave function modifications on the calculated frequencies are reported in Table IV. Effects with absolute values smaller than 10 cm^{-1} are not included in this table.

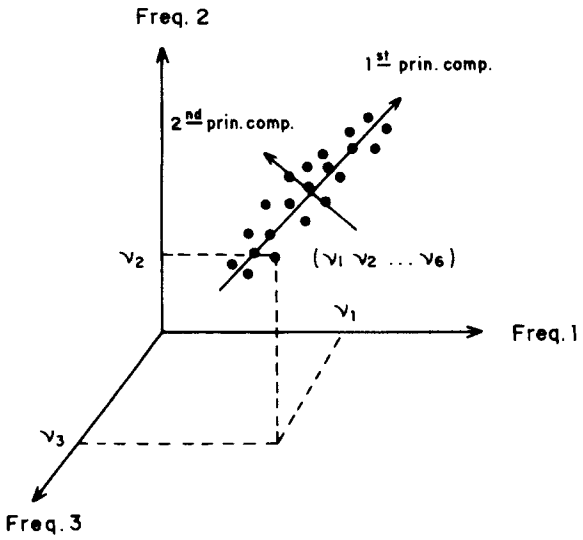

FIGURE 1. A three-dimensional representation of the six-dimensional frequency space. The six-dimensional intensity space has coordinate axes specifying the intensity values rather than the frequency values, as in this figure.

The inclusion of second-order Møller-Plesset perturbation (MP2) is the most important of the four wave function modifications. This procedure lowers the calculated frequencies by 90 to 120 cm^{-1}, as can be confirmed easily by inspection of the values in Table II. For any of the frequencies, a comparison between results of calculations differing only in the fourth sign—that is, (− − − −) and (− − − +), (+ − − −) and (+ − − +), etc.—shows differences of about 100 cm^{-1}, with the MP2 results always lower than the corresponding Hartree-Fock values. The importance of applying the Møller-Plesset procedure to wave functions used to calculate frequency values has already been pointed out by several researchers.[1-4] For methyl fluoride, Sosa and Schlegel[12] have shown that the MP2/6-311G(*d*, *p*) frequencies are, on average, about 50 cm^{-1} lower than those calculated with a HF/6-311G(*d*, *p*) function. Their results differ from ours because they used a fixed equilibrium geometry calculated from a CISD/6-31G(*d*) for both the Hartree-Fock and the Møller-Plesset calculations.

TABLE II.
Calculated and Experimental Fundamental Vibrational Frequencies for CH_3F (cm^{-1}).

Wave Function				ν_1 A_1 CH_3 str.	ν_2 A_1 CH_3 bend	ν_3 A_1 CF str.	ν_4 E, CH_3 str.	ν_5 E, CH_3 def.	ν_6 E, CH_3 def.
−	−	−	−	3245.6	1633.4	1097.4	3344.5	1659.2	1267.2
+	−	−	−	3212.4	1621.4	1080.8	3309.0	1644.7	1264.0
−	+	−	−	3203.5	1637.0	1186.1	3286.3	1633.0	1307.2
+	+	−	−	3190.3	1625.2	1169.9	3271.6	1613.7	1301.1
−	−	+	−	3251.7	1608.6	1047.2	3359.5	1644.7	1254.3
+	−	+	−	3209.4	1602.3	1049.2	3310.8	1638.5	1257.2
−	+	+	−	3214.9	1618.2	1153.6	3304.3	1623.8	1298.7
+	+	+	−	3193.5	1610.9	1156.6	3276.9	1614.0	1295.5
−	−	−	+	3096.2	1534.6	996.2	3202.2	1569.6	1172.4
+	−	−	+	3049.3	1522.6	975.7	3162.8	1549.0	1161.8
−	+	−	+	3132.8	1554.5	1112.5	3237.5	1563.1	1227.0
+	+	−	+	3087.6	1536.6	1105.6	3186.3	1519.1	1224.0
−	−	+	+	3105.0	1498.4	923.9	3225.6	1548.3	1147.8
+	−	+	+	3050.4	1495.4	924.4	3171.9	1541.9	1144.6
−	+	+	+	3143.5	1525.3	1056.7	3259.6	1548.5	1211.8
+	+	+	+	3093.5	1516.5	1075.7	3197.6	1520.6	1215.3
	Expt.[a]			3031.2	1490.2	1059.2	3131.5	1497.8	1206.4

Sign combinations as in Table I.
[a] Estimated harmonic frequencies obtained from observed anharmonic frequencies. See ref. 7.

TABLE III.
Calculated and Experimental Fundamental Infrared Intensities for CH_3F (km mol^{-1}).

Wave Function				A_1 CH_3 str.	A_2 CH_3 bend	A_3 CF str.	A_4 CH_3 str.	A_5 CH_3 def.	A_6 CH_3 def.
−	−	−	−	18.9	5.1	95.3	76.4	8.2	2.6
+	−	−	−	23.9	6.2	108.8	85.6	10.6	4.4
−	+	−	−	28.3	10.9	126.5	85.0	2.2	6.2
+	+	−	−	31.4	7.9	140.0	113.8	6.0	6.8
−	−	+	−	22.7	3.0	118.2	52.6	14.6	3.6
+	−	+	−	26.5	4.1	121.9	69.0	14.0	4.2
−	+	+	−	38.4	6.5	153.3	90.0	6.8	6.6
+	+	+	−	37.2	5.7	157.5	95.0	8.8	6.8
−	−	−	+	18.6	5.2	53.3	73.4	5.2	0.8
+	−	−	+	20.9	5.0	61.9	69.8	7.8	1.8
−	+	−	+	31.4	10.5	85.3	113.8	1.8	2.8
+	+	−	+	28.3	6.7	94.1	82.0	5.2	3.4
−	−	+	+	20.7	2.5	75.4	42.4	10.8	1.2
+	−	+	+	22.9	3.1	78.2	48.8	11.0	1.6
−	+	+	+	32.8	5.1	111.2	54.2	6.4	2.8
+	+	+	+	32.6	4.3	114.0	60.4	8.0	3.0
	Expt.[a]			24.7	0.9	95.3	61.0	8.7	2.7

Sign combinations as in Table I.
[a] Average values of the measured intensities reported in refs. 4 and 5.

TABLE IV.
Main and Interaction Effects of Wave Function Modifications on the Fundamental CH_3F Vibrational Frequencies (cm^{-1}).[a]

Effects	ν_1 CH_3 sym. str.	ν_2 CH_3 sym. bend	ν_3 CF str.	ν_4 CH_3 asym str.	ν_5 CH_3 asym bend	ν_6 CH_3 rock
Basis set	−38.4	—	—	−41.6	−18.6	—
Polarization functions	—	13.4	115.2	—	−20.0	51.4
Diffuse functions	—	−23.8	−42.2	13.2	—	−12.4
MP2 correlation	−120.4	−96.6	−96.2	−102.4	−89.0	−92.6
Correlation–polarization	34.2	—	17.3	38.0	—	11.5

[a] Effects with absolute values less than 10 cm^{-1} are not included in this table.

The other three wave function modifications are also important for accurate frequency estimates, although their effects are less significant and vary noticeably from one vibrational mode to another. For example, the change from a 6-31G to a 6-311G basis set reduces the CH_3 symmetric and asymmetric stretches by about 40 cm^{-1}, on average, but lowers the CH_3 asymmetric bending by only 19 cm^{-1} and leaves the other modes practically unaffected.

The main effects of including polarization and diffuse functions have opposite signs for the ν_2, ν_3, and ν_6 vibrational modes. The inclusion of polarization functions, which correspond to positive effects, tends to increase the frequencies of these modes, especially the CF stretching frequency, while the inclusion of diffuse functions tends to decrease them. The CH_3 stretching modes, on the other hand, are relatively insensitive to these effects, except for the asymmetric mode, whose frequency is slightly raised by the inclusion of diffuse functions.

The only significant interaction effect involves the polarization orbitals and the second-order Møller-Plesset perturbation procedure. This effect can best be interpreted with the diagrams presented in Figure 2. In these diagrams the four possible combinations of the two levels of the modifications (including or not polarization functions and applying or not the MP2 correction) are located at the corners of a square. The numerical values in each corner are the average frequencies corresponding to the respective level combinations. Differences between these values taken along one side of the square represent the effect corresponding to a particular wave function modification. As shown in Table IV, both CH_3 stretching modes have positive correlation–polarization interaction effects of about 35 cm^{-1}. They also exhibit very large correlation main effects, but no significant polarization main effects. The diagrams in Figure 2 help explain why this is so. When the MP2 correction is introduced, there is a lowering of both stretching frequencies, regardless of the inclusion of polarization functions in the basis set. When these functions are present, however, the Møller-Plesset main effect is approximately halved, falling from −154.7 cm^{-1} to −88.5 cm^{-1} in the symmetric mode and from −140.4 cm^{-1} to −64.5 cm^{-1} in the asymmetric mode, as a consequence of the strong interaction between the two wave function modifications. The effect of introducing polarization functions is also significant for a given level of the MP2 factor. However, this effect is positive when the MP2 correction is applied and negative when it is not. Since the magnitude of

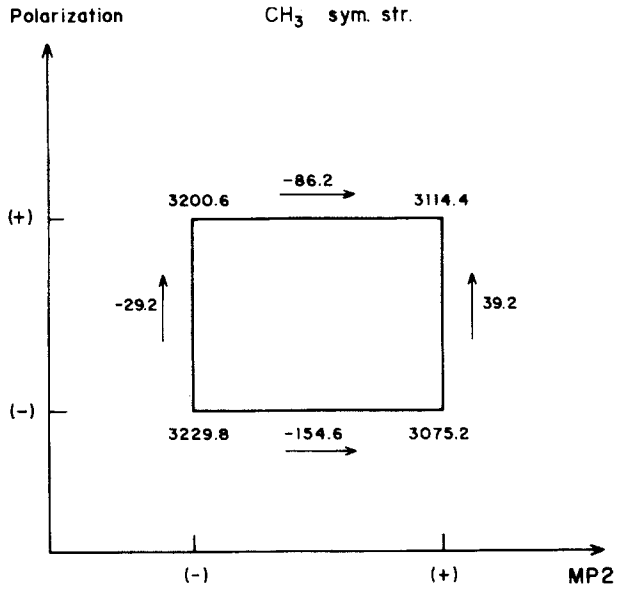

FIGURE 2. Analysis of the MP2 and polarization main and interaction effects on the calculated frequencies of CH_3F. (a) CH_3 symmetric stretching.

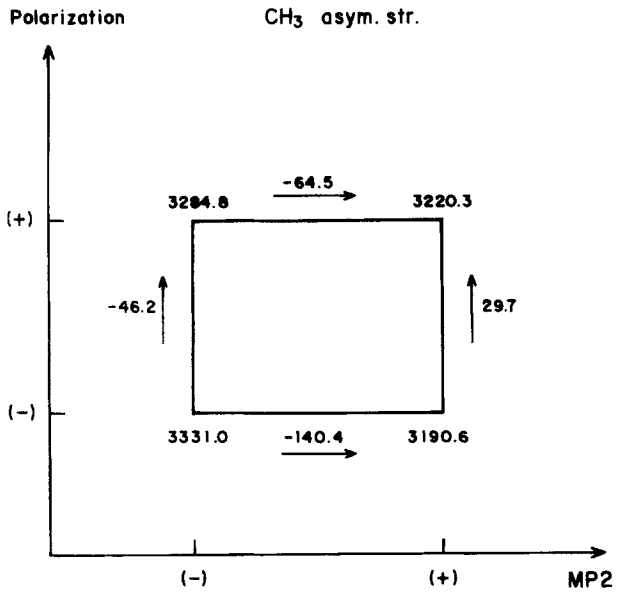

FIGURE 2. —(*Continued*). (b) CH_3 asymmetric stretching.

these two effects is more or less the same, they cancel each other out in the overall calculation, resulting in a nonsignificant polarization main effect, as indicated in Table IV. This could be interpreted as meaning that the inclusion of polarization functions has no effect on the calculated frequencies. The significant value observed for the polarization–correlation interaction effect acts as a warning against such an interpretation.

The effects of the wave function modifications on the infrared intensity values are more varied than those observed for the frequencies. As can be seen in Table V, the MP2 correction has negative main effects on all intensities. For the CH_3 asymmetric stretch and the CF stretch, they are significant: -23.2 km mol^{-1} and -43.6 km mol^{-1}. These are large values indeed, considering that the experimental intensities measured for these vibrations are 61.0 and 95.0 km mol^{-1}. For the CH_3 rock, the MP2 main effect is only -3 km mol^{-1}, but this band is very weak and has an experimental intensity value even smaller, 2.7 km mol^{-1}. A similar MP2 effect is calculated for the CH_3 symmetric stretching that is nearly 10 times stronger. This is in contrast with the regularity of the Møller-Plesset effects calculated for the vibrational frequencies, which were relatively constant, varying from 3 to 10% of the harmonic frequency values.

Substituting the 6-311G for the 6-31G basis has comparatively small effects on the intensity values, but the inclusion of polarization orbitals produces significant changes. The CH_3 asymmetric bending intensity is lowered by 4.6 km mol^{-1}, whereas all the other intensities are increased by values ranging from 2.2 km mol^{-1} for the CH_3 rock to 33.6 km mol^{-1} for the CF stretch. The inclusion of diffuse functions has opposite effects on the CH_3 asymmetric stretch and the CF stretch, lowering the former intensity by 24.2 km mol^{-1} and raising the latter by 20.6 km mol^{-1}. Again, most of these effects correspond to substantial contributions to the calculated intensity values. The only significant interaction occurs for the CH_3 asymmetric stretch and involves the correlation and polarization main effects, as in the frequency analysis.

The values of the effects in Tables IV and V help explain why accurate intensities are more difficult to calculate than accurate frequencies. Not only are the effects on the intensities, relative to their experimental values, much larger than those for the frequencies, but there is more variation from one vibrational mode to another. For example, while the MP2 correction results in calculated frequencies in better agreement with the harmonic frequencies for all modes, for the intensities the same effect leads to significant improvement in the A_3,

TABLE V.
Main and Interaction Effects of Wave Function Modifications on the Fundamental CH_3F Infrared Intensities (km mol^{-1}).[a]

Effects	A_1 CH_3 sym. str.	A_2 CH_3 sym. bend	A_3 CF str.	A_4 CH_3 asym str.	A_5 CH_3 asym bend	A_6 CH_3 rock
Basis set	—	—	7.2	—	1.9	—
Polarization functions	10.6	3.0	33.6	22.8	−4.6	2.2
Diffuse functions	4.0	−2.8	20.6	−24.2	4.2	—
MP2 correlation	−3.2	—	−43.6	−23.2	−1.9	−3.0
Correlation–polarization	—	—	—	−11.0	—	—

[a] Effects with absolute values less than 1 km mol^{-1} are not included in this table.

A_4, and A_6 values only, and yields slightly worse values for the A_5 band.

The existence of interaction effects can explain deviations from the additive models that are sometimes used to estimate molecular properties for complex wave functions based on results obtained from simpler wave functions.[20,21] One such example is the calculation done by Sosa and Schlegel[12] to estimate frequency and intensity values for CH_3F at the MP2/6-311 + + G(3*d*,3*p*) level. Their procedure is conceptually equivalent to adding only the main effect values to the HF/6-31G wave function results. If significant interaction effects are present, one expects to find discrepancies between the results given by the additive model and those obtained from the full molecular orbital calculation.

In Table VI estimated MP2/6-311 + + G(*d*, *p*) frequency and intensity values calculated using the HF/6-31G results in Tables II and III plus the main effects in Tables IV and V are compared with the MP2/6-311 + + G(*d*, *p*) and experimental values. All the estimated frequency values are within 10 cm^{-1} of the values calculated from the full MP2/6-311 + + G(*d*, *p*) wave function except ν_4 and ν_5, for which the estimated values are 16 and 11 cm^{-1} larger than the corresponding molecular orbital results. In general, the estimated values are almost in as good an agreement with the experimental results as are the results calculated from the MP2/6-311 + + G(*d*, *p*) wave function. A similar situation holds for the intensities, although in this case the agreement of both the estimated and the molecular orbital values with experiment is less impressive. One may conclude, then, that for methyl fluoride an additive model can be successfully employed to estimate results from more sophisticated quantum chemical calculations and the savings in computer time compensate for the small loss of accuracy in the estimated results.

Principal Component Results

The first two principal component equations for the calculated frequencies and intensities are presented in Table VII. For the frequencies, they explain 97.1% of the total variance in the original data. Since the remaining four components together account for only 2.9% of this variance, the corresponding equations are not shown in Table VII. The high amount of variance described by the first two components implies that a plot having them as coordinate axes provides an accurate projection of the six-dimensional frequency data. In other words, the swarm of data points in the original multidimensional frequency space, each point locating the six calculated frequencies for each of the 16 wave functions in Table I, has an almost exactly planar structure, with the plane

TABLE VI.
Comparison of MP2 / 6-311 + + G(*d*, *p*) and Experimental Results with Estimates Calculated Using the HF / 6-31G Results and the Factorial Design Main Effect Values.

	HF / 6-31G[a]	MP2 / 6-311 + + G(*d*, *p*)[b] Estimated	MP2 / 6-311 + + G(*d*, *p*)[c]	Expt.
ν_1	3245.6	3086.8	3093.5	3031.2
ν_2	1633.4	1526.4	1516.5	1490.2
ν_3	1097.4	1074.2	1075.2	1059.2
ν_4	3344.5	3213.7	3197.6	3131.5
ν_5	1659.2	1531.6	1520.6	1497.8
ν_6	1267.2	1213.6	1215.3	1206.4
A_1	18.9	30.3	32.6	24.7
A_2	5.1	5.3	4.3	0.9
A_3	95.3	113.1	114.0	95.3
A_4	76.4	51.8	60.4	61.0
A_5	8.2	9.7	8.0	8.7
A_6	2.6	1.8	3.0	2.7

[a] Calculated values using a HF / 6-31G wave function.
[b] Estimated values using the HF / 6-31G result and the principal effect values in Tables IV and V.
[c] Calculated values using the MP2 / 6-311 + + G(*d*, *p*) wave function.

TABLE VII.
Principal Component Equations for the Calculated Frequencies and Intensities.[a]

Frequencies	Explained Variance
$PC_1 = 0.432\ \nu_1 + 0.437\ \nu_2 + 0.340\ \nu_3 + 0.407\ \nu_4 + 0.407\ \nu_5 + 0.418\ \nu_6$	83.5%
$PC_2 = -0.234\ \nu_1 + 0.712\ \nu_3 - 0.386\ \nu_4 - 0.385\ \nu_5 + 0.374\ \nu_6$	13.6%
Intensities	**Explained Variance**
$PC_1 \cong 0.904\ A_3 + 0.393\ A_4$	74.0%
$PC_2 \cong -0.397\ A_3 + 0.904\ A_4$	24.3%

[a] Equations for the third through sixth principal components are not included because they explain insignificant amounts of the total variance.

being defined by the first two principal components. The distribution of the frequency points in this bidimensional projection is presented in Figure 3.

The first principal component, which corresponds to the abscissa in Figure 3, is essentially an average of all six frequency values. The MP2 results occupy the left portion of the plot, whereas the Hartree-Fock values are on the right. This separation could have been anticipated, since the first principal component accounts for most of the statistical variance and the MP2 effect values in Table IV have large magnitudes for all the frequencies. The Møller-Plesset results are located on the left side of the plot because all the MP2 effects are negative, while the PC_1 coefficients are all positive.

The second principal component discriminates between frequencies obtained from wave functions with polarization orbitals and frequencies calculated without these orbitals. Large positive polarization effects were determined for the ν_3 and ν_6

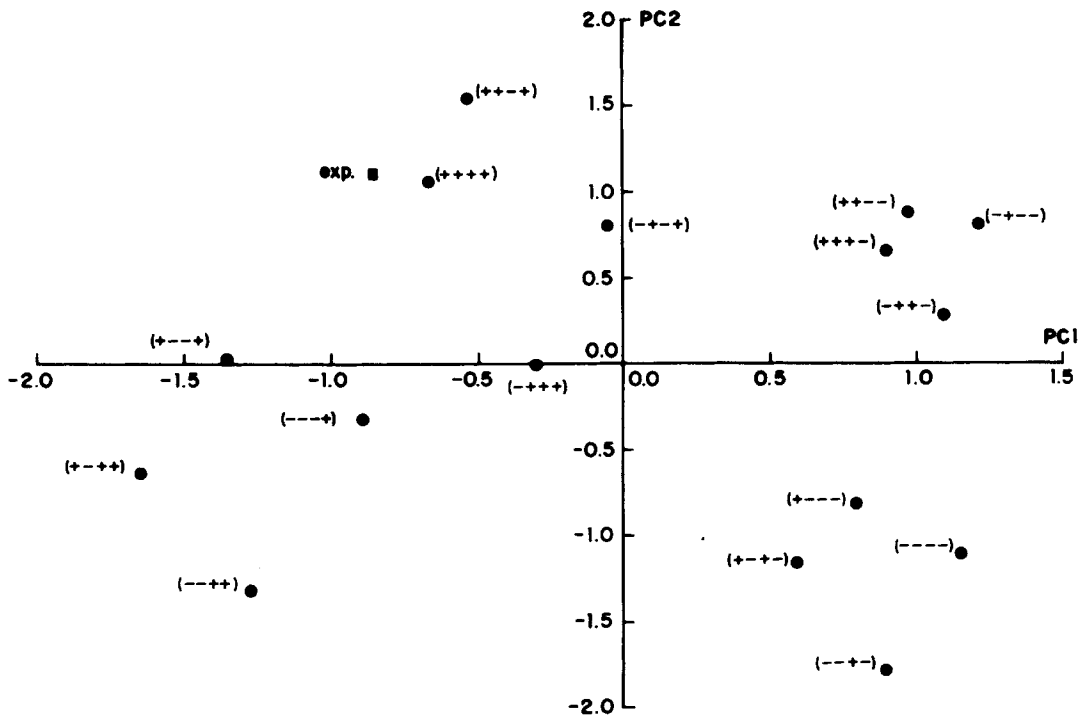

FIGURE 3. Principal component plot of the calculated and experimental frequencies of a CH_3F.

frequencies. Since the ν_3 and ν_6 terms in the second principal component have positive coefficients, the results calculated with polarization orbitals have larger PC_2 scores than those calculated without polarization orbitals. They are therefore located in the upper part of the plot.

The coordinates of the point corresponding to the experimental values are obtained by simply substituting the autoscaled harmonic frequency values in the principal component equations. Since this point is in closest proximity to the (+ + + +) point, the experimental values are in best agreement with the frequencies obtained with the MP2/6-311 + + G(d, p) wave function. This is confirmed by the values in Table II. The four points closest to the experimental one all correspond to results calculated at the MP2 level with wave functions containing polarization orbitals. Of the four factors studied here, these two are thus the most important to obtain better overall agreement between the calculated and the experimental frequencies of CH_3F.

The six-dimensional intensity data can also be accurately represented by a bidimensional principal component projection, which explains 98.3% of the total data variance. The plot of the scores on the first two principal components, shown in Figure 4, is an accurate projection of the almost planar distribution of the intensity data points in the original multidimensional intensity space.

The two principal component equations for the intensities contain significant contributions from the A_3 and A_4 intensities only. All other terms have coefficients with absolute values less than 0.15 and are not shown in Table VII. This can be easily understood by examining the calculated intensity values in Table III. Since the A_3 and A_4 values have much larger statistical variances than the calculated intensities for the other vibrational modes, their contribution to the first component should be correspondingly dominant. The first principal component (PC_1) accounts for 74% of the variance and is a linear combination of A_3 and A_4, with a larger contribution from A_3. The results for wave functions at the MP2 level occupy the left-hand portion of the plot and the Hartree-Fock results appear toward the right, although this separation is not as clearcut as for the frequencies. This geometrical arrangement is consistent with the factorial design results, indicating that the MP2 correction is responsible for the largest of the four main effects calculated. Since the MP2 treatment lowers the A_3 intensities and A_3 enters into the PC_1 equation with a positive sign, the right–left

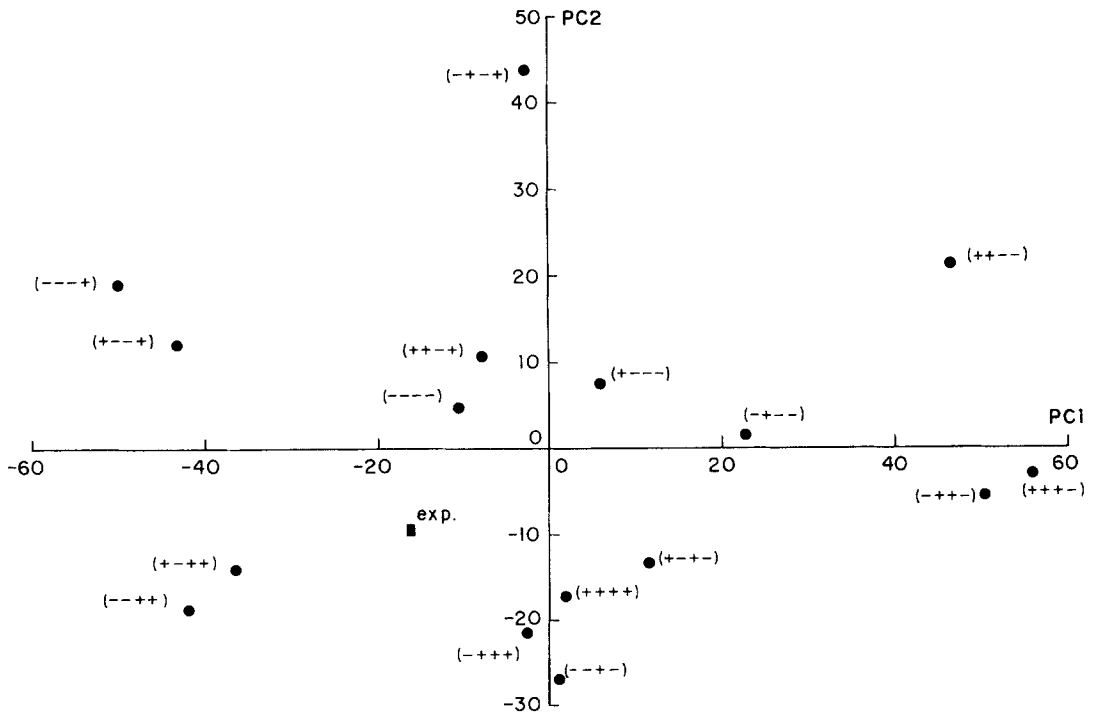

FIGURE 4. Principal component plot of the calculated and experimental intensities of CH_3F.

distribution of the data points could have been anticipated. The other term also tends to enhance this horizontal distribution, due to the negative MP2 effect on A_4.

The separation, along the PC_1 axis, of results calculated with and without polarization orbitals is less well defined. Intensities calculated using polarization orbitals are situated more to the right of the plot, because both A_3 and A_4 have positive polarization factorial effects. This separation, however, is evident only for points on the extreme left and right sides of the graph. The middle portion contains results from both types of calculation.

The second principal component explains 24.3% of the variance and is also a linear combination of A_3 and A_4 (the A_4 intensity is the dominant contribution). The largest factorial effect for A_4 is a negative one for the diffuse functions, while the A_3 intensities have a significant positive diffuse function effect. Since the coefficients of A_3 and A_4 in the second principal component have opposite signs, these effects reinforce each other and displace toward the lower part of the graph the points representing calculations with diffuse functions.

The point representing the experimental intensity values is also shown in Figure 4. Its location is easily calculated by substituting the experimental A_3 and A_4 intensity values of Table III (mean-centered) in the second set of principal component equations in Table VII. The theoretical point closest to the experimental one is identified by the $(- - - -)$ sign combination, which surprisingly corresponds to a simple HF/6-31G calculation. The intensities obtained from this wave function present the smallest sum of squares deviation from the experimental values, as can be verified using the values in Table III. The $(- + + +)$ and $(+ + + +)$ points, corresponding to the MP2/6-31 + + G(d, p) and MP2/6-311 + + G(d, p) wave functions, are the next closest to the experimental point. These results contrast with those derived from the PC frequency analysis. Whereas the experimental frequencies are in closest agreement with the values obtained in MP2 calculations with polarization orbitals in the basis set, the experimental intensities agree best with the results of very diverse calculations, the HF/6-31G and MP2/6-311 + + G(d, p) functions. This clearly demonstrates that optimizing molecular wave functions to calculate accurate CH_3F intensities is more difficult than achieving an optimization aimed at accurate frequencies. A similar result can probably be expected for most other molecules, in view of the difficulties usually found in attempts to calculate accurate infrared intensities from molecular orbital wave functions.

Acknowledgments

The authors acknowledge partial financial support from the agencies Conselho Nacional de Pesquisa (CNP_q), Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP), and Fundação de Apoio à Ciencia de Estado de Pernambuco (FACEPE). A. E. O. is indebted to CNP_q for a fellowship.

References

1. J. F. Stanton, W. N. Lipscomb, D. H. Mayers, and R. J. Bartlett, *J. Chem. Phys.*, **90**, 3241 (1989).
2. Y. Yamaguchi, M. Frisch, J. F. Gaw, H. F. Schaefer, and S. Binkley, *J. Chem. Phys.*, **84**, 2262 (1986).
3. E. D. Simandilus, R. D. Amos, and N. C. Handy, *Chem. Phys.*, **114**, 9 (1987).
4. M. D. Miller, F. Jansen, O. L. Chapman, and K. N. Honk, *J. Phys. Chem.*, **93**, 4495 (1989).
5. G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
6. (a) K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979, chap. 8; (b) E. Suto, M. M. C. Ferreira, and R. E. Bruns, *J. Comp. Chem.*, **12**, 885 (1991).
7. G. M. Barrow and D. C. Mckean, *Proc. R. Soc. London Ser. A*, **213**, 27 (1952).
8. J. W. Russell, C. D. Needham, and J. Overend, *J. Chem. Phys.*, **45**, 3383 (1966).
9. S. Kondo and S. Saëki, *J. Chem. Phys.*, **76**, 809 (1982).
10. C. Dilauro and I. M. Mills, *J. Mol. Spectrosc.*, **21**, 386 (1966).
11. J. Duncan, D. C. Mckean, and G. K. Spiers, *Mol. Phys.*, **24**, 553 (1972).
12. C. Sosa and H. B. Schlegel, *J. Chem. Phys.*, **86**, 6937 (1987).
13. E. Suto, H. P. Martins, Jr., and R. E. Bruns, *J. Mol. Struct. (THEOCHEM)*, **282**, 81 (1993).
14. C. E. Blom and A. Muller, *J. Mol. Spectrosc.*, **70**, 449 (1978).
15. W. B. Person, In *Vibrational Intensities in Infrared and Raman Spectroscopy*, W. B. Person and G. Zerbi, Eds., Elsevier, Amsterdam, 1982, p. 271.
16. J. H. Newton and W. B. Person, *J. Chem. Phys.*, **64**, 3036 (1976).
17. W. B. Person and J. Overend, *J. Chem. Phys.*, **66**, 1442 (1977).
18. I. S. Scarminio and R. E. Bruns, *Trends Anal. Chem.*, **8** 326 (1989).
19. M. J. Frisch, J. S. Binkley, H. B. Schlegel, K. Raghavachavi, C. F. Melius, R. L. Martin, J. J. P. Stewart, F. W. Bobrowicz, C. M. Rohlfing, L. R. Kahn, D. J. Defrees, R. Saeger, R. A. Whiteside, D. J. Fox, E. M. Fleuder, and J. A. Pople, *Gaussian 92*, Revision C, Gaussian Inc., Pittsburgh, PA, 1992.
20. J. A. Pople, B. T. Luke, M. J. Frisch, and S. J. Binkley, *J. Phys. Chem.*, **89**, 2198 (1985).
21. J. A. Pople and L. A. Curtis, *J. Phys. Chem.*, **91**, 155 (1987).