Automatic and semi-automatic assignment and fitting of spectra with PGOPHER.

Colin M Western[a], Brant E Billinghurst[b]

Abstract

Two new tools for computer assisted assignment of rotational spectra with the PGOPHER program are presented, aimed particularly at spectra where many individual lines are resolved. The first tool tries many different assignments, presenting a small number for possible refinement and a preliminary version of this has already been presented. The second tool, the nearest lines plot (a new style of residual plot) provides a clear indication as to whether a trial calculation is plausible, and allows rapid assignment of sets of related lines. It gives good results even for dense spectra with no obvious structure and in the presence of multiple interfering absorptions. The effectiveness of these tools is demonstrated by the analysis of high resolution IR spectra of 8 bands of *cis-* and *trans-*1,2-dichloroethene where, including hot bands and isotopologues, 31 vibrational transitions and 158,316 individual lines have been analysed, including perturbations for the higher energy states. Walkthroughs are presented to show this process is rapid.

Introduction

Molecular spectra provide an invaluable source of information on structure, bonding and other properties of molecules. Many spectroscopic methods, including microwave, Fourier transform and laser-based methods can produce well resolved spectra with many thousands of individually resolved lines, which contain much information, but also a significant challenge in analysis to extract the information present. This is exacerbated by recent technological advances, such as broad band microwave[1] and frequency comb[2] spectroscopy meaning that ever more spectra are available, and the speed of analysis limits the possible studies.

To understand the complexities that arise, and how a computer can help, consider the process of assigning a rotationally resolved spectrum. Typically a trial simulation might be set up, using *ab initio* geometries or other estimated values, and compared to experiment. If the spectrum has an obvious pattern, comparison of the experiment and simulation is straightforward, and the trial can be accepted or rejected. An interactive program, such as PGOPHER [3] makes this trial and error process quick, and also simplifies the next step, trying specific assignments, fitting constants and checking for an improved simulation. For closed shell linear molecules the pattern of lines produced is normally very simple, being only determined by one or two rotational constants, and an incorrect trial is easily identified and incorrect assignments are likely to only be J values off by a small number. This process becomes more difficult as the spectrum becomes more complicated, such as for asymmetric tops or where unpaired electrons are introduced. As the pattern of lines becomes more complicated not only does the number of possible miss-assignments increase, but also

[a] School of Chemistry, University of Bristol, UK, C.M.Western@bristol.ac.uk
[b] Canadian Light Source, University of Saskatchewan, Canada, Brant.Billinghurst@lightsource.ca

comparing the overall pattern of lines becomes more difficult. The most significant consequence of this is that it becomes more difficult to detect trial simulations that are partially correct, which is an essential part of assignments where several constants are important for the overall form of the spectrum. The method presented in this paper is particularly aimed at very complex spectra, with many thousands of lines, with little or no obvious pattern and an average line spacing of the order of the linewidth, where an initial assignment is difficult and a reasonable (if partial) simulation can be very hard to identify.

Computer assistance is essential for such complex spectra, and a recent development has indicated a possible method. The group of Brooks Pate presented the AUTOFIT algorithm[4], specifically aimed at determining three rotational constants from broadband microwave spectra of asymmetric tops. This essentially involved trying every possible assignment for selected lines and ranking the resulting fits. In a preliminary version of the current work we presented[5] a version of this method incorporated into the PGOPHER program[3] using a similar basic algorithm but written in a more general way and so it can be applied to a much wider range of systems. This paper is a follow up to this, presenting some refinements to the algorithm, but also addressing the complementary problem of assessing the quality of trial simulations.

The key figure of merit for ranking the trial fits in both the Pate and earlier PGOPHER automated assignment methods is the root mean square observed – calculated line positions. Both methods work by just looking at a subset of the available lines (typically only a very small number) and fitting just a subset of the possible constants. A good match between observed and calculated positions is of course essential for a model, but not sufficient for dense spectra where it is easy to find a random set of lines that will fit to a few calculated lines, so other quality assurance mechanisms are required.

Intensity information provides an obvious additional check, but both the automatic methods above make little direct use of it, mainly because intensity information is typically much less accurate than frequency information. Comparing intensities is thus best done by manual inspection that can see that trends are correct, even if not quantitatively predicted. Some specific plots have in fact been recently added to PGOPHER[6] to assist in this, but they are not (yet) useful for spectra of the complexity presented here, mainly because the number of overlapping transitions mean there are few unblended lines. The intensity information is thus best used at the end of the process, when most lines have been assigned, and overall trends can be verified.

Other traditional spectroscopic tools for making and verifying assignments include combination differences[7], applicable if one of the vibrational states in the transition to be analysed is already known. For example, if the ground state constants are known, differences between transitions with a common upper level can be exactly calculated. This is used in many spectroscopic analyses, and indeed the Automated Spectra Assignment Procedure (ASAP)[8] uses this to pick assignments. However, as discussed below, the ASAP technique has some limitations and the line density and lack of transitions with a common upper level mean it is not suited to the spectra analysed here.

Double resonance techniques provide very direct confirmation of assignments but are typically limited to a relatively small number of transitions, require more equipment and are anyway not applicable to every technique. In the context of analysis of complex spectra it is worth mentioning the work of Martin-Drumel et al on automated microwave double resonance spectroscopy[9] which can in principle tackle spectra of comparable complexity to those discussed here.

Genetic algorithms[10] should also be included in this discussion of assignment methods. They have been used to assign complex spectra but require a complete model of the spectrum in the region to be analysed and are thus problematic for the spectra analysed here which contain overlapping bands from isotopologues and hot bands, and perturbations. A key feature of the methodology introduced in this paper is that only a small subset of transitions, selected by quantum numbers, not frequency, are examined at any one time, and the process is insensitive to interfering transitions. Another general technique for looking at complex problems, artificial neural networks, have also recently been explored for determining constants from spectra [11], but is not yet at a stage to be applied to real spectra, with the presence of multiple species noted as a particular complicating factor in that work.

In the current paper a new plot type is introduced, termed a nearest lines plot, which gives a particularly clear indication of partially correct assignments and allows new assignments to be made rapidly. In design it bears some similarity to Loomis-Wood plots[12, 13] in that the same spectrum is plotted multiple times with a set of selected displacements in frequency, but the density of lines in the spectra tackled here is such that the traditional Loomis-Wood method of picking progressions is ineffective.

In this paper, the new tools are demonstrated on new high-resolution (0.001 cm^{-1}) Fourier transform spectra of *cis-* and *trans-* 1,2-dichloroethene taken using the Canadian Light Source[14]. These spectra provide an excellent testbed, in that the spectra are all very dense, with an average line spacing of the same order as the line width and have largely been neglected in previous high-resolution studies. For the *trans* species there have been two studies of the 898 cm^{-1}[15, 16] band, which gave full analyses of this band ignoring perturbations, but no other high resolution work has been reported. For the *cis* species microwave and millimetre-wave spectra are available [17, 18], but no high resolution infra-red work apart from our preliminary publication[5] on the v_{11} band at 570 cm^{-1}. This paper presents an improved analysis of this latter band, now including the effects of a perturbation, and extends the analysis to all the other vibrational bands of both species below 1000 cm^{-1}. The set of bands analysed provides excellent tests of the methodology, as several complicating factors are present, including hot bands, the $^{35}Cl_2$ and $^{35}Cl^{37}Cl$ isotopologues and several perturbations. The overall procedure allows rapid analysis of all these bands. Accompanying this paper is a new release of the PGOPHER program[19] incorporating the changes described here, and the documentation accompanying the program includes detailed walk-throughs for selected bands of the fitting process.

Experimental

Spectra of *cis*-and *trans*- 1,2-dichloroethene were collected by making use of the high brightness[14] provided by the synchrotron radiation from the far-infrared beamline at the Canadian Light source using a Bruker IFS125HR spectrometer. A 2-meter White type gas cell was used with a path length of 72 meters operated at ~200 K and a selection of pressures in the range 0.002 – 0.4 Torr to allow weak and strong features to be seen. The spectra were collected using a scanner velocity of 80 kHz and averaging over up to 550 scans. The Fourier transformed spectra are included in the supplementary information[20], which also details the conditions for each spectrum. The compounds were obtained from commercial sources, and subject to freeze-thaw cycles for purification before filling the spectrometer.

The Automatic Assignment Algorithm

The automatic assignment algorithm used is rather similar to that described in our preliminary publication[5], but the description is repeated here for completeness. In principle it can be used at any stage of the fitting process where there is a choice of assignments, but the workflows described below only use it for initial assignments. It requires a numerical list of observed line positions, and a set of parameters (rotational constants here) giving a qualitative simulation of the spectrum. In its most general form the steps are then:

1. Select n_{par} parameters to refine and (optionally) a range for some or all of them

2. Select a small number of (calculated) transitions to fit, n_{fit}, sufficient to determine the parameters selected in step 1, and a search window, w_S, around each of these transitions.

3. Select some additional check transitions, n_{check}, and an acceptance window, w_a, associated with these.

These three selections are key to the success of the process, and where judgement is required. The original AUTOFIT paper[4] applied specifically to asymmetric top microwave spectra and used the three rotational constants, *A*, *B* and *C*, for the parameters with three transitions chosen for fitting, though the specific choice of transitions was not that critical. Our implementation is more much general, and any combination of parameters can be used, though practical application requires restricting the number of parameters, and restricting the search to transitions involving a carefully selected range of quantum numbers, as discussed in the example applications below.

Give the required selections, the program then searches through all the possible sets of assignments of the fit transitions to the observed transitions, limited only by the requirement that each individual assignment has the observation within $\pm w_S$ of the initially calculated fit transition. For each set of assignments a standard least squares fit of the n_{fit} transitions to the n_{par} selected parameters is performed. If a range has been specified for a parameter, and the fit moves the parameter outside the range, then the assignment is rejected. The assignment is also rejected if the average error of the fit is more than the acceptance window. If the assignment is not rejected the position of each of the check transitions is calculated from the fitted parameters and a count, n_{OK}, is made of those with an observed transition within the acceptance window, and the root mean square difference,

σ_{check}, between the calculated and observed positions of the accepted transitions is calculated. The sum of the intensity of the accepted lines, I_{check} is also determined, if intensity information is present in the experimental line list. Having cycled through all the possible assignments, the best assignments are presented, ranked by n_{OK} and σ_{check}. A refinement introduced here is to rank assignments with different n_{OK} separately, and the user interface now allows the results for a particular n_{OK} to be displayed, or all n_{OK} above a particular value. In either case the display is sorted by σ_{check} and I_{check} and the fitted parameters are also displayed.

There are some additional options to the process above; already known assignments can be included in the process, and these are included as additional observations in the individual least squares fits but are not included in the n_{OK} or σ_{check} statistics. There is also the possibility of putting some limits on the possible assignments if required. These currently include a limit on the number of observed transitions that can be assigned to any one fitted line, and a requirement that the relative order of observed and calculated transitions be preserved.

Even when searches are well set up cycling through all the possible assignments can be very time consuming, so the program has some key features to make the search as fast as possible. Firstly, testing an individual assignment is independent of testing any other, so the search process is very easily parallelised, and the current implementation will make full use of all the cores on a multi-CPU system. Secondly, most of the first cycle of each of the fits can be precalculated, specifically the initial calculated values and derivatives, so the first cycle can be very rapid, only involving some simple matrix operations on small matrices (no larger than $n_{fit} \times n_{par}$). The second cycle is speeded up by re-using the initial derivatives and only on subsequent cycles are the derivatives re-calculated in full. While most spectroscopic fits are formally non-linear the changes in parameters are typically small so convergence is normally rapid, often within a single cycle. Additionally, testing for rejection of a given fit (on the grounds of a parameter moved too far or overall fit quality) is done after every cycle including the first. Any rejection mechanism therefore speeds the search process by reducing the number of fit cycles, and also by skipping the testing of the check transitions.

The Nearest Lines Plot

As will become clear from the examples below, the search process will in general be aimed at a subset of the overall spectrum and additional measures of success other than residuals are needed. Successful fits on a small subset of lines in a dense spectrum with no obvious pattern cannot be verified by simple comparison of the experimental and simulated spectrum. A special plot, which we term the *nearest lines plot*, is introduced here for this specific purpose. The idea of the plot is that, for each calculated line, the observed − calculated value is plotted for each observation within a certain range. Formally, given a set of calculated lines at c_i and observed lines at o_j a point is plotted at $y = o_j − c_i$ for every i and j. There are various choices for the x axis, but J' is a good initial choice. The net effect is to plot vertically a separate copy of the experimental spectrum for each calculated line, with each copy offset vertically by the position of the calculated line. Each point on the plot corresponds to a potential assignment of an observed line to a calculated transition. If the calculation is

perfect, then the result will be a horizontal line of points at $y = 0$, and a random set of points above and below. As can be seen in the examples below, a calculation that is close to correct will typically show a curve (reflecting a systematic trend in residuals) surrounded by a sea of random points. A key additional feature is some selectivity in the calculated lines considered; while plotting every line is useful for a nearly complete model, for the initial stages some selectivity is essential.

To see how this works in practice, consider the simple infra-red spectrum of the v_2 band of N_2O spectrum shown in Figure 1. The top of figure 1a shows an experimental line list, and below that an initial simulation set up by using known ground state constants, and rough manual adjustment of the upper state origin and B to give a qualitatively correct plot. Figure 1b is a nearest lines plot based on these two plots, where the calculated lines have been restricted to the P and R branch lines. The horizontal axis chosen here is essentially J though formally is line number, with the lines sorted by upper state J, so P(J+1) and R(J-1) lines are adjacent, rather than on top of each other. In some senses this example is too simple but does serve to demonstrate some of the essential features. Several curves are clear on this plot, and the eye can easily distinguish the better curves. Each point corresponds to an assignment of a specific observation to a particular set of quantum numbers, so selecting a particular curve corresponds to making multiple assignments which will show a systematic trend in residuals, and so are likely to give a good fit after adjusting constants. In the current case, the correct assignment is the heavy line curving downwards from obs-calc = +0.1 on the left to line number 80 on the bottom. The user interface allows selecting lines within a region, or on a line, and here about 30 observations on the correct line can be assigned with a single mouse action. Fitting the lines assigned in this way to the band origin, B' and D' and re-doing the plot gives a horizontal line in the centre (see Figure 1c, where the greyed out points indicate the assigned lines), allowing the remaining correct assignments to be easily picked out with the mouse.

Figure 1 does indicate potential pitfalls, in that there are several other curves that could be selected and require using other information to reject. Some of these can be simply understood on physical grounds. The heavy curves parallel to the correct assignment correspond to shifting all the correct assignments up or down one in J and would not give the correct position of the central band head following fitting. Other curves correspond to the weaker lines in the spectrum which in this case arise from hot bands, so can be distinguished by intensity.

Additional selectivity can be applied to these plots using combination differences to filter the points plotted, if information is available for one of the states involved. To do this a search is performed for each potential point plotted (with a particular o_j and c_i) looking for another calculated line, c_k, with one end of the transition in common and the point is only plotted if there is another observation o_l such that $o_j - c_i = o_l - c_k$. The end of the transition in common must be in the state for which the constants are not known. For the infra-red spectrum considered here the lower state constants could be known from microwave spectroscopy, so the difference R(J−1) − P(J+1) is known for any J. Applying this restriction to the nearest lines plot (Figure 1d, to be compared with Figure 1b) rejects most of the plotted

points, and the correct assignment becomes even clearer. This is a particularly favourable example, but the asymmetric top examples below indicate that more generally applying the combination difference filter is useful, but not essential.
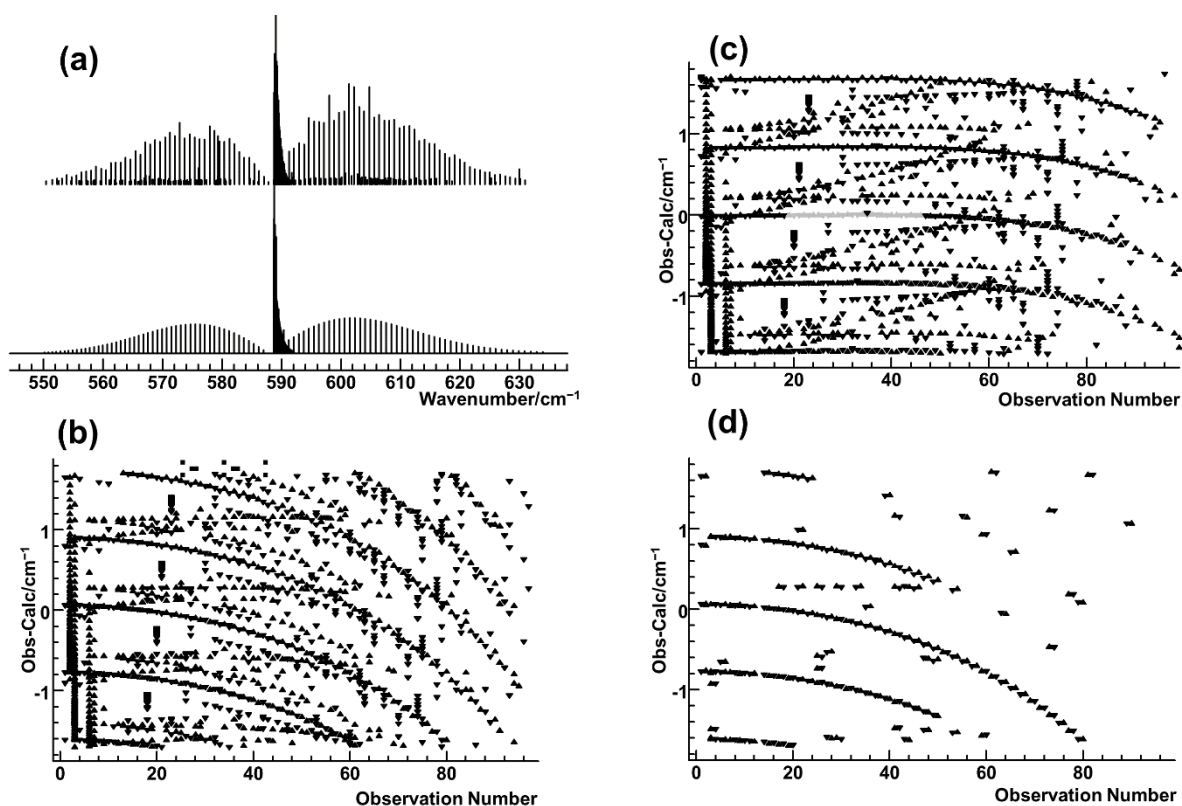

Figure 1. The nearest lines plot applied to the ν_2 band of N_2O. (a) shows the experimental line list (upper trace) and initial simulation (lower trace). (b) shows the corresponding nearest lines plot where, for each calculated line in (a), the distance away of each observed line is shown. The vertical axis is observed − calculated line position, and the horizontal axis is calculated line, sorted by J and ΔJ. (c) shows the same plot after an initial fit; note the clear horizontal line at obs−calc = 0. (d) shows (b) after adding the requirement that transitions satisfy ground state combination differences.

Application to the ν_5 band of *cis*-1,2-dichloroethene

To provide a concrete example of the automatic assignment and nearest lines plot in action the first analysis of the ν_5 band of *cis*-1,2-dichloroethene at 169 cm^{-1} is presented here. Even at the low temperature used (207 K) the spectrum (Figure 2) shows no obvious pattern other than the broad P and R branch envelopes, though the experienced spectroscopist can pick out some band heads on close examination of parts of the spectrum. The $^{35}Cl_2$ species will give ~14,000 lines with a strength > 5% of the strongest and, as discussed below, other species and hot bands each contribute a similar number of lines with similar intensity. The starting point is the ground state rotational constants from microwave spectra by Leal *et al*[17], with the upper state constants taken to be the same as the ground state, and the band origin manually adjusted to give the band centre in approximately the right place.
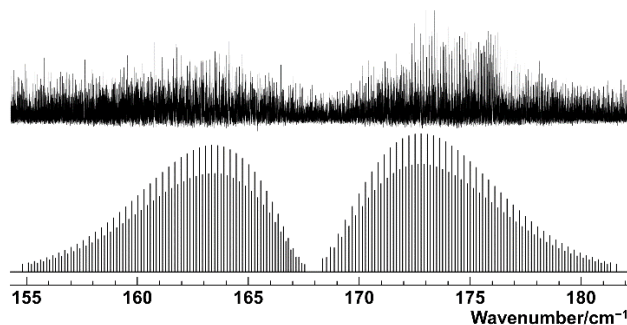
7

Figure 2 Central portion of the ν_5 band of *cis*-1,2-dichloroethene recorded at a temperature of 207 K (upper trace). The lower trace shows a simulation of $K'_a = 0$ lines with $\Delta J \neq 0$

The automatic fitting process works best if only a subset of lines is considered. Figure 2 shows the $K'_a = 0$ lines with $\Delta J \neq 0$, which can be seen to be approximately equally spaced, a consequence of the energy levels approximately following $E = \bar{B}J(J+1)$ where $\bar{B} = \frac{1}{2}(B + C)$. The formula is not exact, but over a small range of J a reasonable fit can be obtained by adjusting $\bar{B}$ for the upper state, assisted by the use of estimated values for other constants that will predict some of the deviations from linearity. Given the density of lines, even an experienced spectroscopist is very unlikely to be able to pick out this $K'_a = 0$ sequence, but it makes a good candidate for an automatic search. In this walk through all the transitions with $J' = 10 - 20$ were included in the search process, with $^pR_{1,9}(9)$ and $^pR_{1,19}(19)$ as the fit transitions, and the remaining 20 as the check transitions. This allows two parameters to be found by the search process, specifically $\bar{B}'$ and origin which are sufficient to fit the $K'_a = 0$ lines over a small range. Given a search window, w_s, of 1 cm^{-1} (approximately the accuracy to which the band origin can be estimated) and an acceptance window, w_a, of 0.001 cm^{-1} (the linewidth) leads to 5×10^5 trials which takes about 5 minutes to work through on a standard desktop computer. The top 5 results of this search are shown in Table 1, ranked by the mean observed − calculated values for the check transitions, σ_{check}, where at most one check transition is not found.

Table 1 Top 5 results of the autofit search on $K'_a = 0$ with $J' = 10 - 20$ for the ν_5 band of *cis*-1,2-dichloroethene

n_{OK}	σ_{check}/cm^{-1}	I_{check}	Origin	$\bar{B}'$ / cm^{-1}	Trial number
20	.000159	1.7131	168.672175	.0770223194	408789
19	.000349	.8806	168.420107	.0780489217	452779
19	.000353	.9765	167.507875	.0804024463	470713
19	.000367	.7926	166.737863	.0802260692	254455
19	.000370	.7787	167.475687	.0779197132	195896

In this case further analysis shows that the top ranked fit is correct, but this is by no means obvious from the table above. At first glance all the assignments are reasonable, with average errors < 1/3 of the line width, reasonable total intensity (I_{check}) and rotational constants. Simply looking at normal simulations with these constants does not provide much

discrimination, and at this point the nearest lines plot is invaluable. Consider the nearest lines plots for $K'_a = 0$ lines shown in Figure 3. Figure 3 (a) uses the constants from the top row in Table 1 and Figure 3 (b) the constants from the second row. The lighter symbols correspond to the $n_{OK}+2$ transitions assigned by the automatic process. The first plot shows a clear line extending beyond the assigned region indicating a set of assignments to higher J which will show a clear trend in residuals, strongly suggesting this is the right assignment. The second plot shows no such trend and indicates that the assignment is simply chance. The plots do serve to indicate that such a chance assignment is quite likely given the line density; almost any curve drawn on the plot will have a transition within a linewidth for most observation numbers, but the top plot shows an assignment giving a much better error than random chance.

Given the successful initial assignment an additional program feature allows the set of assignments to be rapidly extended as each point on the curve in Figure 3 (a) corresponds to a good assignment. The required points are selected using a mouse, and the program can then add the selected assignments to the line list used for fitting. Adding the obvious points and re-fitting allows the assignment to be rapidly extended to $J' \sim 60$ with an average error (after re-fitting the origin and $\bar{B}'$) of 0.0030 cm^{-1}. The resulting nearest lines plot still shows come curvature reflecting the fact that $\bar{B}'J(J+1)$ is only approximately correct; floating both $\delta' = B' - C'$ and D'_J reduces the error and allows the fit to be extended to $J' \sim 85$ with an average error of 0.0013 cm^{-1} at which point the curve on the nearest lines plot fades out as the transitions weaken. Overall it indicates the subset chosen for the search is sensible; deviations from the levels calculated with approximate constants are smooth and the range of J chosen for the search is not likely to be critical. There is nothing special about the choice made here and it is also likely that other subsets of lines would give good starting points for searches; $K'_a = 0$ lines are likely to be a good starting point in general as they are typically the strongest, are spread over a wide region avoiding band heads, and are least likely to be affected by perturbations but other K'_a or J' ranges are also likely to give good results.
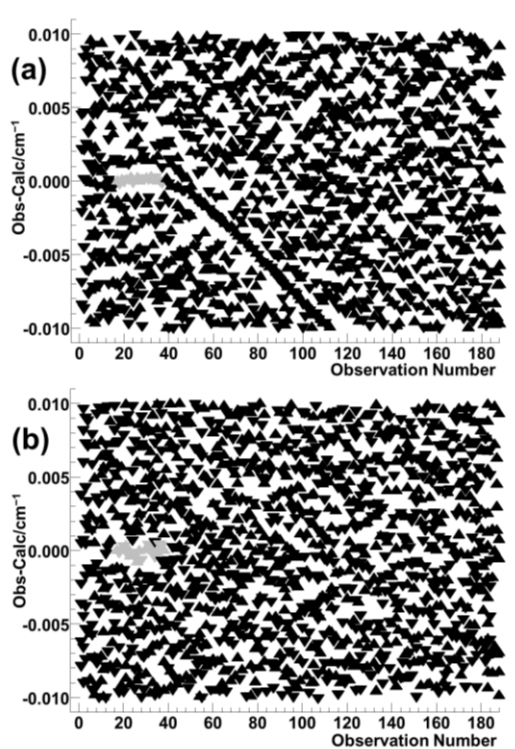
Figure 3 Nearest lines plot for two trial simulations of $K'_a = 0$ lines in the ν_5 band of *cis*-1,2-dichloroethene. (a) is a successful trial, (b) an unsuccessful one. The greyed-out points indicate lines included in the trial fits.

The nearest lines plots can give even clearer results if the combination difference filter can be applied. Figure 4 shows this, derived from Figure 3 (a) by only plotting points satisfying a ground state combination difference. This additional selectivity is clearly not essential, even with the high line density considered here, but provides additional selectivity to the assigned lines.
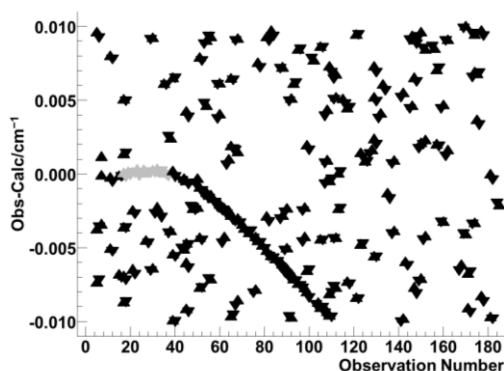
Figure 4 Nearest lines plot as Figure 3a, but requiring plotted assignments to satisfy ground state combination differences.

Given that most of the $K'_a = 0$ lines have been assigned, the obvious next step is to consider $K'_a = 1$. In principle the automatic search process could be used for $K'_a = 1$ but in practice the nearest lines plot will typically indicate the correct assignments immediately. Figure 5 shows a nearest lines plot including both $K'_a = 0$ and 1 lines using the constants from

the previous step. The horizontal axis is $K'_a + J'/J_{max}$, chosen to separate points by K'_a and also indicate the trend with J'. The left half of the plot shows the good fit to $K'_a = 0$ lines described above, and the right half of the plot suggests two possible sets of assignments for $K'_a = 1$. In fact these correspond to the two possible symmetries for $K'_a = 1$ states, and separate nicely if the points plotted are further selected by symmetry. Selecting the assignments corresponding to these two curves allows the assigned line list to be extended to $K'_a = 1$, and now all three rotational constants (A, B and C) can be determined.
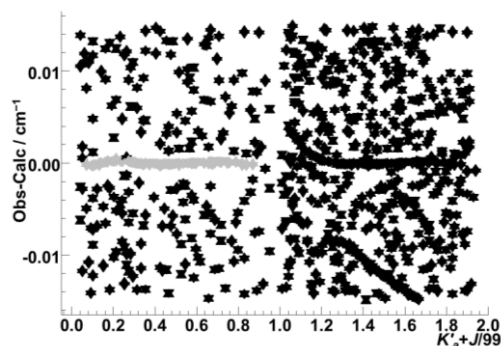

Figure 5 Nearest lines plot $K'_a = 0$ lines (left half) and $K'_a = 1$ lines (right half) for the ν_5 band of *cis*-1,2-dichloroethene. All the points plotted satisfy ground state combination differences.

Repeating the process for other K'_a values is straightforward; adding $K'_a = 2$, 10, 20 and 30 to the line list for fitting gives enough data to allow the five quartic centrifugal distortion parameters to be determined. At this stage the nearest lines plot (Figure 6) gives sensible results if all the transitions are included. Note the very small vertical range ($\pm$ 3×line width) and the horizontal axis is again $K'_a + J'/J_{max}$, separating transitions by J and K. The dark area in the middle indicates many observed lines have assignments within 1/3 of the line width, and these can all be added to the fitting line list from the plot. This allows some 8000 transitions to be added in one go and gives a good quality final fit with an average error of 0.00016 cm^{-1}.
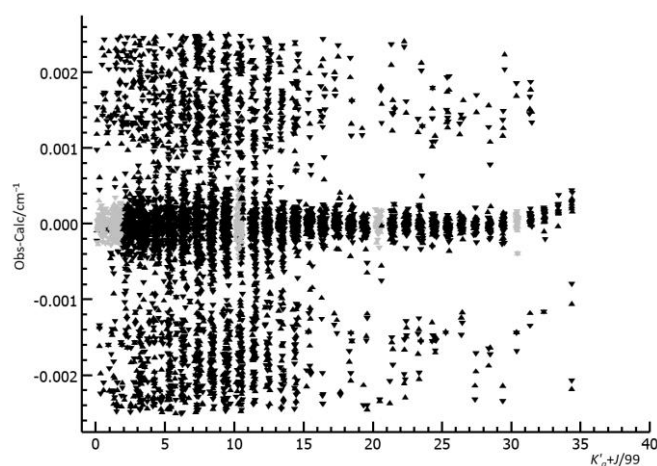

Figure 6 Nearest lines plot including all lines for the ν_5 band of *cis*-1,2-dichloroethene.

Table 2 Rotational constants[a] (/cm⁻¹) for cis 1,2-dichloroethene

	Ground State[b]	v_5	$2v_5$	$3v_5$	v_{11}	v_5+v_7
$C_2H_2{}^{35}Cl_2$						
Origin	0	168.6726210(65)	337.745260(10)	507.221000(16)	570.7668047(74)	574.62436(199)
A	0.384205580(33)	0.386388268(66)	0.38857336(11)	0.39076078(17)	0.38388703(12)	0.3810981(722)
$\frac{1}{2}(B+C)$	0.0771851191(88)	0.0769963157(97)	0.076813261(16)	0.076635677(25)	0.0771620861(78)	0.07748553(256)
$(B-C)\times10^3$	15.43212386(635)	15.395420(25)	15.361146(39)	15.329018(65)	15.543830(23)	14.8480(11)
$\Delta_K\times10^6$	1.3800964(374)	1.41649(15)	1.45364(29)	1.49058(43)	1.32723(144)	-3.0189(671)
$\Delta_{JK}\times10^7$	-3.669745(245)	-3.65065(44)	-3.63145(72)	-3.6142(11)	-3.601342(908)	2.6421(551)
$\Delta_J\times10^8$	4.778408(388)	4.66975(39)	4.56461(69)	4.4680(10)	4.785903(291)	3.495(33)
$\delta_K\times10^8$	9.84208(651)	10.1969(53)	10.5350(90)	11.017(13)	10.201(12)	c
$\delta_J\times10^8$	1.186577(121)	1.15901(28)	1.13320(46)	1.10872(76)	1.19125(20)	c
$\Phi_K\times10^{11}$	2.37503(355)	2.484(13)	2.560(24)	2.481(38)	-4.736(139)	c
$\Phi_{KJ}\times10^{12}$	-8.4906(327)	-8.74(13)	-8.484(74)	-6.96(30)	2.985(132)	c
$\Phi_{JK}\times10^{13}$	6.115(108)	6.72(37)	5.79(13)	1.384(827)	1.595(144)	c
$\Phi_J\times10^{15}$	32.720(514)	33.19(43)	34.44(86)	39.7(12)	33.399(296)	c
$\phi_K\times10^{12}$	9.197(244)	11.3(10)	c	-3.6(23)	4.494(423)	c
$\phi_{JK}\times10^{14}$	6.41(165)	c	c	c	5.7(15)	c
$\phi_J\times10^{14}$	1.6300(203)	1.653(31)	1.789(56)	1.921(89)	1.673(17)	c
					$<v_{11}\|J_a\|v_5+v_7>$	0.013792(11)K
n_{obs}	176	14872	7699	5479	5017	
σ	1.3×10⁻⁶	0.00016	0.00022		0.00013	
$C_2H_2{}^{35}Cl^{37}Cl$						
Origin	0	166.5371047(73)	333.4665077(95)		568.386926(10)	572.91066(96)
A	0.3815011123(430)	0.383639458(60)	0.385779349(77)		0.38117105(12)	0.3809632(182)
$\frac{1}{2}(B+C)$	0.0753387724(121)	0.0751574129(84)	0.074981394(11)		0.075318164(11)	0.07526325(21)
$(B-C)\times10^3$	14.81160962(994)	14.777830(26)	14.746126(36)		14.921442(29)	14.76001(89)
$\Delta_K\times10^6$	1.3575546(425)	1.39256(13)	1.42850(17)		1.34721(33)	c
$\Delta_{JK}\times10^7$	-3.566218(327)	-3.54242(31)	-3.52702(41)		-3.55417(33)	c
$\Delta_J\times10^8$	4.580192(559)	4.47786(34)	4.37977(49)		4.58807(30)	4.5946(16)
$\delta_K\times10^8$	9.51253(755)	9.396(17)	9.845(27)		9.5762(40)	c
$\delta_J\times10^8$	1.120267(209)	1.09532(24)	1.07054(35)		1.12474(24)	c
$\Phi_K\times10^{11}$	2.3260(124)	1.960(11)	2.108(17)		1.836(16)	c
$\Phi_{KJ}\times10^{12}$	-8.1669(1509)	-2.14(10)	-2.95(19)		-7.273(73)	c
$\Phi_{JK}\times10^{13}$	5.901(455)	-10.56(28)	-8.870(520)		5.160(202)	c
$\Phi_J\times10^{15}$	32.768(779)	26.39(38)	27.16(59)		32.70(23)	c
$\phi_K\times10^{12}$	9.135(1260)	-40.73(82)	-35.66(154)		6.923(594)	c
$\phi_{JK}\times10^{14}$	3.27(187)	-119.7(24)	-126.5(41)		c	c
$\phi_J\times10^{14}$	1.4810(345)	1.271(23)	1.257(34)		1.474(20)	c
					$<v_{11}\|J_a\|v_5+v_7>$	0.016635(11)
n_{obs}	148	13932	6809		4196	
σ	1.3×10⁻⁶	0.00018	0.00020		0.00017	

[a] Figures in brackets are one standard deviation in units of the last significant figure. The calculations use the Ir representation with Watson's A reduced centrifugal distortion Hamiltonian[21]

[b] Refit to millimetre wave transitions measured by Leal *et al*[17]

[c] Constrained to ground state value

Table 2 (Continued) Rotational constants[a] (/cm^{-1}) for cis 1,2-dichloroethene

	v_{12}	v_4	v_5+v_{11}	$2v_5+v_7$	v_{10}	v_5+v_{12}
$C_2H_2{}^{35}Cl_2$						
Origin	695.7468747(72)	723.3931772(76)	739.378190(12)	743.7827(19)	858.866175(32)	864.4235(58)
A	0.382409678(63)	0.385757366(76)	0.38607076(27)	0.382759(131)	0.38328164(71)	0.38326(11)
$\tfrac{1}{2}(B+C)$	0.077190574(10)	0.076829393(11)	0.076972369(15)	0.07729536(245)	0.077118924(39)	0.0769542(31)
$(B-C)\times10^3$	15.397360(48)	15.249889(33)	15.506053(29)	14.8359(20)	15.504067(80)	15.2552(79)
$\Delta_K\times10^6$	1.24617(14)	1.55164(16)	1.32347(420)	-4.615(129)	1.3466(42)	c
$\Delta_{JK}\times10^7$	-3.46971(56)	-3.90922(64)	-3.55205(194)	2.397(439)	-3.7404(34)	c
$\Delta_J\times10^8$	4.772546(880)	4.786683(532)	4.676768(357)	3.5642(681)	4.78963(90)	c
$\delta_K\times10^8$	8.2136(487)	10.718(30)	10.5373(87)		9.065(16)	c
$\delta_J\times10^8$	1.20172(67)	1.18694(39)	1.16355(11)	c	1.19654(45)	c
$\Phi_K\times10^{11}$	1.0356(464)	4.1410(278)	-10.708(338)	c	-3.40(67)	c
$\Phi_{KJ}\times10^{12}$	-7.970(652)	-13.536(356)	5.851(900)	c	-12.33(70)	c
$\Phi_{JK}\times10^{13}$	15.886(2018)	12.76(101)	3.464(344)	c	6.68(52)	c
$\Phi_J\times10^{15}$	-14.26(214)	9.24(82)	32.63(27)	c	23.75(73)	c
$\phi_K\times10^{12}$	37.36(523)	13.73(295)	c	c	c	c
$\phi_{JK}\times10^{14}$	-330.7(142)	-37.9(65)	c	c	c	c
$\phi_J\times10^{14}$	5.04(12)	0.394(46)	$\langle v_5+v_{11}|J_a|2v_5+v_7\rangle$	0.019453(22)	c	c
	$\langle v_{12}|J_bJ_c|v_4\rangle$	6.9807(35)×10^{-5}	$\langle v_5+v_{11}|J_c|v_4\rangle$	4.1288(33)×10^{-3}	$\langle v_5+v_{12}|J_b|v_{10}\rangle$	2.3791(76)×10^{-3}
n_{obs}	5642	5384	1924		3269	
σ	0.00015				0.00049	
$C_2H_2{}^{35}Cl^{37}Cl$						
Origin	695.656721(15)	720.225007(15)	734.873473(25)	740.0647(15)	857.098091(41)	862.013(14)
A	0.37957075(16)	0.38320639(17)	0.383291811(880)	0.380107(498)	0.38061641(43)	0.38218(26)
$\tfrac{1}{2}(B+C)$	0.075344085(37)	0.074990157(27)	0.075135760(28)	0.07504419(851)	0.075274859(58)	0.0751707(24)
$(B-C)\times10^3$	14.77788(12)	14.633305(45)	14.886282(54)	14.9144(132)	14.88092(12)	14.709(17)
$\Delta_K\times10^6$	1.185905(477)	1.56924(61)	1.3266(128)	-7.18(137)	1.35897(88)	c
$\Delta_{JK}\times10^7$	-3.32641(369)	-3.8368(12)	-3.52532(344)	-4.202(570)	-3.5851(39)	c
$\Delta_J\times10^8$	4.58154(390)	4.5458(14)	4.475010(946)	4.378(90)	4.6129(20)	c
$\delta_K\times10^8$	7.992(276)	9.079(35)	9.097(32)		9.174(42)	c
$\delta_J\times10^8$	1.1322(20)	1.09638(38)	1.09503(33)	c	1.13475(100)	c
$\Phi_K\times10^{11}$	-1.474(329)	4.509(63)	-11.76(160)	c	c	c
$\Phi_{KJ}\times10^{12}$	17.684(4455)	c	c	c	c	c
$\Phi_{JK}\times10^{13}$	-57.08(1244)	-4.06(36)	c	c	17.40(92)	c
$\Phi_J\times10^{15}$	24.6(37)	29.5(18)	30.47(70)	c	c	c
$\phi_K\times10^{12}$	-161.8(357)	c	c	c	c	c
$\phi_{JK}\times10^{14}$	c	c	c	c	c	c
$\phi_J\times10^{14}$	c	c	$\langle v_5+v_{11}|J_a|2v_5+v_7\rangle$	0.020206(50)	c	c
	$\langle v_{12}|J_bJ_c|v_4\rangle$	7.673(16)×10^{-5}	$\langle v_5+v_{11}|J_c|v_4\rangle$	4.2266(50)×10^{-3}	$\langle v_5+v_{12}|J_b|v_{10}\rangle$	2.280(12)×10^{-3}
n_{obs}	3650	3427	1269		2361	
σ	0.00020				0.00060	

Given the ease noted above with which a spurious set of assignments can give an apparently reasonable fits further checks are appropriate. Simulation of selected parts of the experimental spectrum indicate good agreement, as shown in Figure 7. Some discrepancies are expected at this stage, as the $^{35}Cl^{37}Cl$ species would be expected to have similar intensity.
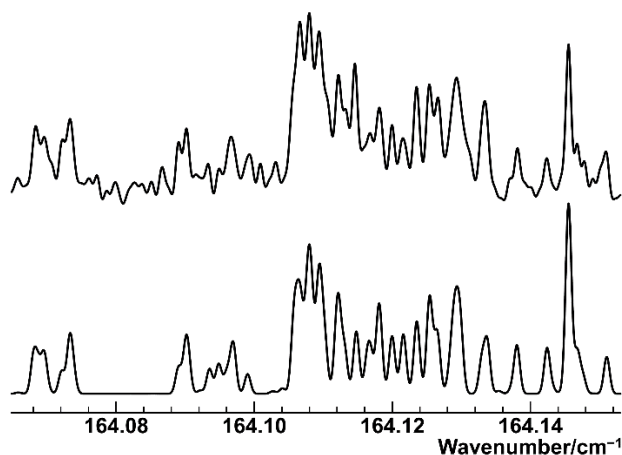
Figure 7 Simulation (lower trace, including $^{35}Cl_2$ cold band only) and experimental spectrum (upper trace) for a small part of the v_5 band of *cis*-1,2-dichloroethene.

The process above has assumed ground state constants are available, which will not always be the case. The same process will work in the absence of ground state constants, but the first step is potentially rather slower. The initial search must now be a three parameter fit (Origin, $\bar{B}'$ and $\bar{B}''$) so an additional fitted transition is required, and the number of trials rises sharply. If the $^PP_{1,21}$ transition is added to the initial search described above the number of trials required rises to 4×10^8 if nothing else is changed, perhaps 1 day on a fast desktop. This is still useful, though running on a mainframe computer is then more appropriate. If more care is taken in setting up the calculation the run time can be significantly reduced; for example limiting the search range on rotational constants to 1% of their values results in more than 98% of the trials being rejected and reduces the search time to ¾ of an hour on a fast desktop machine. This specific search gives the correct assignment as the top trial.

Assigning and fitting transitions from $^{35}Cl^{37}Cl$ is the obvious next step and repeating the procedure above for this species yields a similarly extensive set of assignments for this species. The only difference worthy of note in the process is the initial search process; searching for the same $K'_a = 0$ lines as before does work, but the correct trial is number 10 on the list, rather than the first. Making small changes to the fit and check transitions can switch this to top of the list or push it much further down. The experience over the 31 vibrational transitions analysed in this paper indicates that the correct assignment is found after a few searches, if not the first one, and doing several searches is anyway straightforward with the interactive system. There is probably scope for adjusting the search strategy to automate this, but this is left as a project for future work.

Having assigned the two most abundant isotopologues, there are still unassigned lines. The $^{37}Cl_2$ species was not found, unsurprising giving an abundance of 6% compared to 56% for $^{35}Cl_2$, but $v_5 = 2 \leftarrow v_5 = 1$ hot bands for both the $^{35}Cl_2$ species (1/3 of the intensity of the cold band) and the $^{35}Cl^{37}Cl$ species (1/5 the intensity of the strongest band) were also assigned and analysed using the same procedure. (For the hot bands the lower state constants are constrained to the values from the cold band fits.) We were also able to tentatively identify the $v_5 = 3 \leftarrow v_5 = 2$ hot band (1/10 the intensity of the strongest band) though this is sufficiently weak that the number of isolated lines that are mainly from this species is small

and many of the lines used will be blends. However, the overall simulation including all 5 bands is excellent – see Figure 8 for a 0.03 cm^{-1} region that shows some clear lines from the weakest band.
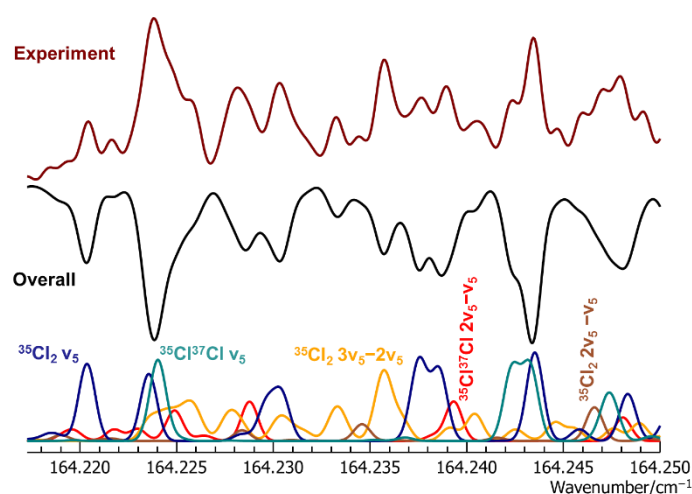

Figure 8 Simulation and experimental spectrum (upper trace) for a small part pf the v_5 band of *cis*-1,2-dichloroethene. The middle trace is an (inverted) simulation summing all isotopologues and hot bands, and the lower traces shows the individual components.

The above provides an outline of the process; a detailed walk through of the process used here, including screen shots, the spectrum used, intermediate and final data files and a line list has been added to the set of files included with the version 10.1 PGOPHER distribution. This is freely available from the website[22] and has been deposited in the University of Bristol data repository[19].

Application to other bands of *cis*-1,2-dichloroethene

To further test the procedure described above we also applied the above process to the 4 other bands available from the spectra recorded, with similarly good results, though most of the higher energy bands are affected by perturbations which slightly complicates the analysis process but is still straightforward with the tools described here.

We started by revisiting the v_{11} band at 571 cm^{-1}, which was the subject of our preliminary communication on automatic fitting[5]. The earlier work was done without the nearest lines plot and relied on simply comparing simulation and experiment over small regions showing characteristic sub-bands. This made it difficult to pick out progressions at higher *J* as the sub-bands merge. As a result, the initial analysis was only able to assign lines with $J' < 40$ and $K'_a < 30$ for the $^{35}Cl_2$ species, and $J' < 30$ and $K'_a < 25$ for the $^{35}Cl^{37}Cl$ species. Also, some evidence for local perturbations was found, but not analysed. To validate the process presented here, a fresh analysis was performed on the room temperature spectra analysed in the previous work. As for the v_5 band, initial constants were taken as the ground state constants with some manual adjustments to the origin and $\bar{B}'$ to give a qualitative match to the overall spectrum. This band is easier to work with, as there is a central Q branch and the origin can easily be identified from the spectrum to better than 0.1 cm^{-1}, and the automatic search runs 10 times faster if this is used as a constraint. (That constraint also

moves the correct fit into top position from number 15.) Assignment and fitting of lines with J' up to 98 for $K'_a = 0$ was then straightforward, following the process in the previous section. The nearest lines plot for $K'_a = 1$ lines with the resulting constants, shown in Figure 9 was then very revealing. Figure 9 (a) shows upper states with one specific symmetry, with upper states $J_{1,J}$, where the model was already giving good results. The plot of the other possible symmetry for $K'_a = 1$ states ($J_{1,J-1}$), Figure 9 (b), shows shows clear evidence of a local perturbation around $J' = 37$, a region above the earlier assignments.
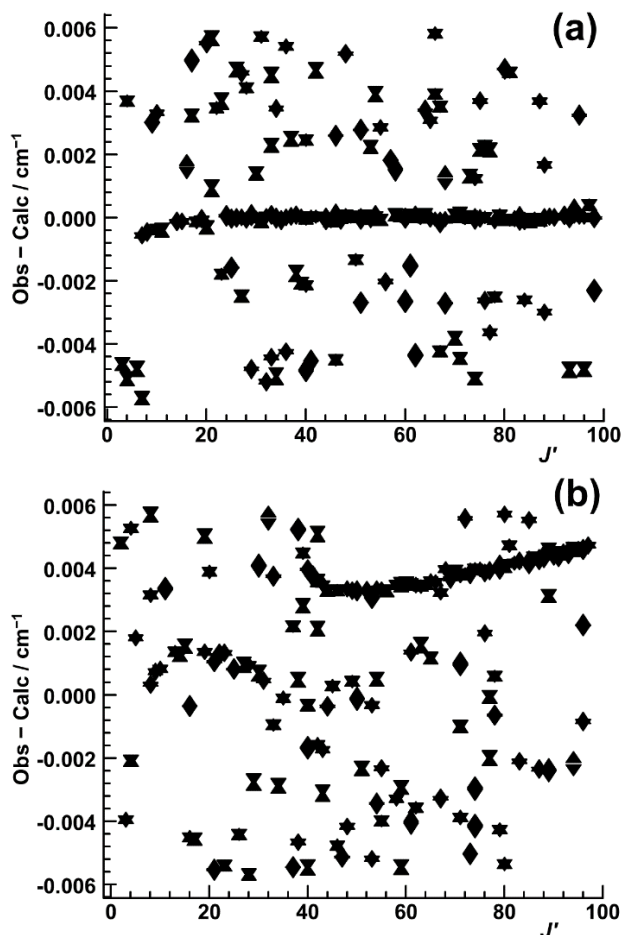

Figure 9 Nearest lines plot for $K'_a = 1$ lines for the ν_{11} band of *cis*-1,2-dichloroethene. (a) shows $J_{1,J}$ levels, which follow the standard model whereas (b) shows $J_{1,J-1}$ levels indicating a local perturbation peaking at $J' = 37$.

As the perturbation affects low K'_a the perturbing state must have an origin close to that of ν_{11}, and considering the possible[23] states by far the closest state is $\nu_5+\nu_7$. This has A_2 symmetry overall and, as the ν_{11} state has B_1 symmetry a J_a type Coriolis interaction will mix the two states. The $\nu_5+\nu_7$ state was therefore added to the Hamiltonian matrix in the upper state, and a matrix element proportional to $\delta_{K,K'}K$ added between the states. The initial rotational constants for $\nu_5+\nu_7$ were again taken from the ground state, and the origin adjusted to give a crossing at upper state level $36_{1,35}$. Fitting the origin and the interaction gave good results for all the $K'_a = 1$ transitions, and the analysis was easily extended to higher K'_a. It was not quite as straightforward as for ν_5, as some care was required at several stages to ensure the correct state ordering between ν_{11} and $\nu_5+\nu_7$ for each K'_a, but the analysis was extended

up to J' < 100, K'_a < 30 for v_{11}, limited only by population of the lower state levels. The final constants are given in Table 2. Only the main constants could be determined for the v_5+v_7 state, as the mixing between states was not strong enough to make transitions directly to the v_5+v_7 state visible, and as a result only the position of selected rotational levels of v_5+v_7 were determined. The final fit is as good as for the v_5 band. The analysis was confirmed by repeating for the $^{35}Cl^{37}Cl$ species, using exactly the same method, and also resulted in a good fit. It was also possible to identify the $v_5+v_{11}-v_5$ hot band for each species and analyse as for the cold bands including an interaction between the v_5+v_{11} and $2v_5+v_7$ states though the full analysis of these hot bands required consideration of v_4 and v_{12}, discussed below.

Comparison of the procedure used here with that in the preliminary paper[5] is instructive. The key difference is that the earlier method relied on being able to compare simulations directly, which for a spectrum as complicated and extensive as studied here means just looking at a small region. This is possible for some regions of the v_{11} band, which show clear sub-bands from only one species, but the range of application even to v_{11} is limited, as most transitions do not appear grouped into sub-bands, and this explains why the earlier analysis only covered some of the band. Direct comparison of simulations for the v_5 band as a whole, where only some transitions are correct, is not possible, as it is not possible to pick out just the transitions that should be correct. The nearest lines plot provides a much better way of testing simulations, taking into account the entire range of the spectrum. It is also instructive to consider the nearest lines plot applied to constants from the previous work. Figure 10 shows the same plots as Figure 9, but using the constants from the previous work. (Note also the slightly different vertical scale) Figure 10 (a) shows that the assignment of the unperturbed K'_a = 1 lines can be smoothly continued above the J' = 40 limit of the previous work. However, Figure 10 (b) shows that, while effective constants have resulted in a good fit below the perturbation crossing point, the resulting constants have no predictive value above the crossing point, and it is very difficult to pick out the correct assignment from the nearest lines plot. The correct assignments have been greyed out in Figure 10 (b) as otherwise they could not be distinguished from the background of random points. This illustrates a fundamental requirement of the nearest lines plot, specifically of the constants used; the calculated assignment sequences must appear reasonably close to horizontal on the plots for the eye to be able to pick them out. A steeply sloping curve will only have a few points visible if the zoom is high, and at compressed vertical scales there are simply too many points.
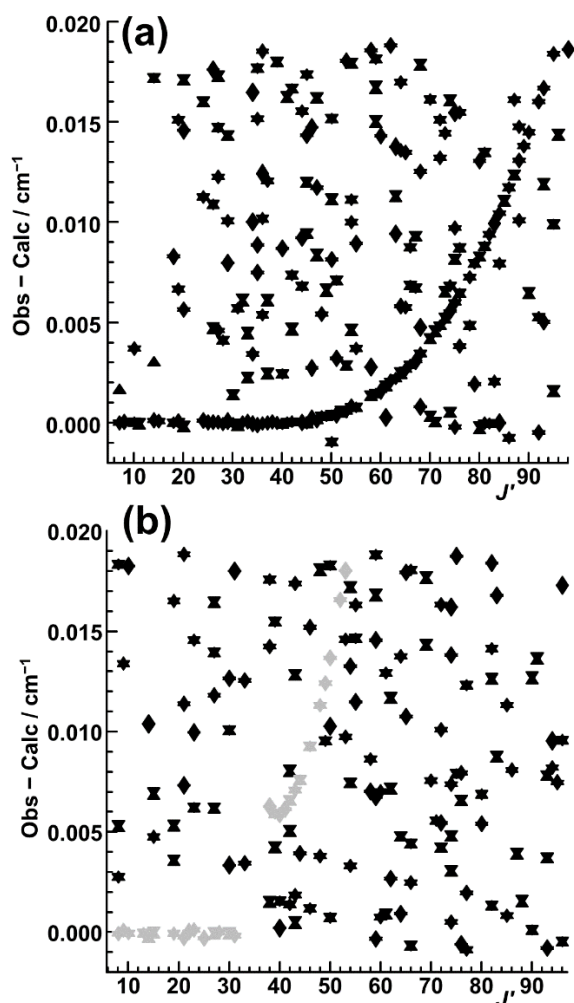
Figure 10 Nearest lines plot for $K'_a = 1$ lines for the ν_{11} band of *cis*-1,2-dichloroethene, using the constants in the preliminary paper. (a) shows J_{1J} levels, and (b) shows $J_{1\,J\text{-}1}$ levels, with the correct assignments indicated in lighter colours in (b).

At this point, it is also possible to compare assignments using the more traditional Loomis-Wood plot[12, 13]. Figure 11 shows a Loomis-Wood plot that might be used to assign the unperturbed ${}^qP_{1,J}(J)$ branch that makes up half the lines in Figure 9 (a) and Figure 10 (a). The bottom plot is a small portion of the experimental spectrum centred on the ${}^qP_{1,28}(28)$ transition and the plots above are similar portions with the centre shifted down in frequency by 0.143 cm^{-1} from the plot below, which centres the plots on successive transitions of the same branch, i.e. ${}^qP_{1,29}(29)$, ${}^qP_{1,30}(30)$.... If the Loomis-Wood approach was working well there should be an obvious vertical sequence, but the line density is such that it is difficult to pick out a branch by eye. In addition, the issue noted above for the nearest lines plot also applies to Loomis-Wood plot, in that the lines must be reasonably close to the vertical to be able to pick them out from the sea of other lines. The automated search procedure used here assists with this, and could indeed be used in conjunction with conventional Loomis-Wood plots though the approach here is much more streamlined.

There are four important improvements in the nearest lines plot as compared with the Loomis-Wood approach; firstly, all transitions from different branches are plotted

18

together so, in this case, P and R branch lines with the same upper state are plotted together, increasing the selectivity. Secondly the calculated line positions from the current trial Hamiltonian are used, rather than an assumed polynomial, which is particularly helpful in moving to related branches, and in the presence of perturbations. In addition, several related sets of states can be examined simultaneously, as in Figure 5, or even the entire band, as in Figure 6. Finally, the combination difference filter can be used to reduce the number of spurious lines plotted.
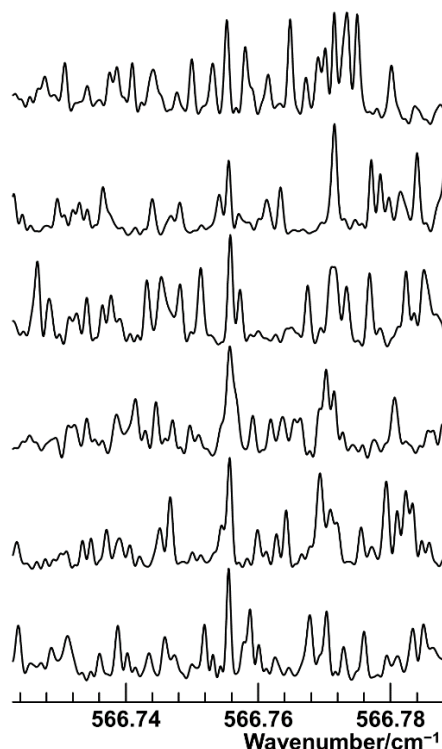

Figure 11 Loomis Wood plot for the v_{11} band of *cis*-1,2-dichloroethene; the lowest plot is cantered on the $^qP_{1,28}(28)$ transition and each successive plot is shifted down by by 0.143 cm^{-1}, which will centre the $^qP_{1,J}(J)$ branch transitions.

In the region covered in this work there are three other bands, all of which were assigned and fitted using the methods introduced here, and none have been previously recorded at high resolution or analysed. The v_{12} band at 696 cm^{-1} and the v_4 band at 723 cm^{-1} are strong, and low K'_a levels could be easily assigned for both. However these bands are sufficiently close in energy that there is some interaction between them, and detailed assignment indicated a further interaction with the v_5+v_{11} and $2v_5+v_7$ pair of states mentioned above. The final model thus included four interacting states, and assignments from the $v_5+v_{11}-v_5$ hot band were included in the final fitted line list.

Also analysed was the v_{10} band at 859 cm^{-1}. Assignment was again straightforward, and again the band was perturbed. One of the perturbing states could be assigned as v_5+v_{12}, but the discussion above implies mixing with v_5+v_4, $2v_5+v_{11}$ and $2v_5+v_7$ is to be expected. Our model was limited to the v_{10} and v_5+v_{12} states, resulting in a fit showing some clear residual errors at high K'_a, though the overall average error was no more than ½ the line width. A global model might well be able to improve this.

The final constants for all the states discussed above are given Table 2. These are the results of a selection of fits combining observations from more than one band. For example, the v_5 constants result from a fit to the fundamental band frequencies, and combination differences from other bands with v_5 as the lower state. Details of each fit, including the line list and residuals, are available in the supplementary data[20]. Overall the fits include 116,351 transitions from 23 vibrational bands fitting with an average error of 0.0002 cm^{-1}.

Application to *trans*-1,2-dichloroethene

We also took a series of spectra covering the same region for *trans*-1,2-dichloroethene. As the molecule has no dipole moment no microwave spectra are available, but there have been two studies of the v_6 band at 898 cm^{-1}[15, 16]. The first paper [15] was a conventional room temperature Fourier transform study at a resolution of 0.002 cm^{-1} covering $C_2H_2{}^{35}Cl_2$, $C_2H_2{}^{35}Cl^{37}Cl$, $^{13}C_2H_2{}^{35}Cl_2$, $C_2D_2{}^{35}Cl_2$ and $C_2H_2{}^{35}Cl^{37}Cl$. Rotational constants were presented for the upper and lower states for each molecule but, while perturbations were noted, they were not included in the analysis. The second paper[16] used a special sample cell to obtain a low temperature (180 K) spectrum leading to an improved analysis for $C_2H_2{}^{35}Cl_2$ and $C_2H_2{}^{35}Cl^{37}Cl$ though again perturbed transitions were noted and excluded from fits of the upper states. In the context of the current work it should be noted that, of all the bands discussed in this paper, the v_6 band is the only one that is tractable by conventional spectroscopic trial and error, as the band type (*c* dipole, $\Delta K_a = \pm 1$) means it has a series of sharp K'_a sub-bands allowing the band origin and A to be readily determined for both states and making the subsequent analysis much easier. In contrast the adjacent (and stronger) v_{11} band at 829 cm^{-1} is like the v_5 band for the *cis* species in having no recognizable structure, and even though it is clear that the band was recorded by the earlier workers, no analysis was presented.

Starting with this latter band, and using the published ground state constants[16], the assignment and analysis was straightforward using the methods described above. A search for $J' = 10 - 20$, $K'_a = 0$ transitions for the main isotopologue yielded the correct assignment as the top candidate, though for this case the search range had to be extended (to ±2 cm^{-1}) as the lack of clear structure meant the band origin could only be estimated very approximately. The fit could be easily extended up to $K'_a = 13$ using the methods described above, leading to 3100 assigned transitions. The $^{35}Cl^{37}Cl$ isotopologue was more difficult, with searches for $K'_a = 0$ transitions not yielding good results. However, searching on $K'_a = 1$ transitions yielded good results immediately, and 2500 assignments extending to $K'_a = 13$ could be easily made. This indicated that the low J, $K'_a = 0$ lines were perturbed by up to 0.01 cm^{-1}, with some smaller shifts (of the order of the line width) for $K'_a = 1$ transitions. It is not immediately obvious what the perturbing state is, and there is not enough information about the state from the observations, so $K'_a = 0$ lines were excluded from the fit to give the constants in Table 3. For this band the hot bands were not obvious and, as indicated below, are likely to be complicated so an analysis of these was not attempted for this, or any of the other bands analysed for the trans species.

The v_6 band could also be analysed in a straightforward way using the methods introduced in this paper. The different band type here implies 6 branches for a given upper state, so there

is a wider choice of transitions for searching. For $K'_a = 0$ qualitative simulations indicated most of the branches were likely to be under the central band heads, so $J' = 15 - 25$ of the ^PP and ^rR branches was used for searching, with effective results. Extending the assignments as above confirmed, as found in the previous work, the presence of a perturbation with the greatest effect at $K'_a = 7$. The assignment was nevertheless straightforward, and it was possible to model the perturbation by introducing a state of Bu symmetry connected by a $\Delta K_a = 0$ matrix element proportional to K_a. The assignment of this state is not immediately obvious, so it is denoted "P" in Table 3. The state mixing for $K'_a = 7$ is sufficiently strong that transitions directly to the perturbing state can be seen, and these were included in the final model. A second perturbation for the very highest K'_a was also identified, peaking at $K'_a = 14$ though this was not modelled, and transitions with $K'_a \geq 13$ were excluded from the fit. The $^{35}Cl^{37}Cl$ isotopologue was similar, though here the $K'_a = 3$ transitions were the most perturbed. The analysis was similar, though the results (Table 3) were not quite as satisfactory; note the negative value for $B-C$ and a small systematic error for $K'_a = 0$. This suggests other states are perturbing v_6, perhaps as a result of the reduced symmetry, and the analysis for $^{35}Cl^{37}Cl$ must be regarded as preliminary, to be improved by a global analysis. As for the main isotopologue a second perturbation was identified at high K'_a and transitions with $K'_a \geq 13$ were excluded from the final fit. Overall the constants for both isotopologues were essentially consistent with the previous work, with differences down to the range of transitions fitted, increased from 2490 to 5989 lines for the most abundant isotopologue, mainly reflecting the inclusion of perturbed levels.

The v_6 band is a potential candidates for using the Automated Spectral Assignment Procedure (ASAP) technique[8] for assignment. The ASAP technique relies on picking out a set of transitions with a common upper state by cross-correlating a set of experimental spectra separated by ground state combination differences. Locations where a transition occurs for each displacement will reinforce and, with sufficient ground state combination differences, stand out, allowing easy assignment. The example given in the published description had 9 transitions to a single upper state. For perpendicular type bands like v_6 up to 6 strong transitions are present to each upper state, so the ASAP technique might be feasible. Experiments with cross-correlating the spectrum analysed here gave some indications of the method working, but not sufficient to give immediately obvious assignments. The ASAP method requires a good baseline, and in our spectra there are few regions with no absorption, which makes determining an accurate baseline hard and also limits the selectivity by cross-correlating. This suggests the line density is too high here for the ASAP technique to be effective. (The other bands considered in this paper have only 2 or 3 transitions to any given upper state which does not provide anything like the required selectivity for ASAP.)

The low frequency band around 230 cm^{-1} was the most awkward to analyse. In appearance it is a single band, with a clear series of K band heads visible to one side only. Attempts to fit the band heads gave large residuals with errors linear in K'_a. This can only be explained if there are two near degenerate states, mixed by a large term proportional to K_a, much as in the v_6 band discussed above, but affecting all levels rather than a limited range of K'_a. There are only two low frequency modes [23], v_7 at 212 cm^{-1} and v_{12} at 240 cm^{-1} and the implication is that the observed single band arises from both of these two states with a strong

Coriolis interaction between them. The Hamiltonian used therefore included these two states, and the analysis started with a rough fit to the band heads to determine the two band origins and the mixing matrix element, with the A rotational constants constrained to the ground state value. This gave a qualitatively reasonable simulation, provided significant intensity was given to each band with the relative signs adjusted to give cancelation of intensity at the longer wavelengths. The detailed assignment could then be made as before, starting with $K'_a = 0$ and working upwards. The fit is then straightforward, with one complicating factor, a strong correlation between the upper state A rotational constants and the mixing matrix element. This made the fit somewhat unstable, requiring a large number of figures needed to specify the error for these parameters for the main isotopologue in Table 3, given the use of Watson's sensitivity parameter[24] to control the precision of errors given. For the $^{35}Cl^{37}Cl$ species the fit would not converge if the mixing parameter was floated, and a value of 3.334(1) cm^{-1} for the mixing matrix element was determined from the minimum error from a series of fits where the value was constrained to a selection of values around 3.33 cm^{-1}. The error bars on some of the other parameters must therefore be taken as significant underestimates. The final fits are good, given this proviso.

As the starting ground state constants were based on data with similar precision, the ground state rotational constants were re-determined by fitting to ground state combination differences using data from all three band analyses described above. The results are given in Table 3, and are essentially consistent with the values from Wugt Larsen et al[16], with error bars smaller by a factor of 5 – 10. The final set of assignments included 42,025 transitions from 8 vibrational bands.

Table 3 Rotational constants[a] (/cm^{-1}) for trans 1,2-dichloroethene

C$_2$H$_2$^{35}Cl$_2$	Ground State	v$_7$	v$_{12}$	v$_{11}$	v$_6$	P
Origin		212.0044384(81)	239.6894451(63)	829.1623114(86)	897.9507126(51)	907.578394(7787)
A	1.74726275(25)	1.74577780(19912)	1.74874390(19910)	1.74468086(64)	1.73796534(36)	1.5388485(1589)
½(B+C)	0.050826028(19)	0.050878393(15)	0.0508388214(296)	0.0506303785(68)	0.0508385687(30)	0.050824946(33)
(B−C)×10^3	1.4904515(86)	1.382199(213)	1.462260(178)	1.481644(16)	1.456396(17)	1.2334(72)
Δ$_K$×10^5	1.90753(20)	0.75242(230)	3.08460(219)	1.88128(173)	1.92141(72)	c
Δ$_{JK}$×10^7	−1.60881(43)	−0.1550(1934)	−3.0636(1940)	−1.6140(12)	−1.58217(31)	c
Δ$_J$×10^9	3.8846(18)	3.9003(25)	3.7996(16)	3.8896(16)	3.89249(56)	c
δ$_K$×10^8	3.21(10)	−22.72(59)	−32.86(32)e	3.75(16)	2.85(21)	c
δ$_J$×10^{10}	1.5380(30)	0.2590(81)	0.3701(63)	1.5487(41)	1.5023(41)	c
Φ$_K$×10^{10}	7.162(56)	−4.53(179)	21.90(172)	−23.94(174)	−9.17(36)	c
Φ$_{KJ}$×10^{11}	0	−3.84(18)	2.95(15)	0.136(58)	0	0
Φ$_{JK}$×10^{11}	0	2.8892(219)	−2.8619(229)	−0.0026(10)	0	0
Φ$_J$×10^{15}	0	5.66(18)	−6.419(97)	−0.493(90)	0	0
L$_K$×10^{12}	0		−1.497(71)	12.947(550)	0	0
					<P|J$_a$|v$_6$>	0.0185752(15)
					<P|J$_a$J^2|v$_6$>	2.600(68)×10^{-8}
n$_{obs}$	14285	12793^b		3100	5989	
σ	0.00011	0.00012		0.00012	0.00012	

C$_2$H$_2$^{35}Cl$_2$	Ground State	v$_7$	v$_{12}$	v$_{11}$	v$_6$	P
Origin		211.238643(15)	238.8267106(97)	826.042717(13)	897.947826(15)	900.60136(413)
A	1.74664962(33)	1.74234597(90)	1.75088101(68)	1.74430237(71)	1.73733415(96)	1.446529(461)
½(B+C)	0.049487094(30)	0.049542724(153)	0.049507646(124)	0.0492972396(51)	0.049499412(11)	0.04951038(156)
(B−C)×10^3	1.414879(37)	1.37042(40)	1.45528(25)	1.407365(39)	1.376174(87)	−1.3178(57)
Δ$_K$×10^5	1.90086(25)	0.05237(214)	3.72831(197)	1.9700(10)	1.9128(16)	c
Δ$_{JK}$×10^7	−1.56387(48)	−2.9796(1213)	−0.1368(1190)	−1.5501(13)	−1.5774(13)	25.73(171)
Δ$_J$×10^9	3.6721(33)	3.7322(62)	3.6503(41)	3.6864(13)	3.3151(45)	3.433(74)
δ$_K$×10^8	14.42(49)	11.91(104)	12.86(101)	17.94(51)	−128.9(13)	c
δ$_J$×10^{10}	1.750(17)	1.136(32)	1.357(31)	1.953(16)	1.492(45)	c
Φ$_K$×10^{10}	5.684(52)	−53.506(824)	45.536(562)	29.71(43)	−14.61(74)	c
Φ$_{KJ}$×10^{11}	0	4.816(466)	−5.771(263)	1.121(87)	0	0
Φ$_{JK}$×10^{13}	0	5.86(53)	−2.920(566)	0	0	0
Φ$_J$×10^{15}	0	0.48(51)	−1.17(28)	0	0	0
					<P|J$_a$|v$_6$>	0.0421995(77)
					<P|J$_a$J^2|v$_6$>	3.215(56)×10^{-7}
n$_{obs}$	12954	12948^b		2394	4294	
σ	0.00016	0.00018		0.00020	0.00025	

[a] Figures in brackets are one standard deviation in units of the last significant figure. The calculations use the Ir representation with Watson's A reduced centrifugal distortion Hamiltonian[21]

[b] Pair of interacting states with the following perturbation matrix elements (/cm^{-1}):

	C$_2$H$_2$^{35}Cl$_2$	C$_2$H$_2$^{35}Cl^{37}Cl	
<v7|J$_a$|v12>	3.34611334(82297)	3.334	δ$_{KK'}$K
<v7|J$_b$|v12>	0.0198050(296)	−0.016177(235)	δ$_{K±1K'}$ ½(J(J+1)−KK')^{½}
<v7|J$_a$J$_c$|v12>	−7.988535(3863)×10^{-3}	−1.51512(4041)×10^{-3}	δ$_{K±1K'}$ ±½(J(J+1)−KK')^{½}(K+K')
<v7|J$_b$J$_c$|v12>	−1.41515(360)×10^{-4}	2.456(33)×10^{-5}	δ$_{K±2K'}$ ±½(J(J+1)−½K(K+K'))^{½} (J(J+1)−½K'(K+K'))^{½}
<v7|J$_a$J^2|v12>	−2.59640(1046)×10^{-5}	0	δ$_{KK'}$KJ(J+1)

[c] Constrained to ground state value

Conclusions

This paper presents the analysis of 8 vibrational bands involving 18 vibrational states of two isotopologues of *cis*- and *trans*- dichloroethane, including all the allowed fundamental bands below 1000 cm^{-1} for both species and several combination and hot bands. In each case,

automated search for assignments applied to K'_a = 0 or 1 with a range of J' values around 15 leads to good assignments on the first search tried in favourable cases, and after at most a few searches in all cases. The nearest lines plot introduced in this work gives a very clear diagnostic as to the quality of assignments and allows rapid extension of assignments to related transitions. In comparison to the preliminary version of the paper, the presence of identifiable structure in the spectrum is not required, as the nearest lines plot picks out the required lines. As the examples above demonstrate the process is robust with respect to perturbations and the presence of other species, and indeed key to the overall method is only working on a subset of the lines for the band of interest at any one step. This is in contrast to genetic algorithms[10] which must look at the entire spectrum and there is no convenient way of excluding lines from other, unknown species.

Areas still to be explored in future developments of the program include making more formal use of the intensity information and exploring enhancements to the search algorithm to reduce the amount of trial and error necessary. The method is general purpose and should be applicable to any spectrum showing resolved structure, including electronic transitions, and trying it out on a wider variety of systems would also help to develop the algorithm.

Supplementary Files

The complete program described here, including source, executables and documentation for using the PGOPHER program is available from the program website [22] and has been deposited in the University of Bristol Research Data Repository [19]. A separate deposit[20] includes the raw spectra, input files and fit output for all the analyses described in this paper.

Conflicts of Interest

There are no conflicts to declare.

Acknowledgements

References

1.   G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman and B. H. Pate, *Review of Scientific Instruments*, 2008, **79**, 053103.
2.   J. L. Hall, *Reviews of Modern Physics*, 2006, **78**, 1279-1295.
3.   C. M. Western, *J. Quant. Spectrosc. Radiat. Transf.*, 2016, **186**, 221-242.
4.   N. A. Seifert, I. A. Finneran, C. Perez, D. P. Zaleski, J. L. Neill, A. L. Steber, R. D. Suenram, A. Lesarri, S. T. Shipman and B. H. Pate, *J. Mol. Spectrosc.*, 2015, **312**, 13-21.
5.   C. M. Western and B. E. Billinghurst, *Phys. Chem. Chem. Phys.*, 2017, **19**, 10222-10226.
6.   C. M. Western, L. Carter-Blatchford, P. Crozet, A. J. Ross, J. Morville and D. W. Tokaryk, *J. Quant. Spectrosc. Radiat. Transf.*, 2018, **219**, 127-141.
7.   J. Tellinghuisen, *J. Mol. Spectrosc.*, 2003, **221**, 244-249.

8. M. A. Martin-Drumel, C. P. Endres, O. Zingsheim, T. Salomon, J. van Wijngaarden, O. Pirali, S. Gruet, F. Lewen, S. Schlemmer, M. C. McCarthy and S. Thorwirth, *J. Mol. Spectrosc.*, 2015, **315**, 72-79.

9. M.-A. Martin-Drumel, M. C. McCarthy, D. Patterson, B. A. McGuire and K. N. Crabtree, *The Journal of Chemical Physics*, 2016, **144**, 124202.

10. W. Leo Meerts and M. Schmitt, *International Reviews in Physical Chemistry*, 2006, **25**, 353-406.

11. D. P. Zaleski and K. Prozument, *Journal of Chemical Physics*, 2018, **149**, 104106.

12. F. W. Loomis and R. W. Wood, *Physical Review*, 1928, **32**, 223-236.

13. B. P. Winnewisser, J. Reinstädtler, K. M. T. Yamada and J. Behrend, *J. Mol. Spectrosc.*, 1989, **136**, 12-16.

14. M. Winnewisser, B. P. Winnewisser, F. C. De Lucia, D. W. Tokaryk, S. C. Ross and B. E. Billinghurst, *Phys. Chem. Chem. Phys.*, 2014, **16**, 17373-17407.

15. N. C. Craig, R. A. Appleman, H. E. Barnes, E. Morales, J. A. Smith, S. Klee, M. Lock and G. C. Mellau, *Journal of Physical Chemistry A*, 1998, **102**, 6745.

16. R. W. Larsen, F. Hegelund, A. Engdahl, P. Uvdal and B. Nelander, *J. Mol. Spectrosc.*, 2007, **243**, 99-102.

17. L. A. Leal, J. L. Alonso and A. G. Lesarri, *J. Mol. Spectrosc.*, 1994, **165**, 368-376.

18. H. Takeo, M. Sugie and C. Matsumura, *J. Mol. Struct.*, 1988, **190**, 205-214.

19. C. M. Western, *PGOPHER version 10.1*, University of Bristol Research Data Repository, 10.5523/bris.3mqfb4glgkr8a2rev7f73t300c, 2018.

20. C. M. Western and B. E. Billinghurst, *High resolution spectra of cis- and trans- 1,2-dichloroethene*, University of Bristol Research Data Repository, doi.org/10.5523/bris.16lvnq33mt9ea24wclci640zhu, 2018.

21. J. K. G. Watson, *Journal of Chemical Physics*, 1967, **46**, 1935-1949.

22. C. M. Western, http://pgopher.chm.bris.ac.uk PGOPHER, a Program for Simulating Rotational, Vibrational and Electronic Spectra, http://pgopher.chm.bris.ac.uk.

23. K. S. Pitzer and J. L. Hollenberg, *J. Am. Chem. Soc.*, 1954, **76**, 1493.

24. J. K. G. Watson, *J. Mol. Spectrosc.*, 1977, **66**, 500-502.