Numerical Pattern Recognition Analysis of CO Atmospheric Simulation Experiments

Matthew P. Jacobson, Stephen L. Coy, and Robert W. Field*

Department of Chemistry and George R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Steven J. Lipson, Ronald B. Lockwood, David L. Vititoe, and William A. M. Blumberg

Air Force Research Laboratory, Space Vehicles Directorate, Hanscom AFB, Massachusetts 01731-3010

Peter S. Armstrong

Stewart Radiance Laboratory, 139 Great Road, Bedford, Massachusetts 01730

Received: July 20, 1999; In Final Form: October 20, 1999

A technique entitled Hybrid Linear Pattern Analysis (HLPA), which represents a combination of model-based and pattern recognition-based approaches to the analysis of spectroscopic data, is introduced and applied to the analysis of time-resolved infrared emission spectra of ground electronic state ($X^1\Sigma^+$) CO obtained in atmospheric simulation experiments. The spectra are highly congested and consist of incompletely resolved, overlapping $v' - v'' = 1$ emission bands from $v' = 1$ up to at least $v' = 12$. The analysis of the time dependence of the emission intensity in the various vibrational bands had been stymied by a severe optical opacity effect in the $v = 1 \rightarrow 0$ emission, which is difficult to simulate; thus, conventional least-squares fitting could not be used confidently to determine the time-dependent emission intensity of this band, or that of at least three other emission bands that overlap strongly with it. The HLPA technique permits an alternate approach in which the $v = 1 \rightarrow 0$ emission band is considered to be an unknown pattern that is identified by the Extended Cross Correlation (XCC) pattern recognition technique (*J. Chem. Phys.* **1997**, *107*, 8349). The intensity profiles of the other bands, however, can be predicted accurately based on the experimental parameters, and this knowledge is used in conjunction with the results of the XCC to determine the time dependence of *all* of the vibrational bands, and the intensity profile of the $v = 1 \rightarrow 0$ emission band.

I. Introduction

We introduce in this article a technique entitled Hybrid Linear Pattern Analysis (HLPA), which represents a combination of model-based and pattern recognition-based approaches to the analysis of spectroscopic data.

Consider a spectroscopic data set in which each spectrum is a linear superposition of a finite number of patterns. (These patterns might be associated with, for instance, different chemical species,[1,2] polyad quantum numbers,[3] or as in this work, different vibrational bands of a single species.) A "model-based" analysis of such a data set is possible if the patterns (relative intensity vs frequency) that are contained in the spectra can be predicted. The relative amplitude of each pattern in each spectrum can then be determined by conventional optimization procedures. Least-squares fitting algorithms[4] are by far the most commonly used for such optimizations, although robust methods of estimation[5−8] can reduce the sensitivity of the fit to outliers (which could be due to either experimental artifacts or deficiencies in the model).

If, on the other hand, it is not possible to predict the number or appearance of the patterns, then pattern-recognition techniques may provide a successful approach to the analysis of the data set. A wide variety of pattern-recognition techniques by now have been applied to spectroscopic data. Among the most common are techniques based on principal component analysis (PCA) (see, for instance, refs 9−12); other approaches include applications of neural networks,[13] genetic algorithms,[14] covariance mapping,[15] and the recently introduced Extended Cross Correlation (XCC) technique.[1,2] These various pattern-recognition approaches differ greatly in terms of their assumptions about the data to be analyzed, their realms of applicability, and the numerical algorithms used. However, most of these share the goal of identifying unknown patterns that are repeated either within one spectrum or among multiple spectra.

The case in which some of the patterns contained in a data set can be predicted a priori, but others cannot, has received less attention in the literature. One possible approach to this type of problem is to apply one of the pattern-recognition techniques discussed above in an attempt to identify all the patterns in the data set, including those that are previously known. However, it is of course advantageous to incorporate any knowledge of the patterns into the spectral analysis. The HLPA technique makes this possible. It uses techniques of pattern recognition to identify the unknown patterns but explicitly incorporates a priori knowledge of the remaining patterns.

In Section II, we illustrate the use of this technique with synthetic data, and in Section III with time-resolved emission spectra of CO that were recorded in atmospheric emission simulation experiments conducted at the LABCEDE facility at

* To whom correspondence should be addressed. E-mail: rwfield@mit.edu.

10.1021/jp9924732 CCC: $19.00
Published on Web 12/16/1999

the Air Force Research Laboratory, Hanscom AFB. Details of the experimental setup have been reported previously.[16,17] In brief, an electron beam is pulsed through a 10-mTorr sample of CO in a cryogenic chamber. Time-resolved infrared emission spectra of CO are recorded both during the 10-ms electron pulse and for 86 ms afterward. The time-resolved spectra provide information about the nascent populations and collisional deactivation rates of the excited vibrational levels of the ground state of CO. At least 12 excited vibrational levels of the electronic ground state of CO can be observed to be populated by processes initiated by the electron beam, and at the resolution of the experiment (~ 2.5 cm^{-1}), the overlapping emission bands are only partially resolved.

This data set has been partially analyzed previously with standard least-squares procedures. As discussed in Section III, the rotational distribution of the sample can be modeled to a good approximation by a Boltzmann distribution that is independent of time. Thus, vibrational emission band "basis functions" can be constructed (the emission frequencies are well-known from spectroscopic studies of CO), and least-squares algorithms in principle can be used to fit the relative contribution of each basis function to each spectrum and thereby to recover the intensity of emission from the various vibrational bands as a function of time.

However, this analysis has been stymied by an optical opacity effect in the $v = 1 \rightarrow 0$ emission band. That is, although the rotational distribution of the $v = 1$ vibrational level can be assumed to be described by the same Boltzmann distribution as the higher vibrational levels, the frequency dependence of the emission from this band *as seen at the detector* cannot be predicted easily because of strong self-absorption of the emission by ground vibrational state molecules, which are present at much higher concentration than all excited vibrational levels. The inability to construct a basis function for this $v = 1 \rightarrow 0$ emission band prohibits the use of least-squares fitting to determine the time dependence of the emission intensity *and that of every vibrational band that overlaps substantially with it*. Thus, before this work, only a limited analysis of this data set has been possible.

Here we report a complete analysis of the CO atmospheric simulation data using the HLPA technique. The optically thick $v = 1 \rightarrow 0$ emission band is considered to be a pattern that is repeated in more than 100 time-resolved spectra. As a first step in the HLPA technique, the XCC pattern-recognition technique is used to determine the time-dependent intensity of the $v = 1 \rightarrow 0$ emission with no knowledge of its band profile. The HLPA technique also permits a statistically rigorous determination of the time dependences of the remaining (known) emission bands, and the band profile of the optically thick emission.

II. Description of the HLPA Technique

In the CO atmospheric simulation experiments that are analyzed later in this article, the information that we wish to extract from the data is kinetic in nature: how do the emission intensities of various vibrational emission bands change with time? To illustrate how pattern recognition (in particular, XCC) and the HLPA technique can assist in this analysis, we define in Figure 1 a simulated data set that mimics several key properties of the real experimental data. Each synthetic spectrum is a linear superposition of the two "emission bands" in the top panel of Figure 1. That is, if we designate the spectra by numbers and the individual emission bands by letters, then

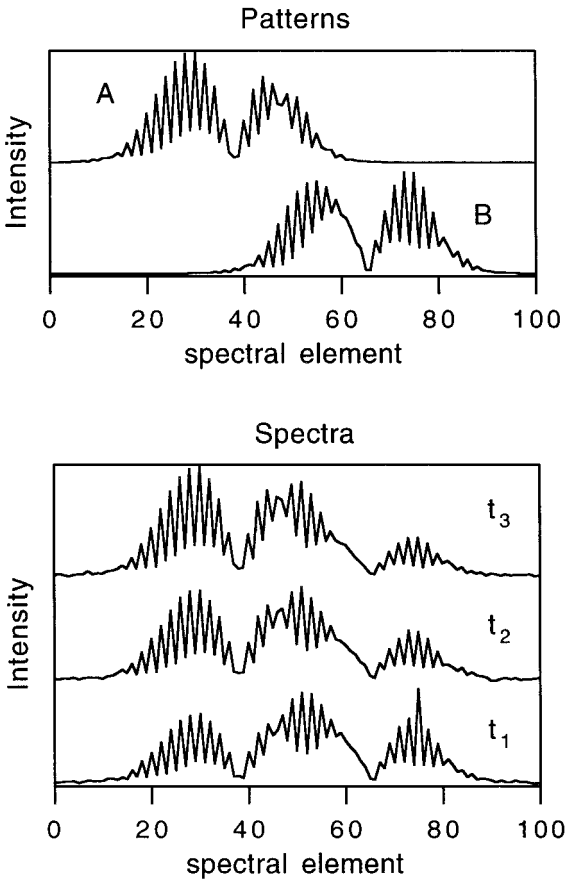

Figure 1. Simulated patterns (top panel) and spectra (bottom panel) for purposes of illustrating the HLPA technique. Each of the three spectra in the bottom panel are linear superpositions of the two patterns, plus noise.

$$I_1(\omega) = a_1 I_a(\omega) + b_1 I_b(\omega)$$
$$I_2(\omega) = a_2 I_a(\omega) + b_2 I_b(\omega) \quad (1)$$
$$I_3(\omega) = a_3 I_a(\omega) + b_3 I_b(\omega)$$

in which ω represents frequency (spectral element) and the coefficients $\{a\}$ and $\{b\}$ represent the relative amplitudes of the emission bands in each spectrum. Gaussian random noise has been superimposed on each of the spectra so that the signal-to-noise is approximately 100. In addition, in spectrum 1, we have increased the intensity of one of the spectral elements (number 75) by 50%. The purpose of this deliberate corruption of the data is to illustrate the way in which the XCC and HLPA methods deal with deviations from eq 1 that are neither small nor random; the ability of these techniques to identify patterns even in the presence of such "corruptions" of the data will be critical to the analysis of the CO data set in Section III.

The simulated spectra can thus be considered to represent three time-resolved emission spectra, and the goal of the analysis is to determine the time dependence of the two emission bands, as represented by the coefficients $\{a\}$ and $\{b\}$. Note that we are assuming here that the intensity vs frequency profiles of the bands do not change as a function of time; a similar assumption will be made in the analysis of the CO data set in Section III. If the intensity profiles of the emission bands were known a priori, then the task of obtaining the coefficients would be straightforward using standard linear least-squares fitting routines. However, in this article we are concerned with the case in which at least one of the two emission bands A and B is unknown. We consider first the case in which both patterns

are unknown, and demonstrate the way in which XCC can be used to determine the coefficients $\{a\}$ and $\{b\}$, as well as the frequency dependence of the emission bands $[I_a(\omega),I_b(\omega)]$, if so desired.

From the standpoint of the XCC, the emission bands constitute patterns that are repeated in three different spectra. The XCC is a model-free pattern-recognition technique, and in principle no prior knowledge of either the form or the number of patterns that are present in the data is necessary. However, there is one important condition for the success of the XCC: some portion of the features in each pattern must not be overlapped with any other pattern. There is no limit to the number of patterns that can be identified in a set of spectra, as long as each of the patterns conforms to this condition.

In the synthetic data in Figure 1, note that although the two patterns overlap heavily in the central portions of each spectrum, the intensity in the "wings" of the spectrum arises almost entirely from one of the two patterns. For example, consider spectral elements 0 through 30. To a good approximation, the intensity in each of the spectra over this range arises solely from vibrational band A:

$$I_k(\omega) \approx a_k I_a(\omega); \quad (k = 1, 2, 3)$$

Conversely, for spectral elements 70−100,

$$I_k(\omega) \approx b_k I_b(\omega); \quad (k = 1, 2, 3)$$

Thus, one could imagine determining the $\{a\}$ and $\{b\}$ coefficients by, for instance, simply integrating the intensities of the spectra over these spectral ranges. However, we are assuming that the patterns are unknown, which makes it difficult to judge, in the absence of some numerical tool, precisely which regions of the spectra contain contributions from just one pattern (the real experimental data analyzed in the next section are even more challenging).

The XCC incorporates principles of robust estimation[6] to identify, in a numerically rigorous and automated way, those regions of spectra that can be accounted for by just one pattern. The methods of robust estimation used by the XCC have been described in detail previously.[1,2] Here, we describe only the general strategy of the XCC, which is to repeat iteratively the following two steps: (1) Postulate that a pattern exists in the spectra with a given set of (relative) amplitudes in each of the spectra (the set of relative amplitudes is called a "ratio direction"). (2) Evaluate this postulate using a figure-of-merit function, which we will call G.

If the set of "guessed" ratio directions is close to that for a pattern that is actually present in the data, then G is large; otherwise, it is small. In practice, optimization routines are used to search the space of all possible ratio directions and identify all patterns present in a data set; strategies for performing this search in high dimensionality spaces (many spectra) are described in Ref 2. Note that the dimensionality of this search is $N - 1$, where N is the number of spectra. That is, only the relative amplitude of a pattern among several spectra is meaningful; the absolute amplitude of a pattern is arbitrary.

In the top panel of Figure 2 we illustrate the use of the XCC merit function G to identify the patterns that are present in spectra 1 and 3 of the synthetic data set. Because only two spectra are used for this illustration, the search for patterns occurs in a 1D space that represents the relative amplitude of each pattern in the two spectra (we choose, arbitrarily, I_3/I_1 to represent the ratio direction). Two pronounced maxima can be observed in the merit function, which occur at $I_3/I_1 = 1.604$ and 0.604; these correspond to vibrational bands A and B, respectively, and compare quite favorably with the values of 1.6 and 0.6, respectively, which were used to construct the synthetic spectra. The XCC can also be used to identify the patterns in all three of the spectra simultaneously (this is somewhat more difficult to illustrate in a single figure); the results are similarly excellent, with the relative pattern amplitudes (ratio directions) again determined by XCC to within 1% of the correct values.

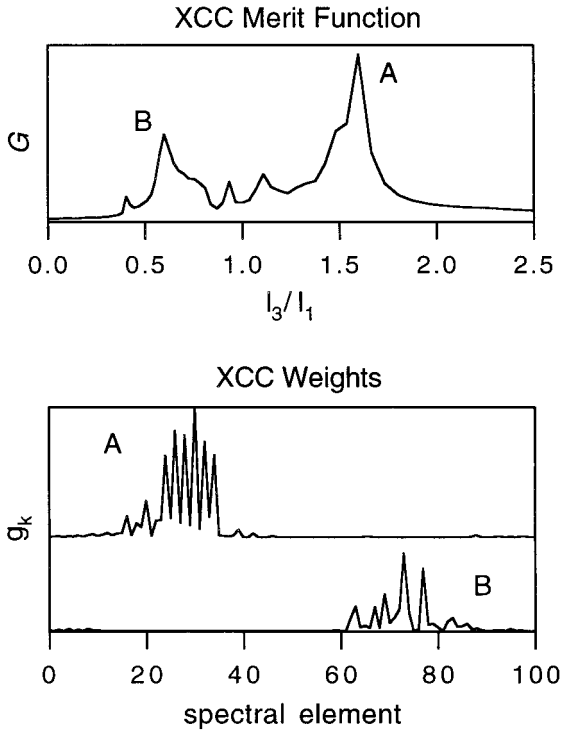

Figure 2. Application of the XCC technique to spectra 1 and 3 of Figure 1. Top panel: the XCC merit function G, as a function of I_3/I_1, which represents the relative amplitudes of a pattern in the two spectra. Bottom panel: weight functions, g_k, computed at the two largest maxima in the XCC and plotted as a function of spectral element. The weight functions indicate those portions of the spectra identified by the XCC as containing contributions from only one pattern.

The definition of the XCC merit function G also makes it easy to identify those portions of the spectra which represent fragments of the patterns that are uncorrupted by overlap with other patterns. Specifically, G is defined as a sum of weight functions (g) which are computed for each spectral element k:

$$G(\vec{\alpha}) = \sum_k g_k(\vec{\alpha})$$

where $\vec{\alpha}$ represents the ratio direction. If the relative intensities corresponding to the point k are close to the postulated ratio direction, then g_k is large. (The definition of g_k includes a Gaussian function of the deviation between the observed and postulated ratio directions.[1]) Thus, a maximum in G appears if there are many points with roughly the same relative intensities (i.e., regions of the spectra that can be accounted for by one pattern).

In the bottom panel of Figure 2 we plot the weight functions g_k at the ratio directions corresponding to the two largest maxima in the merit function G that is displayed in the upper panel. Note that neither set of weight functions includes any substantial amplitude over the central spectral elements (40−60) in which the two patterns overlap substantially. Note also that spectral element 75, which was deliberately corrupted in spectrum 1,

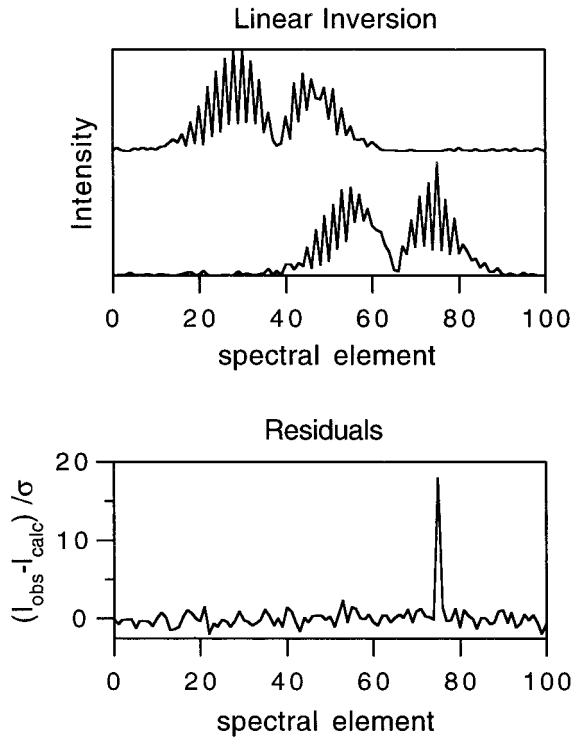

Figure 3. Top panel: results of the inversion of the patterns (vibrational emission bands) from the spectra using the coefficients determined by XCC. Bottom panel: weighted residuals of the inversion procedure for spectrum 1. The residuals are small and random (i.e., comparable with the noise amplitude) over all the spectrum except at spectral element 75, which was corrupted deliberately.

has a weight of nearly zero in both traces; the XCC has automatically excluded this point from its determination of the time dependence of the pattern amplitudes. This insensitivity of the XCC to nonidealities in the data and its ability to identify multiple patterns simultaneously are consequences of its definition as a redescending *robust* estimator.[6]

At this point we have succeeded in using the XCC to determine the time dependence of the emission from the two different vibrational bands with no knowledge of what the vibrational bands look like. We now demonstrate a straightforward extension of the XCC that permits the determination of the band profiles of the two vibrational bands from the time-dependent emission intensity. Specifically, because the coefficients $\{a\}$ and $\{b\}$ are now known from the application of the XCC, the set of equations labeled as eq 1 above is overdetermined. That is, at any given value of ω, there exist three equations with only two unknowns ($I_a(\omega)$ and $I_b(\omega)$), and the spectra can be inverted (in a least-squares sense) to determine the patterns, one spectral element at a time (note that this technique of inversion from spectra to patterns is applicable whenever the number of spectra is greater than or equal to the number of patterns).

Figure 3 depicts the result of this inversion process. The emission band patterns (top panel) that are inverted from the spectra are nearly identical with the patterns in Figure 1 that were used to construct the synthetic spectra originally. One notable discrepancy is observable at spectral element 75; the intensity of this spectral element in the reconstructed emission band B differs significantly from its true value observed in Figure 1. This discrepancy, of course, is caused by the deliberate corruption of spectrum 1 that we performed at spectral element 75. We have noted previously that the XCC was *insensitive* to this corruption of the data in its determination of the ratio directions (time dependence of the emission bands). It may seem paradoxical that the inversion from spectra to patterns is sensitive to the corruption of the data, but it should be kept in mind that the inversion is mathematically equivalent to a linear least-squares fit, which implicitly assumes that any deviations from the model conform to a Gaussian distribution; clearly this is not true for spectral element 75. However, we can determine from the inversion itself that spectral element 75 should be treated with suspicion. In the bottom panel of Figure 3 are plotted the residuals of the fit for spectrum 1 as a function of spectral element. The residuals are mostly comparable to the noise level in the spectra, except at spectral element 75, indicating that the model (eq 1) cannot accurately represent the data at this spectral element.

Thus, the pattern-recognition approach to the synthetic "time-resolved emission spectra" proceeds by first identifying the number of patterns (vibrational bands) present in the data set along with their time-dependent amplitudes, and then using this time dependence to extract, if desired, the frequency dependence of the bands (i.e., their shape). This pattern-recognition approach is conceptually distinct from the least-squares fitting techniques that are commonly applied to this type of data set. In the standard least-squares approach, the intensity profiles of the emission bands A and B are calculable from a model, and these "basis functions" are fit to each individual spectrum to determine the time-dependent emission intensities of the vibrational bands.

A hybrid between the least-squares and pattern-recognition approaches is possible if a situation arises in which *some*, but not all, of the patterns that are represented in a data set are known a priori. We already suggested in the Introduction that such is the case with the CO atmospheric simulation data. To illustrate the hybrid approach, we now assume that the frequency dependence of vibrational band A in our synthetic example can be predicted, but that of vibrational band B cannot. We begin by writing the emission intensity in the spectra as a function of both frequency (spectral element) and time (t):

$$I(\omega,t) = a(t)I_a(\omega) + b(t)I_b(\omega) \qquad (2)$$

in which a and b refer to the two emission bands as before. Experimentally, the frequency and time dependences of the emission are sampled only at discrete intervals. In the synthetic data, there are three time intervals, and 101 spectral elements (frequency intervals). Thus, 303 equations of the form

$$I(\omega_j,t_k) = a(t_k)I_a(\omega_j) + b(t_k)I_b(\omega_j) \qquad (3)$$

are necessary and sufficient to describe the data set. Of the parameters in this set of equations, $b(t_1)$, $b(t_2)$, and $b(t_3)$ can be determined from the application of the XCC, and the full set of $\{I_a(\omega_j)\}$ are assumed to be known. The parameters to be determined are $a(t_1)$, $a(t_2)$, and $a(t_3)$, and the set of $\{I_b(\omega_j)\}$: a total of 104 parameters. Thus, the system of equations represented by eq 3 is overdetermined, and standard least-squares algorithms can be used to determine the 104 parameters of interest from the set of 303 equations. This somewhat unconventional application of linear least-squares is what we refer to as Hybrid Linear Pattern Analysis.

The time-dependent amplitude of vibrational band A is determined by the HLPA approach to be $a(t_1) = 0.992$, $a(t_2) = 1.303$, and $a(t_3) = 1.593$, in close agreement with the values of 1.0, 1.3, and 1.6, respectively, which were used to construct the synthetic spectra. The remaining 101 parameters determined from the fit represent the frequency dependence of vibrational band B, and are not depicted because they are nearly identical

with the parameters determined from the purely pattern-recognition approach (Figure 3).

III. HLPA Analysis of CO Atmospheric Simulation Experiments

Before presenting the HLPA analysis of the CO atmospheric emission simulation experiments, we review briefly the motivation and experimental conditions for these experiments. Evidence for very high rotational excitation, with more than 2 eV of energy in rotation alone, has been observed in infrared spectra of the diatomic molecules NO[18] and OH[19] in the atmosphere. To model the rotationally excited populations of such species in the atmosphere, it is necessary to measure the formation and loss rates for the relevant molecular states, and thus it is necessary to produce measurable populations of excited molecules in the laboratory. The present experiment on CO represents one approach to the study of these rotationally excited species. CO itself is an important infrared-active species in the upper atmosphere, and significant effects on the infrared spectra of atmospheric CO caused by optical opacity, isotopic concentrations, and the temperature structure of the atmosphere have been observed.[20,21]

The experimental apparatus for the CO atmospheric emission simulation experiments (the LABCEDE facility at the Air Force Research Laboratory at Hanscom AFB) has been described previously.[16,17] In the data to be analyzed here, a 4.0-kV, 10.0-ms electron beam is pulsed through a sample of CO at 10.0 mTorr in the cryogenic chamber. A Michelson interferometer is used to obtain time-resolved, infrared emission spectra at 0.25-ms intervals, both during the electron beam excitation pulse, and for ~86.0 ms after the electron beam pulse is terminated. Two of these spectra are depicted in Figure 4. The time-resolved emission spectra consist of overlapping $\Delta v = 1$ emission bands from at least 12 excited vibrational states ($v' = 1-12$) of the ground electronic state ($X^1\Sigma^+$) of CO. At the ~2.5 cm^{-1} resolution of the spectra, the individual emission bands are not well resolved in the data.

Above 2200 cm^{-1}, several prominent band heads are evident in the emission spectra. The band heads in CO are known to occur at $J \approx 90$ in the R branch, and thus the rotational distribution of the CO resulting from processes initiated by the electron beam includes very high J states. On the other hand, the majority of the emission below 2200 cm^{-1} can be accounted for by low J ($J < 15$) emission. Thus, the rotational distribution of the CO is believed to be bimodal. This behavior is consistent with the expectation that the rates of rotational relaxation of the high and low J levels are quite different. For the low J levels, rotational equilibration occurs quickly (μs) with respect to the time scale of the experiment (ms), and thus the low J rotational distribution is expected to conform to a Boltzmann distribution. Further, this distribution is, to a good approximation, invariant throughout the experiment, because only a small fraction of the CO molecules are excited by the electron beam, whereas the "bath" of molecules that are not excited are rotationally equilibrated to the temperature of the walls of the cryogenic chamber. An effective rotational temperature of 90 K can be determined empirically to reproduce optimally the observed low J emission (other than the $v = 1 \rightarrow 0$ emission; see below). The high J molecules experience rotational relaxation at a slower rate than lower J molecules because the level spacings at high J are comparable with vibrational spacings. Thus, a small fraction of the excited-state population can become "trapped" in the high J states, although the exact rotational distribution (and its time dependence) are difficult to predict.

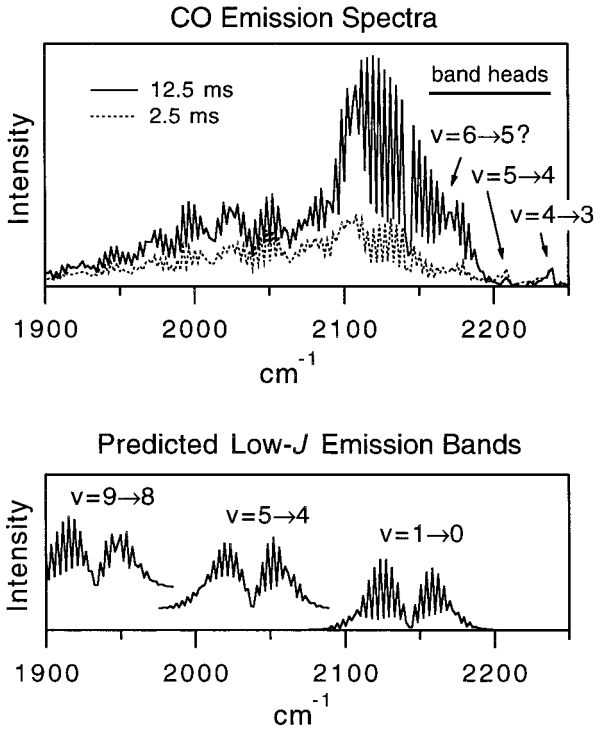

Figure 4. Top: Two examples of time-resolved emission spectra shown over a limited frequency range. Three additional band heads can be observed in the complete spectra at higher frequencies than are depicted here ($v = 3 \rightarrow 2$, $v = 2 \rightarrow 1$, $v = 1 \rightarrow 0$). Bottom: Three predicted low J vibrational emission bands, assuming an effective rotational temperature of 90 K. The predicted emission bands from $v = 2$ and higher are expected to represent the data accurately, but the predicted $v = 1 \rightarrow 0$ emission band does not take into account the effects of optical opacity.

The relative populations in the high J and low J levels are a function of v and excitation time. For example, consider the $v = 5 \rightarrow 4$ emission bands. At the time corresponding to termination of electron-beam excitation, the population in the high J levels responsible for the $v = 5 \rightarrow 4$ band head feature is approximately 8% of the population that radiates in the $v = 5 \rightarrow 4$ low J emission band (see Figure 4). If the populations for $v = 2-5$ are added together, the high J to low J ratio increases to about 0.16.

The five "band heads" observed in the data above 2200 cm^{-1} do not overlap each other substantially, and can be integrated easily to determine their time-dependent amplitudes. The band heads are not explicitly analyzed in the analysis presented here, except insofar as they overlap with the low J bands. A more serious obstacle to the analysis is an optical opacity effect that is associated with the $v = 1 \rightarrow 0$ low J emission band. Although the $v = 1$ molecules are expected to have an effective low J rotational temperature that is identical with that of the $v \geq 2$ molecules, the large relative concentration of $v = 0$ molecules leads to an optical opacity in the $v = 1 \rightarrow 0$ emission band *as seen at the detector*. This optical opacity is extremely difficult to model under the measurement conditions because, although the density of molecules in the ground vibrational state is essentially uniform throughout the chamber, the spatial distribution of excited-state molecules is highly nonuniform and forms an irregular cloud of varying density along the one-meter path of the electron beam. This path in turn obliquely intercepts the conical viewing region.

The optical opacity in the $v = 1 \rightarrow 0$ emission is clearly evident in Figure 4. Most of the emission observed in the time-resolved spectra between 2130 and 2190 cm^{-1} is caused by the

$v = 1 \rightarrow 0$ emission, but the observed band shape does not even approximately match the band shape predicted by a Boltzmann-distribution assumption, neglecting self-absorption. The inability to predict the frequency dependence of the $v = 1 \rightarrow 0$ emission band implies that the time dependence of this emission cannot be determined in a simple fashion by least-squares fitting. In addition, the $v = 1 \rightarrow 0$ emission overlaps substantially with several other emission bands, particularly the $v = 2 \rightarrow 1$, $v = 3 \rightarrow 2$, and $v = 4 \rightarrow 3$ emission bands. For this reason, the time dependence of the emission from these bands is not easily determined using standard least-squares techniques, despite the fact that their band profiles are known.

The HLPA technique that was outlined in Section II provides an alternative data analysis approach, in which the $v = 1 \rightarrow 0$ emission band can be treated as a pattern to be identified by XCC. The remaining low *J* vibrational emission bands, however, need not be treated by pattern recognition, because their shapes have been accurately determined by synthetic spectral fitting. Thus, the HLPA technique permits the utilization of the known band shapes, together with the time dependence of the $v = 1 \rightarrow 0$ band determined by pattern recognition, to determine simultaneously the remaining parameters of interest: the frequency dependence of the $v = 1 \rightarrow 0$ band and the time dependences of the remaining low *J* bands.

The pattern-recognition approach to the CO data set implicitly assumes that the band shape of the optically thick $v = 1 \rightarrow 0$ emission does not change as a function of time. The exact band shape is governed by the relative populations of the $v = 1$ and $v = 0$ states, the oscillator strengths of the various $v = 1 \rightarrow 0$ rotational transitions, collisional deactivation rates, transport of molecules out of the field of view, the rotational distribution of the $v = 1$ molecules, and geometrical considerations. Of these parameters, only the population of the $v = 1$ excited state is expected to change with time. (As with the other excited vibrational states, the low *J* rotational distribution of the $v = 1$ molecules is expected to conform to a time-independent Boltzmann distribution.) However, the population of the $v = 1$ excited state remains a small fraction of that of the ground vibrational state, and thus all changes in the band shape are expected to be minor. The results of the HLPA analysis presented below support this argument.

The first step in the HLPA analysis of the CO data set is the pattern-recognition determination of the time-dependent amplitude of the optically thick $v = 1 \rightarrow 0$ band. As explained in Section II, the XCC identifies patterns within a data set by searching for fragments of the patterns that are repeated uncorrupted (by overlap with other patterns) in each of the spectra. As is clear in Figure 4, the optically thick $v = 1 \rightarrow 0$ emission band overlaps heavily with other emission bands. At frequencies above ~2165 cm^{-1}, the $v = 1 \rightarrow 0$ emission band should be *relatively* free from overlap, although it almost certainly overlaps with the $v = 5 \rightarrow 4$ band head above 2190 cm^{-1}. In addition, the existence of a $v = 6 \rightarrow 5$ band head around 2180 cm^{-1}, obscured by the much stronger $v = 1 \rightarrow 0$ band, cannot be ruled out. Finally, it is difficult to determine the exact frequency at which the optically thick band can be assumed to be free from overlap with the $v = 2 \rightarrow 1$ emission band.

We have used the XCC to determine the amplitude of the $v = 1 \rightarrow 0$ emission band in more than 100 spectra, a result that will be discussed below. First, however, we illustrate this application of the XCC with just two spectra, in particular, the two spectra depicted in the top panel of Figure 4, across the frequency range 2165−2215 cm^{-1}. Two maxima are observed

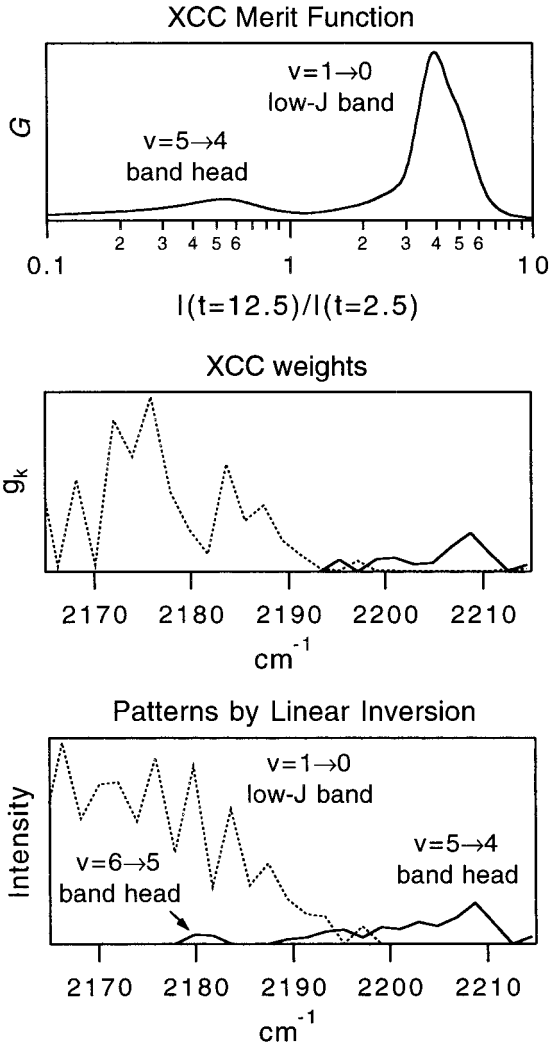

Figure 5. Top: XCC merit function *G* as a function of ratio direction for the two spectra shown in Figure 4 over the frequency interval 2165−2215 cm^{-1}. The two maxima correspond to the $v = 5 \rightarrow 4$ band head emission, and the optically thick $v = 1 \rightarrow 0$ low *J* emission. Middle: XCC weight functions plotted as a function of frequency at ratio directions corresponding to the two maxima in the top panel. Bottom: Results of the linear inversion method.

in the XCC merit function (Figure 5, top panel), indicating the presence of two "patterns", and in the middle panel of Figure 5, the weight functions corresponding to each of these maxima are plotted as a function of frequency. As explained in Section II, the weight functions indicate those portions of patterns that are repeated, uncorrupted by overlap with other patterns, in each of the spectra. Thus, on the basis of these weight functions, the two maxima observed in the merit function are clearly assignable as the $v = 1 \rightarrow 0$ optically thick emission and the $v = 5 \rightarrow 4$ band head emission. The $v = 2 \rightarrow 1$ emission band is not identified as a distinct pattern, because no regions of the spectra exist within which this emission band is uncorrupted by overlap with other vibrational bands. Notice, however, that the weight functions for the $v = 1 \rightarrow 0$ optically thick pattern are very nearly zero at two resolution elements below 2170 cm^{-1}. This observation is consistent with the optically thick pattern overlapping with the "tail" of the $v = 2 \rightarrow 1$ emission band below this frequency.

In addition, the XCC weights provide evidence for the existence of a $v = 6 \rightarrow 5$ band head near 2180 cm^{-1}. Although the XCC weights for the optically thick pattern are not zero near 2180 cm^{-1}, they are substantially lower than the weights

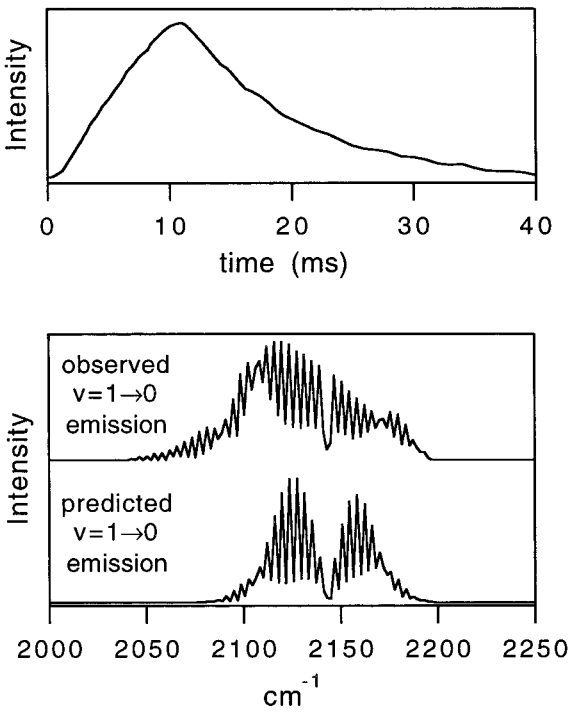

Figure 6. Time and frequency dependences of the optically thick $v = 1 \rightarrow 0$ emission band. Top: time dependence of the optically thick band as determined by XCC. Bottom: frequency dependence of the optically thick band as determined by the HLPA method, with the frequency dependence of the band predicted by neglecting the effects of optical opacity, for comparison.

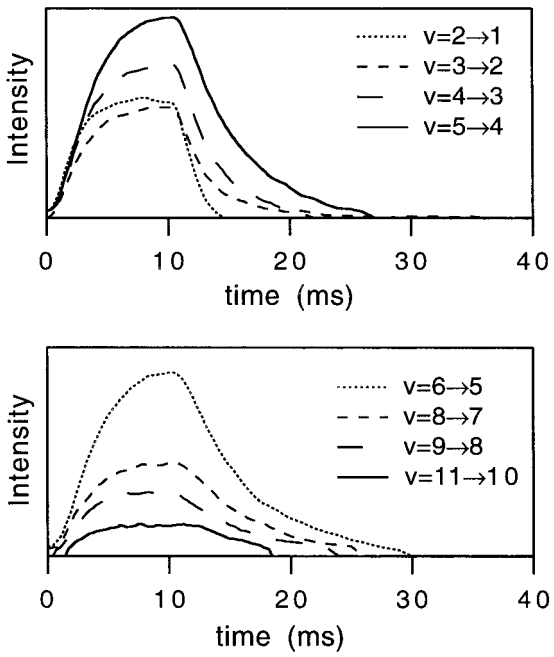

Figure 7. Time dependence of the emission intensity from selected low *J* vibrational bands as determined by the HLPA technique.

at higher and lower frequencies; this observation is consistent with a slight "corruption" of the pattern by overlap with a weak band head. This evidence is augmented by the application of the linear inversion technique for pattern reconstruction (Section II) to this frequency region. In this case, the two identified patterns are not expected to account for 100% of the intensity observed within the frequency interval chosen, but the linear inversion method can be applied naively to this region anyway, and the results are depicted in the bottom panel of Figure 5. As expected, the "tail" of the $v = 5 \rightarrow 4$ band head is observed to extend to below 2190 cm^{-1}. A small bump is also observed in this "reconstructed pattern" around 2180 cm^{-1}. Although it is possible that this bump might represent an unexpected feature in the $v = 5 \rightarrow 4$ band head, the bump occurs at the frequency at which the maximum of the $v = 6 \rightarrow 5$ band head is predicted to be located. Thus, a reasonable inference is that the $v = 6 \rightarrow 5$ band head is present, although the emission is weak, and that it has a time dependence that is approximately the same as that of the $v = 5 \rightarrow 4$ band head. As far as either the XCC or the linear inversion methods are concerned, patterns with identical time-dependent amplitudes (ratio directions) are indistinguishable, and thus the $v = 6 \rightarrow 5$ and $v = 5 \rightarrow 4$ band heads may be lumped together into one pattern.

Having investigated the insights that the XCC provides for the 2165−2215 cm^{-1} region, we now return to the determination of the time-dependent amplitude of the optically thick $v = 1 \rightarrow 0$ pattern. The top panel of Figure 6 depicts the time dependence of the $v = 1 \rightarrow 0$ emission intensity that is determined by applying the XCC to 100 spectra simultaneously. Thus, by adopting a pattern-recognition analysis of the data set, the time dependence of the $v = 1 \rightarrow 0$ emission intensity has been determined without any knowledge of the band profile. At this point, it remains to determine the time-dependent amplitudes of the $v = 2 \rightarrow 1$, $v = 3 \rightarrow 2$, and $v = 4 \rightarrow 3$ emission bands, which also could not be determined by standard least-squares techniques because of their overlap with the optically thick band.

The HLPA approach outlined in Section II provides a conceptually straightforward way to determine these time dependences. Numerically, the HLPA technique relies on the fact that the data set can be described by a set of equations of the form in eq 3, except that there are many more than two patterns in the present application. If all 100 of the spectra were included in the HLPA analysis, the total number of equations would be ∼20,000, with ∼1400 unknown parameters. Obviously, the problem is highly overdetermined, but in practice such a large fit is computationally tedious and unnecessary. The time-resolved spectra change slowly from one time interval to the next, and it is sufficient to choose a small subset of the data that evenly spans the time interval in which the kinetics of interest is played out. The calculations reported here use 10 of the time-resolved spectra, which were chosen at 2.0-ms intervals, from 2.0 to 20.0 ms (after the electron beam is turned on).

The bottom panel of Figure 6 depicts the frequency dependence of the optically thick $v = 1 \rightarrow 0$ band that is determined by the HLPA method, as well as the shape of the band that is predicted by neglecting the effects of optical opacity. This comparison underscores the enormity of the optical opacity and may aid in the modeling of the self-absorption, transport, and deactivation effects that lead to the optical opacity. Figure 7 depicts the time-dependent emission intensity from selected low *J* vibrational bands, as determined by the HLPA technique. As mentioned previously, only 10 time intervals were included in the HLPA calculations, but the intensities of the various vibrational bands at all other times can be determined in a straightforward manner by least-squares fitting (because we have now determined the frequency dependence of the optically thick $v = 1 \rightarrow 0$ band). The smoothness of the curves in Figure 7 constitutes evidence that the HLPA technique provided a physically reasonable solution. The kinetics of the system will be analyzed in a future publication.

Finally, in Figure 8 are depicted the residuals of the HLPA calculation for three of the 10 time-resolved spectra that were

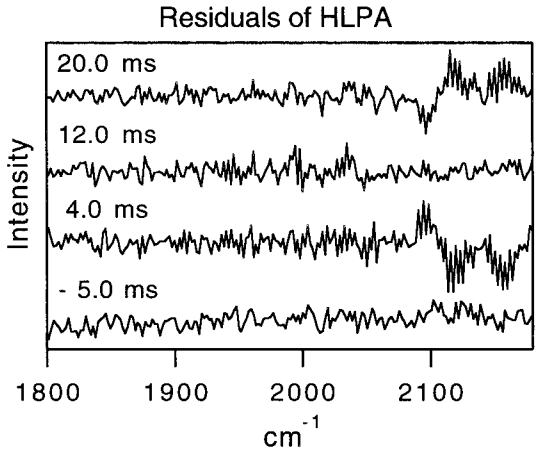

Figure 8. Residuals (observed minus calculated) of the HLPA calculation for three of the 10 time-resolved spectra included in the fit. The bottom trace, for comparison, is a spectrum taken 5.0 ms before the electron beam is turned on, which provides a measure of the experimental noise.

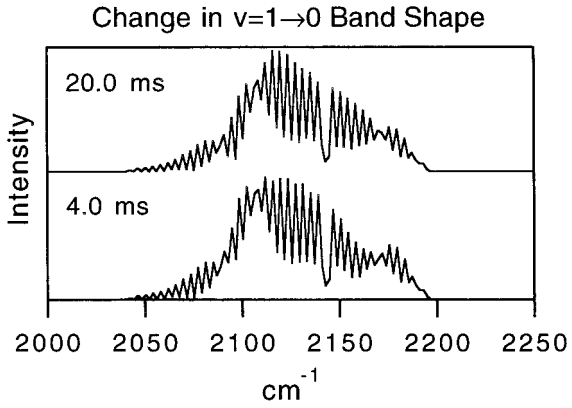

Figure 9. The change in the shape of the optically thick $v = 1 \rightarrow 0$ band shape as a function of time, assuming that the structure in the residuals (Figure 8) is due entirely to the breakdown of the assumption of constant band shape. That is, the band shapes depicted here were generated by adding the residuals in Figure 8 (above 2075 cm^{-1}) to the $v = 1 \rightarrow 0$ band shape that was determined by the HLPA technique (in Figure 6).

included in the analysis, along with a spectrum taken *before* the electron beam was turned on, for comparison. The residuals of the fit are similar in amplitude to the "background" spectrum at most time intervals. However, at early and late times, the residuals above 2075 cm^{-1} are significantly larger than the background noise. The structure in the residuals appears to invert from 4.0 to 20.0 ms, and the two prominent "lobes" in this structure line up in frequency with the *P* and *R* branches of the $v = 1$ low *J* emission band. We believe this implies that the optically thick $v = 1 \rightarrow 0$ band profile changes slightly with time. The structures observed in the residuals are small relative to the integrated intensity of the $v = 1 \rightarrow 0$ band, and the corresponding changes in the shape of the optically thick band are subtle. In Figure 9 we depict the change in band shape of the $v = 1 \rightarrow 0$ emission between 4.0 ms and 20.0 ms, assuming that the structure in the residuals is due entirely to the breakdown of the assumption of constant band shape. That is, the band shapes that are depicted in Figure 9 were generated by adding the residuals in Figure 8 (above 2075 cm^{-1}) to the $v = 1 \rightarrow 0$ band shape that was determined by the HLPA technique (in Figure 6). The difference in band shape between 4.0 ms and 20.0 ms is subtle but noticeable; the band shapes at intermediate times vary smoothly between these extremes.

IV. Conclusion

Numerical pattern recognition algorithms (in particular, the HLPA technique that is introduced here) have played a critical role in the successful analysis of CO atmospheric emission simulation experiments. Similar techniques may also be useful for analyzing "field data", such as the infrared emission spectra of the Earth's atmosphere that were obtained recently from the CIRRIS 1A instrument aboard the Space Shuttle.[20] Analysis of such data is complex, because of the numerous emitting species, each of which may have substantial populations in highly excited rovibrational states. Spectroscopic pattern-recognition techniques such as the XCC and HLPA may be valuable in extracting from such data sets the relative contributions from individual chemical species, as in Refs 1 and 2, and the various vibrational bands belonging to a single species, as we have done here. Other potential applications of these techniques include spectra that are obtained for purposes of surveillance or in atmospheric remote sensing experiments. The spectra that are obtained in both cases frequently contain features which result from unknown species, as well as optical opacity effects. As a result, such spectra often cannot be analyzed with standard least-squares techniques, and the ability of the XCC and HLPA to identify and extract unknown patterns may prove useful in these cases.[22]

Acknowledgment. This research was supported by AFOSR under Grant no. F49620-97-1-0040 and AFOSR Project 2303ES, Task 92VS04COR. M.P.J. acknowledges support from the Department of the Army under a National Defense Science and Engineering Graduate Fellowship, and from the Fannie and John Hertz Foundation.

References and Notes

(1) Jacobson, M. P.; Coy, S. L.; Field, R. W. *J. Chem. Phys.* **1997**, *107*, 8349.

(2) Coy, S. L.; Jacobson, M. P.; Field, R. W. *J. Chem. Phys.* **1997**, *107*, 8357.

(3) O'Brien, J. P.; Jacobson, M. P.; Sokol, J. J.; Coy, S. L.; Field, R. W. *J. Chem. Phys.* **1998**, *108*, 7100.

(4) Press: W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1992; Chapter 15.

(5) Ruckstuhl, A. F.; Stahel, W. A.; Dressler, K. *J. Mol. Spectrosc.* **1993**, *160*, 434.

(6) Stahel, W. A.; Ruckstuhl, A. F.; Senn, P.; Dressler, K. *J. Am. Stat. Assoc.* **1994**, *89*, 788.

(7) Ruckstuhl, A. F.; Dressler, K. *J. Mol. Spectrosc.* **1994**, *168*, 185.

(8) Dabrowski, I.; Tokaryk, D. W.; Watson, J. K. G. *J. Mol. Spectrosc.* **1998**, *189*, 95.

(9) Meloun, M.; Houalla, M.; Proctor, A.; Fiedor, J. N. *PC-Aided Statistical Data Analysis*, Vol. 1 of *Chemometrics for Analytical Chemistry*; Ellis Horwood: New York, 1992.

(10) Hercules, D. M.; Houalla, M.; Proctor, A.; Fiedor, J. N. *Anal. Chim. Acta* **1993**, *283*, 42.

(11) Fiedor, J. N.; Proctor, A.; Houalla, M.; Hercules, D. M. *Surf. Interface Anal.* **1993**, *20*, 1.

(12) Cowe, I. A.; McNicol, J. W.; Cuthbertson, D. C. *Anal. Proc.* **1990**, *27*, 61.

(13) Zupan, J.; Gasteiger, J. *Anal. Chim. Acta* **1991**, *248*, 1.

(14) Larch, D. E. *SPIE* **1996**, *2758*, 2.

(15) Frasinski, L. J.; Codling, K.; Hatherly, P. A. *Science* **1989**, *246*, 1029.

(16) Upschulte, B. L.; Green, B. D.; Blumberg, W. A. M.; Lipson, S. J. *J. Phys. Chem.* **1994**, *98*, 2328.

(17) Holtzclaw, K. W.; Upschulte, B. L.; Caledonia, G. E.; Cronin, J. F.; Green, B. D.; Lipson, S. J.; Blumberg, W. A. M.; Dodd, J. A. *J. Geophys. Res.* **1997**, *102*, 4521.

(18) Armstrong, P. S.; Lipson, S. J.; Dodd, J. A.; Lowell, J. R.; Blumberg, W. A. M.; Nadile, R. M. *Geophys. Res. Lett.* **1994**, *21*, 2425.

(19) Dodd, J. A.; Lipson, S. J.; Lowell, J. R.; Armstrong, P. S.; Blumberg, W. A. M.; Nadile, R. M.; Adler-Golden, S. M.; Marinelli, W. J.; Holtzclaw, K. W.; Green, B. D. *J. Geophys. Res.* **1994**, *99*, 3559.

(20) Dodd, J. A.; Winick, J. R.; Blumberg, W. A. M.; Lipson, S. J.; Armstrong, P. S.; Lowell, J. R. *Geophys. Res. Lett.* **1993**, *20*, 2683.

(21) Dothe, H.; von Esse, F.; Sharma, R. D. *J. Geophys. Res.* **1996**, *101*, 19715.

(22) Vititoe, D. L.; Lipson, S. J.; Allred, C. L.; Lockwood, R. B.; Blumberg, W. A. M.; Armstrong, P. S.; Jacobson, M. P.; Coy, S. L.; Field, R. W. *Eos, Trans. Am. Geophys. Union* **1997**, *78*, F526.