Multi-Resolution error analysis of predicted absorption coefficients. Method and application to the infrared spectrum of methane at high temperature.
Christian Wenger, Jean-Paul Champion

To cite this version:

HAL Id: hal-00473611
https://hal.archives-ouvertes.fr/hal-00473611v2
Submitted on 10 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Multi-Resolution error analysis of predicted absorption coefficients.
Method and application to the infrared spectrum of methane at high temperature.

Ch. Wenger & J. P. Champion

Laboratoire Intercalé Carnot de Bourgogne (ICB), UMR 5209 CNRS - Université de Bourgogne, 9 Av. A. Savary, BP 47870, F-21078 DIJON, France.

Abstract

A general method for the estimation of the confidence interval of molecular absorption coefficients is presented. Statistical numerical experiments are implemented to quantify the propagation of errors from line parameters to absorption coefficients as a function of the resolution. The method uses line parameter predictions (position and intensity) with estimated uncertainties derived from global polyad models. This work is especially intended to provide expert information for applications requiring theoretical predictions for which the present state of the art of line by line high-resolution modeling is not accurate enough to meet the desired precision. As an illustration, it is applied to predict the absorption coefficient of methane at high temperature involving simultaneously well known lower energy levels (cold bands) and higher energy levels with much lower precisions (hot bands). Potential extensions are discussed in the context of atmospheric remote sensing of astrophysical objects.

Keywords: Error analysis; absorption coefficient; methane; infrared; high temperature; high-resolution; vibration-rotation spectroscopy
Introduction

Remote sensing of the atmosphere of astrophysical objects relies essentially on molecular spectroscopy. Astrophysical investigations generally need both complete and accurate spectroscopic databases. Despite continuous efforts in experimental and theoretical spectroscopic investigations, the lack of data in specific spectral regions of interest is one of the principal limitation of the presently available spectroscopic databases. Extrapolations to relevant experimental conditions like high temperature is also a major issue for astrophysical applications. Among other molecules, methane is present in the atmospheres of many astrophysical objects. The modeling of its absorption coefficient in the near infrared and/or at high temperature remains a challenge mainly due to the intrinsic complexity of its ro-vibrational spectrum. Fortunately, for this type of molecule, global effective models have been introduced very early (Champion et al [1]). They were successfully applied to provide a unified description of the successive polyads of the molecule (Boudon et al [2]). Obviously, the extremely large density of levels in highly excited polyads makes it very difficult (if not to say impossible) to envisage line by line modeling above say 10 000 cm$^{-1}$ (Wenger et al [3]). In fact, the present state of the art of high-resolution modeling is limited approximately to the spectral region below around 5 000 cm$^{-1}$ (Albert et al [4]). The present work is aimed to bring tools for astrophysical applications for which completeness and reliability of spectroscopic data are at least as important as the resolution itself. The key feature of this work is an advanced statistical error analysis of multi-resolution global simulations using an effective Hamiltonian approach.

1. General characteristics of absorption spectra

Absorption - Transmission. In order to introduce the notation used in the present work, let us first consider a gas with a single absorbing transition at the wavenumber $\sigma_0$ (in cm$^{-1}$).

The intensity $I(\sigma)$ transmitted through a homogeneous layer of absorbing gas of thickness $l$ undergoing an incident radiation of intensity $I_0(\sigma)$ is governed by the Beer-Lambert law

$$I(\sigma) = I_0(\sigma) \exp[-l\alpha(\sigma)],$$

(1)

where $\alpha(\sigma)$ is the absorption coefficient (expressed in cm$^{-1}$ if $l$ is expressed in cm).

The absorption $A(\sigma)$ and transmission $F(\sigma)$ are defined by

$$A(\sigma) = 1 - F(\sigma) = \frac{I_0(\sigma) - I(\sigma)}{I_0(\sigma)} = 1 - \exp[-l\alpha(\sigma)].$$

(2)

$A(\sigma)$ represents the absorption of the line recorded using an ideal spectrometer with infinite resolution. The absorption $A_{obs}(\sigma)$ effectively observed using a real instrument is derived from the convolution of $A(\sigma)$ with the apparatus function $f_{app}$ according to

$$A_{obs}(\sigma) = \int_0^\infty A(\sigma') f_{app}(\sigma - \sigma') d\sigma'.$$

(3)
with the normalization condition
\[ \int_{-\infty}^{\infty} \dot{f}_{\text{app}}(t) \, dt = 1. \]  \hspace{1cm} (4)

Normalized profile. The normalized profile \( \Phi(\sigma) \) of the spectral line is defined by
\[ \alpha(\sigma) = k_0 \Phi(\sigma), \]  \hspace{1cm} (5)

where the integrated absorption coefficient \( k_0 \) (in \( \text{cm}^{-2} \)) of the line at the wavenumber \( \sigma_0 \) is defined by
\[ k_0 = \int_0^\infty \alpha(\sigma') \, d\sigma', \]  \hspace{1cm} (6)

with
\[ \int_{-\infty}^{\infty} \Phi(\sigma') \, d\sigma' = 1. \]  \hspace{1cm} (7)

Various line profiles have been proposed by different authors. The Voigt profile which is commonly used in a wide range of experimental conditions will be considered here. The Voigt profile is the convolution of the Doppler and Lorentzian profiles. It is characterized by the Doppler and Lorentzian half-widths at half maximum (HWHM), respectively, expressed in \( \text{cm}^{-1} \) as
\[ \gamma_D = \left( \frac{2\ln2 k_B T}{mc^2} \right)^{1/2} \sigma = 3.58 \times 10^{-7} \sqrt{T/M} \sigma, \]  \hspace{1cm} (8)
\[ \gamma_L = \gamma^0 P \xi, \]  \hspace{1cm} (9)

where \( \gamma^0 \) is the pressure broadening coefficient of the gas (expressed in \( \text{cm}^{-1} \cdot \text{atm}^{-1} \)). In this expression \( \xi \) is the volumetric mixing ratio of the absorbing gas. The quantity \( P \xi \) is the partial pressure of the absorbing gas. In principle \( \gamma^0 \) depends on the transition quantum numbers.

Finally, the integrated absorption coefficient \( k_0 \) of an absorption line is related to its intensity \( S_0 \) by
\[ k_0 = S_0 P \xi, \]  \hspace{1cm} (10)

where \( S_0 \) is expressed in \( \text{cm}^{-2} \cdot \text{atm}^{-1} \).

Obviously, real spectra contain many lines. For simplicity, in what follows, \( \gamma_D \) and \( \gamma_L \) will be assumed as constants for given values of the molar mass \( M \), the temperature \( T \) and the total pressure \( P \) of the gas. Similarly, the same Voigt profile (determined from the experimental conditions) will be assumed for all lines. In particular, the dependencies of the profile parameters on the transition quantum numbers will be neglected. Furthermore no line mixing effect will be considered. Under such assumptions, each line is completely determined by two parameters: the position \( \sigma_i \) and the intensity \( S_i \).
Absorption coefficient. The absorption coefficient (in cm$^{-1}$) of the gas at given temperature and pressure is then given by the summation over all the lines

$$\alpha_{T,P,\xi}(\sigma) = P\xi \sum_i S_i \Phi_{\gamma_D,\gamma_L}(\sigma - \sigma_i),$$  \hspace{1em} (11)

where $\Phi_{\gamma_D,\gamma_L}(\sigma - \sigma_i)$ is the Voigt function centered at $\sigma = \sigma_i$ with Doppler and Lorentzian half-widths given as functions of $T$ and $P$ by Eqs. 8 and 9.

Other notations are found in the literature. For instance Nemtchinov et al [5] expresses the absorption coefficient as a function $k_\nu(p,T)$ of the variables $p$ and $T$ depending on the wavenumber $\nu$. In the present work, focused on modeling under variable resolutions (i.e. involving convolution integrals), it is more convenient to consider the wavenumber $\sigma$ as a variable. However, the two notations refer exactly to the same physical quantity:

$$\alpha_{T,P,\xi}(\sigma) \equiv k_\nu(p,T).$$  \hspace{1em} (12)

Absorption coefficient at resolution $\Delta$. The purpose of the next sections is to analyze the propagation of uncertainties from the line positions and intensities into the predicted absorption coefficient as a function of the resolution. It is therefore necessary to predict the absorption coefficient at various resolutions. In the present work we define the absorption coefficient at resolution $\Delta$ as the convolution product of $\alpha_{T,P,\xi}(\sigma)$ with a normalized Gaussian function of half-width $\Delta$. More precisely the absorption coefficient at resolution $\Delta$ is defined by

$$\alpha_{\Delta,T,P,\xi}(\sigma) = P\xi \sum_i S_i \Phi_{\gamma_G,\gamma_L}(\sigma - \sigma_i),$$  \hspace{1em} (13)

where the effective Gaussian half-width is given by

$$\gamma_G = \sqrt{\gamma_D^2 + \Delta^2}.\hspace{1em} (14)$$

It should be emphasized that $\Delta$ is nothing but an adjustable integration parameter to be distinguished from the experimental resolution $\Delta_{exp}$ considered in the next paragraph. The influence of these leading parameters onto error propagation is illustrated below. These parameters are useful to describe the effect of saturation on error propagation.

Absorption - Transmission spectra. At given temperature, pressure, mixing ratio and path length $l$, the predicted absorption is expressed as

$$A_{obs}(\sigma) = \int_0^\infty \left(1 - \exp \left[-lP\xi \sum_i S_i \Phi_{\gamma_D,\gamma_L}(\sigma' - \sigma_i)\right]\right) f_{\Delta_{exp}}^{app}(\sigma - \sigma') \, d\sigma'.\hspace{1em} (15)$$

Similarly the predicted transmission is expressed as

$$T_{obs}(\sigma) = \int_0^\infty \exp \left[-lP\xi \sum_i S_i \Phi_{\gamma_D,\gamma_L}(\sigma' - \sigma_i)\right] f_{\Delta_{exp}}^{app}(\sigma - \sigma') \, d\sigma'.\hspace{1em} (16)$$

The apparatus function is determined by the experimental setup. The resulting broadening of the predicted signal peaks depends on its half-width $\Delta_{exp}$ which characterizes the experi-
mental resolution. In Eqs. (15) and (16), the integration adjustable parameter $\Delta$ introduced in (13) is set to zero. Note that, even if the apparatus function is Gaussian, the absorption or transmission spectra cannot be derived using the effective Gaussian half-width (14). This would apply only at optical densities sufficiently low to justify a linear approximation for the exponential function involved in Eqs. (15) and (16).

2. Statistical simulations

The purpose of the present study is to quantify the propagation of line parameter uncertainties into absorption coefficient uncertainty as a function of the resolution. It implies a non-trivial conversion of a discrete set of two-dimensional error bars into a one-dimensional uncertainty on a continuous function as formally expressed by

$$\left(\delta\sigma_i, \delta S_i\right) \rightarrow \delta \alpha(\sigma).$$

(17)

It is clear that the absorption coefficient $\alpha(\sigma)$ is not a linear function of the line parameters $\sigma_i$ and $S_i$ so that simple statistical relationships cannot be applied. Therefore, our approach relies essentially on statistical numerical experiments. The main features are described below.

2.1. Statistical samples

We denote by $\tilde{\sigma}_i$ and $\tilde{S}_i$ Normal random variables with expectation values, respectively, given by the predicted positions $\sigma_i$ and intensities $S_i$, and standard deviations denoted by $\delta\sigma_i = \zeta_i$ and $\delta S_i = \eta_i$. These quantities are derived from the statistics of the fits of the Hamiltonian model (STDS prediction in our case (Wenger et al [6])). From Eq. (13) one gets a random variable associated with the absorption coefficient

$$\tilde{\alpha}_{\Delta,T,P,\xi}(\sigma) = P\xi \sum_i \tilde{S}_i \Phi_{\gamma\xi;\gamma L}(\sigma - \tilde{\sigma}_i).$$

(18)

The original function $\alpha_{\Delta,T,P,\xi}(\sigma)$ is usually described by a discrete series of $N$ spectral elements at wavenumber $\sigma = \sigma_j$, $J = 1, ..., N$. $N$ is the number of points in the considered spectral window. Accordingly, the above formula defines $N$ random variables $\tilde{\alpha}_{\Delta,T,P,\xi}(\sigma_j)$. Using Pseudo Normal numbers (Pang [7]), we calculate a statistical sample of size $n = 10000$ for each of the $\tilde{\sigma}_i$ and $\tilde{S}_i$ Normal random variables in such a way to simulate independent variables. By injecting these simulated values into Eq. (13) one gets a statistical sample of the same size ($n = 10000$) for the absorption coefficient. And finally, each of the $N$ spectral elements involved in the considered window is described by a statistical sample of size $n = 10000$. Obviously, the corresponding probability laws have nothing to do with a Normal law. Their statistical properties may be derived from the numerical analysis of the samples. Therefore, a rigorous analysis and treatment of the corresponding histograms were achieved using appropriate statistical functions.

As usual, uncertainties on intensities are given as percentages. In the case of large uncertainties (say 50 % or more), it turned out that simulated statistical samples contained non-realistic negative values for the intensities and for the absorption coefficient itself. Several tests have
2.2. Statistical functions

The expectation value for the absorption coefficient \( < \alpha_J > \) was simply estimated from the statistical mean over the \( n = 10000 \) simulated values. This quantity can be derived theoretically from the expression (18)

\[
< \alpha_J > \simeq \hat{\alpha}_{\Delta,T,P,\xi} (\sigma_J) = P \xi \sum_i S_i \Phi_{G_i,\gamma_{\text{eff}}} (\sigma_J - \sigma_i),
\]

where the effective Gaussian half-widths include the effect of the uncertainties on positions through the convolution of three factors

\[
\gamma_{\text{eff}}^2 = \gamma_D^2 + \Delta^2 + G_i^2.
\]

In this expression, \( G_i = \sqrt{2 \ln 2} \zeta_i \) is the half-width of the Gaussian function representing the Normal law assumed for positions with standard deviations \( \zeta_i \).

Numerical simulations with various sample sizes have been made for checking purposes. Fig. 1 shows the results obtained with a sample size \( n = 1000 \) in the case of a single line centered at \( 3000 \text{ cm}^{-1} \) with \( T = 296 \text{ K} \) and \( P = 1 \text{ Torr} \). The uncertainty on the position is \( \zeta_i = 0.050 \text{ cm}^{-1} \). The resolution is \( \Delta = 0.010 \text{ cm}^{-1} \). The theoretical curve in black is the effective Gaussian according to equations (19) and (20). The curve in red represents the
empirical values $\langle \alpha_J \rangle$. The curve in green represents the empirical values $\langle \alpha_J \rangle$ when an uncertainty of $\eta_i = 50\%$ on the intensity is included. It can be seen that, under the considered conditions, the fluctuations are principally due to the position uncertainty whereas the uncertainty on intensity has a smoothing effect. The fluctuations of the simulated curves decrease as the size of the sample increases and the discrepancies between the theoretical curve and the empirical ones decrease. The size $n = 10000$ was chosen as a good compromise between accuracy and computer time consumption and in all subsequent studies $\langle \alpha_J \rangle$ was evaluated empirically from the numerical statistical mean to cover all situations of position and intensity uncertainties.

Similarly, no analytical formulation could be found to approximate the other parameters of the probability law. In fact, the histograms constructed on the simulated samples exhibit a large variability. By applying standard statistical functions to each sample, the relative variance of $\hat{\alpha}_J$ was estimated by

$$\hat{\text{Var}}(\hat{\alpha}_J) = \frac{1}{n-1} \sum_k \left( \frac{\alpha_k - \langle \alpha_J \rangle}{\langle \alpha_J \rangle} \right)^2,$$

and the corresponding relative standard deviation in percent was derived by

$$\hat{\delta}_J = \sqrt{\hat{\text{Var}}(\alpha_J)} \times 100.$$

(22)

In fact such estimated parameters were not found very convenient for practical exploitations. As a matter of fact, except in very restrictive situations of minor interest, the usual properties of Normal variables could not be exploited. Instead, the construction of confidence intervals at various levels (usually at 90\%) denoted by $[\alpha_{\text{inf}}, \alpha_{\text{sup}}]$ was found more useful for practical interpretations. This was achieved using a numerical algorithm based on the defining conditions (in the case of a 90\% confidence level).

$$\int_0^{\alpha_{\text{inf}}} f(\alpha) d\alpha = 0.05 \quad \text{and} \quad \int_{\alpha_{\text{sup}}}^{\infty} f(\alpha) d\alpha = 0.05.$$

(23)

For illustration purposes, this method was applied to a typical region of the high temperature spectrum of methane. The results are described and discussed below.

3. Application to high temperature methane

3.1. Description of the predicted line parameters

The spectral window from 3057 to 3059 cm$^{-1}$, around the $R3$ multiplet of the $\nu_3$ band of methane was chosen to illustrate the method. The vibrational and rotational temperatures were set, respectively, to $T_{\text{Vib}} = 2000$ K and $T_{\text{Rot}} = 296$ K. In such conditions the contribution to the absorption from the hot bands is important and conversely the rotational excitation is limited to relatively low values of $J$ with a maximum value for the intensities around $J = 10$. The mixing ratio was set to $\xi = 1$. The pressure broadening coefficient was set to $\gamma^0 = 0.08$ cm$^{-1}$-atm$^{-1}$ (HWHM). Three sets of line parameters were calculated using a global effective Hamiltonian implemented in the STDS computer program (Wenger et al. [6]). The first set
Table 1: STDS statistics around the R3 $\nu_3$ multiplet of methane

<table>
<thead>
<tr>
<th></th>
<th>Pentad−GS</th>
<th>Octad−Dyad</th>
<th>Tetradecad−Pentad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of lines</td>
<td>3</td>
<td>27</td>
<td>126</td>
</tr>
<tr>
<td>Effective Range / cm$^{-1}$</td>
<td>0.73</td>
<td>1.94</td>
<td>1.98</td>
</tr>
<tr>
<td>Lines per cm$^{-1}$</td>
<td>4</td>
<td>14</td>
<td>64</td>
</tr>
<tr>
<td>Strongest Intensity</td>
<td>0.3</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean Intensity</td>
<td>0.22</td>
<td>0.011</td>
<td>0.003</td>
</tr>
<tr>
<td>Integrated Intensity</td>
<td>0.65</td>
<td>0.30</td>
<td>0.43</td>
</tr>
</tbody>
</table>

The statistics correspond to the window plotted in Fig. 2 with $T_{\text{vib}} = 2000$ K and $T_{\text{rot}} = 296$ K. Intensities are expressed in cm$^{-2}$ atm$^{-1}$.

includes transitions arising from the ground vibrational state, the other two include the hot band systems arising from the first lower two polyads: Dyad and Pentad. Therefore, these calculations involve the ground state and the lower four vibrational polyads of the molecule, namely the Dyad, Pentad, Octad and Tetradecad (polyad number $n = 1, 2, 3$ and 4, respectively) as well as the $\Delta n = 2$ dipole moment. Thanks to our unified global model, all the ro-vibrational levels were calculated from a common set of effective Hamiltonian parameters but using different orders of approximation for the different polyads. Similarly all the $\Delta n = 2$ transition intensities were calculated from a common set of dipole moment operators implying different orders of approximation for the successive hot band systems. Table 1 summarizes the statistics of the lines involved in this window arising from the three considered band systems. Note that presently available spectroscopic databases (Jacquinet-Husson et al [8], Rothman et al [9]) contain no information about the Tetradecad−Pentad band system. Typical key properties can be drawn from this example that are fully representative of general features of high temperature spectra. The intensity of the strongest lines as well as the mean intensity of the lines is a decreasing function of the excitation of the initial polyad of the transitions. Conversely, the effective spectral range is an increasing function of the initial polyad of the transitions. Finally, under the considered conditions, the integrated intensities of the three band systems are comparable as a result of the increasing line densities with the vibrational excitation of the initial energy levels. In parallel, the modeling accuracy is a decreasing function of the vibrational excitation of the initial energy levels. The estimated individual standard deviations for the three line parameter sets are quoted in Table 2.

For clarity, in our numerical experiments, only the 10 lines with intensities larger than 0.025 cm$^{-2}$ atm$^{-1}$ were considered. The individual position and intensity standard deviations were set as quoted in Table 2 and illustrated in Fig. 2. For simplicity each band system was assigned uniform estimated errors. Vertical arrows indicate lines that are representative of three typical situations discussed below.
Table 2: Estimated errors of the STDS positions and intensities

<table>
<thead>
<tr>
<th>Band system</th>
<th>Err. Positions / cm$^{-1}$</th>
<th>Err. Intensities / %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentad – G.S.</td>
<td>0.003</td>
<td>3 %</td>
</tr>
<tr>
<td>Octad – Dyad</td>
<td>0.01</td>
<td>10 %</td>
</tr>
<tr>
<td>Tetradecad – Pentad</td>
<td>0.2</td>
<td>50 %</td>
</tr>
</tbody>
</table>

The quoted errors correspond to one standard deviation.

Figure 2: Stick spectrum plot around the R3 $\nu_3$ multiplet of methane

The error bars reflect the values quoted in Table 2.
3.2. Results and discussion

The results obtained for the above spectral window of hot methane at typical resolutions are displayed hereafter. Figure 3 shows an overview of the 90% confidence band under high resolution ($\Delta = 0.003$ cm$^{-1}$). In this case the sampling of the signal is high enough so that the corresponding error bars appear as a continuous purple band.

Figures 4 and 5 show the signals and the 90% confidence bands at medium ($\Delta = 0.020$ cm$^{-1}$) and low ($\Delta = 0.100$ cm$^{-1}$) resolutions, respectively. Zoomed windows around the typical dominant lines marked by arrows in Fig. 2 are plotted in Figs. 6, 7 and 8. In regions where the absorption coefficient is dominated by strong lines predicted with high accuracy as in Fig. 6, the expectation value is fairly well centered with respect to the confidence interval at all resolutions. To some extent, in this case, an analytical calculation based on the usual Normal assumptions would yield similar results. At the opposite, in regions where the absorption coefficient is dominated by lines predicted with poor accuracy as in Fig. 8, the results are drastically dependent on the resolution. As seen on Fig. 8, the high resolution confidence band (in purple) is practically uniform over the whole window whereas the corresponding expectation value presents a peak extending clearly outside this band. It means that, in this case, the probability of actually observing the signal as predicted is very small (smaller than 10%). However, at lower resolutions the plot shows that the predicted signal becomes statistically reliable. In the intermediate case illustrated in Fig. 7 where the absorption coefficient is dominated by lines predicted with medium accuracy, the predicted signal remains inside the confidence bands at all resolutions. A more detailed interpretation of such typical behaviors is presented in section 3.4 for the purpose of astrophysical applications.
Figure 4: Confidence band for prediction at 0.020 cm$^{-1}$ resolution

The solid line represents the expectation value of the signal at the resolution of 0.020 cm$^{-1}$. The error bars represent the corresponding 90% confidence intervals.

Figure 5: Confidence band for prediction at 0.100 cm$^{-1}$ resolution

The solid line represents the predicted signal at the resolution of 0.100 cm$^{-1}$. The error bars represent the corresponding 90% confidence intervals.
Figure 6: Confidence bands around the dominant $R(3)F_2$ line from the $\nu_3$ fundamental band

The solid lines represent the predicted absorption coefficient at the three typical resolutions (0.003, 0.020 and 0.100 cm$^{-1}$ respectively). The error bars represent the corresponding 90% confidence intervals. The colours distinguishing the resolutions are those introduced in Figs 3, 4 and 5.

Figure 7: Confidence bands around the dominant $R(4)F_1$ line from the $\nu_3 + \nu_4 - \nu_4$ hot band

The solid lines represent the predicted absorption coefficient at the three typical resolutions (0.003, 0.020 and 0.100 cm$^{-1}$ respectively). The error bars represent the corresponding 90% confidence intervals. The colours distinguishing the resolutions are those introduced in Figs 3, 4 and 5.
The solid lines represent the predicted absorption coefficient at the three typical resolutions (0.003, 0.020 and 0.100 cm$^{-1}$ respectively). The error bars represent the corresponding 90% confidence intervals. The colours distinguishing the resolutions are those introduced in Figs 3, 4 and 5.

### 3.3. Details on statistics

In fact, to some extent, the construction of confidence intervals is equivalent to arbitrarily replacing a complex probability law by some trivial uniform law. For a proper understanding of the present analysis, it is of interest to investigate the actual probability laws involved. For reasons mentioned previously, such investigations cannot be performed in a purely analytical way. In this section we describe the main features of the histograms generated by numerical simulations.

Figure 9 shows a medium resolution ($\Delta = 0.010$ cm$^{-1}$) picture around the $R(4)F_1$ line from the $\nu_3 + \nu_4 - \nu_4$ hot band. The histograms associated to three representative spectral elements (marked by arrows) are plotted in Fig. 10.

The expectation value of the absorption coefficient (in magenta in both figures) was calculated in two ways: (i) from the direct average over the $n = 10,000$ samples and (ii) from the analytical expression (19). The two values are very close so that they cannot be distinguished on the plots. Under the assumptions of previous sections, the corresponding line shape for a single line is represented by a Voigt function with an effective Gaussian HWHM reflecting the convolution of three effects: Doppler broadening, position uncertainty and resolution as expressed in Eq. (20). Uncertainties on intensities have no sensible effect in this case. The position of the expectation value within the confidence intervals varies from one spectral elements to the next as the overall shape of the histograms themselves. The same behavior also applies to the maximum probability represented by plain circles in wine in Figs. 9 and 10. For completeness the actual upper and lower values of the simulated samples are also plotted.
The solid line in magenta represents the expectation value of the absorption coefficient at the resolution of $\Delta = 0.010 \text{ cm}^{-1}$. The error bars in purple represent the 90% confidence intervals determined for each spectral element. Plain circles in wine indicate the value of $\alpha(\sigma)$ having the largest probability. Plain triangles in grey indicate the overall interval of each simulated sample. The arrows point to the spectral elements whose histograms are plotted in Fig. 10.

The histograms correspond, respectively, from left to right, to the spectral elements at the wing, side and center of the window shown in Fig. 9. The color conventions are the same. The vertical bar in purple represents the 90% confidence interval. The plain circle in wine points to the maximum probability. The tick in magenta indicates the expectation value (mean).
The black line represents the predicted absorption coefficient at the resolution of 0.100 cm\(^{-1}\). The coloured lines represent the confidence bands at the levels of 90, 80, 70, 60 and 50% at the same resolution.

As mentioned previously, the non-realistic negative values were kept without any specific treatment.

To retrieve the information hidden by the construction of confidence intervals, it is helpful to consider different confidence levels, which implies some kind of discretization of the probability law. Figs. 11 and 12 illustrate the non-linear behavior at typical low-resolutions. Note in particular (Fig. 12) that in the regions dominated by weak lines with large uncertainties (like above 3058.3 cm\(^{-1}\)) the lower bounds of all confidence bands coincide with the zero baseline. Simple rules for the interpretation and the exploitation of such statistical information are presented and discussed in the next section.

3.4. Interpretation for astrophysical applications

Astrophysical applications include the detection of molecular species and the retrieval of concentrations as well as environmental physical conditions. The reliability of theoretical predictions for such purposes can be quantified by considering the position of the predicted absorption coefficient relative to the confidence band, which is drastically dependent on the resolution as shown in Figs. 3, 4 and 5. Three typical situations can be distinguished corresponding to the lines marked by arrows in Fig. 2. They are illustrated in Fig. 13 where the resolution is \(\Delta = 0.001\) cm\(^{-1}\).

**Case 1.** The predicted signal lies inside the confidence band and the confidence band lower limit is strictly positive. In this case, the predicted signal is reliable within the precision indicated by the confidence band. It happens in spectral windows dominated by transitions predicted...
Figure 12: Confidence bands at various levels at 0.050 cm$^{-1}$ resolution

The black line represents the predicted absorption coefficient at the resolution of 0.050 cm$^{-1}$. The coloured lines represent the confidence bands at the levels of 90, 80, 70, 60 and 50% at the same resolution.

Figure 13: Warnings at 0.001 cm$^{-1}$ resolution

The black line represents the predicted signal at the resolution of $\Delta = 0.001$ cm$^{-1}$. The red and blue lines represent upper and lower limits of the corresponding 90% confidence intervals. The traffic lights green, yellow and red symbolize the three typical situations described in the text.
with good accuracy. In this case the confidence band is governed mainly by the uncertainties on intensities.

**Case 2.** The predicted signal lies inside the confidence band but the confidence band lower limit and the baseline coincide. It means that the zero value is included in the confidence band. In other words the predicted absorption coefficient is *not significant* at the confidence level of 90%. It is a warning pointing out the uncertainty of the position of the predicted features. It happens in spectral windows dominated by transitions predicted with an uncertainty larger than the considered resolution. In this case the confidence band is governed mainly by the uncertainties on positions and to a much lesser extent on intensities.

**Case 3.** The predicted signal crosses over the confidence band upper limit (while the lower limit coincides with the baseline). As in the preceding case, the predicted absorption coefficient is *not significant* not only at the 90% confidence level but also at higher confidence level (95% ...) since the values are far away from the 90% confidence band. It is a warning pointing out the large uncertainty of the position of the predicted features. It happens in spectral windows dominated by transitions predicted with an uncertainty much larger than the desired resolution. In this case the confidence band is governed mainly by the uncertainties on positions. Uncertainties on intensities are practically negligible despite their relatively large value of 50%.

It is important to understand that, given line parameter accuracies, the situation is drastically dependent on the resolution. To illustrate this dependency, a series of figures displaying the plots for all the resolutions mentioned on the right part of Fig. 13 is available as supplementary materials.

### 3.5. Transmission confidence band

The procedure described in the preceding sections can be applied to transmission spectra. In this case, the resolution $\Delta$ is replaced by the experimental resolution defined by the HWHM $\Delta_{exp}$ of the apparatus function. When the optical density is low enough, the transmission signal depends almost linearly on the absorption coefficient. Conversely, at high optical density, the exponential dependency produces drastic distortions on the confidence band. A typical example is illustrated on Fig. 14 to show the influence of the saturation. In this simulation a Gaussian apparatus function was used with $\Delta_{exp} = 0.050$ cm$^{-1}$. The pressure was set to $P = 1$ Torr and the path length was set to $l = 1$ km. The interpretation previously given for the absorption coefficient is very similar in the present case except that, in addition to all other effects already described, the transmission level drastically influences the transmission uncertainties as reflected by the predicted confidence band. In particular, as expected, the confidence band loses its interest in the case of saturated features. Note that this conclusion is quite consistent with the well known conditions required for the inverse problem of measuring accurate intensities from transmission spectra where saturated features are useless.

In applying our procedure to the transmission signal it was found necessary to take care of non-realistic random values. As a matter of fact, in the case of large intensity uncertainties (such as 50%), statistical samples may include a few negative values. We mentioned previously that
The black line represents the predicted signal at the resolution of $\Delta_{exp} = 0.050$ cm$^{-1}$. The pressure and path length are, respectively, $P = 1$ Torr and $l = 1$ km. The red and blue lines represent upper and lower limits of the corresponding 90% confidence intervals.

The effect of such values was negligible when applied to the derivation of statistical functions for the absorption coefficient. Now, due to the exponential dependency of the transmission signal, such values often cause the divergence of statistical functions. To overcome this problem, we simply replace all randomly simulated negative values by zero.

4. Conclusion and perspectives

The error propagation from predicted line positions and intensities into simulated absorption coefficient and transmission spectra has been investigated using statistical numerical experiments. The non-trivial statistical properties of the predicted signals have been illustrated on the high temperature spectrum of methane. This example was chosen as representative of a common situation in which spectroscopic databases are incomplete and high-resolution theoretical modeling is missing. Our procedure is applicable to any kind of molecules and may also be easily extended to include the propagation of errors from lineshape parameters. The main limitation of the method is the availability of predicted line parameters with known uncertainties.

The present approach takes advantage from the flexibility of the global effective Hamiltonian model adapted for methane and its STDS computer implementation. These tools can be used for low-precision predictions involving a relatively small number of parameters (Borysov et al [10]). Statistical numerical experiments can then be applied to the model parameters instead of the line parameters in order to account for the correlations among the later. Ab initio predictions can also be exploited to predict line parameters and their respective uncertainties.
and possibly their correlations. Work is in progress in this direction. Also a friendly graphical
user interface is being developed to provide multi-resolution information on methane with
quantified uncertainties. The overall goal of the project is to compensate for the incompleteness
of spectroscopic databases by providing exhaustive information at resolutions reflecting up to
date modeling state of the art with realistically quantified uncertainties.

Acknowledgments

Part of the work was performed in the framework of LEFE-CHAT National Program
CNRS (France). Support from Agence Nationale de la Recherche (France) through the project
CH4@Titan (ref: BLAN08-2 321467) and from GDRI SAMIA Tomsk (Russia), Hefei (China),
CNRS (France) is gratefully acknowledged.
References


