



Contents lists available at ScienceDirect

## Journal of Quantitative Spectroscopy &amp; Radiative Transfer

journal homepage: [www.elsevier.com/locate/jqsrt](http://www.elsevier.com/locate/jqsrt)

# Assessment of the precision, bias and numerical correlation of fitted parameters obtained by multi-spectrum fits of the Hartmann-Tran line profile to simulated absorption spectra



Erin M. Adkins\*, Joseph T. Hodges

Chemical Sciences Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

## ARTICLE INFO

## Article history:

Received 3 December 2021

Revised 21 January 2022

Accepted 24 January 2022

Available online 26 January 2022

## Keywords:

Line shapes

Line profiles

Dicke narrowing

Speed dependence

Monte Carlo analysis

Line lists

## ABSTRACT

Although the Voigt profile has long been used to analyze absorption spectra, the quest for increased precision, accuracy and generality drives the application of advanced models of atomic and molecular line shapes. To this end, the Hartmann-Tran profile is now recommended as a standard for high-resolution spectroscopy because it parameterizes relevant higher-order physical effects, is computationally efficient, and reduces to other widely used profiles as limiting cases. This work explores the uncertainty with which line shape parameters can be obtained from constrained multi-spectrum fits of spectra simulated with this standard profile, varying uncertainty levels in the spectrum detuning and absorption axes, and spanning a range of sampling density, pressure, and line shape parameter values. The analysis focuses on how noise-limited measurement precision of frequency detuning and absorption drive statistical uncertainties in fitted parameters and numerical correlations between these quantities. Also, we quantify the degree of equivalence between the full Hartmann-Tran profile and those derived from it in terms of fitted peak areas and line shape parameters. Finally, we introduce a new open-source software package named Multi-spectrum Analysis Tool for Spectroscopy (MATS), which allows users to fit the HTP and its derived profiles to experimental or simulated absorption spectra to explore the limits of the HTP under actual experimental or user-defined conditions.

Published by Elsevier Ltd.

## 1. Introduction

Line-by-line spectroscopic parameters for absorption transitions of molecular and atomic species are widely tabulated in databases such as HITRAN [1], GEISA [2], CDMS [3], JPL Submillimeter, Millimeter, and Microwave Spectral Line Catalog [4], CDS [5], ExoMol [6], AMES [7], TheoReTs [8], NIST SRD 78 [9], NIST SRD 118 [10], and SMPO [11]. These data serve as inputs to line shape models that are used for calculations of absorption/emission spectra and are used in analyses of spectroscopic measurements in the laboratory and field, remote sensing applications and radiative transfer calculations. Among the intrinsic parameters in these line lists, quantities include upper- and lower-state quantum numbers and degeneracy factors, lower-state energy, transition frequency, intensity, and spontaneous emission coefficient.

Semi-empirical models for spectroscopic line shapes describing light absorption or emission by gaseous molecules (or atoms) generally involve parameters for individual transitions that depend

upon temperature, pressure, and composition of the gas mixture. These line profiles are based on more fundamental studies of line shapes which cannot always be parameterized in terms of analytical solutions to governing equations [12]. There are a variety of effects caused by collisional interactions between the absorber and collisional partners that lead to line broadening, shifting, and asymmetry [13]. Of particular relevance in this study are velocity-changing and internal-state-changing collisions which are usually dominant among the high-order line shape effects. Velocity-changing collisions can be analytically described assuming soft-collisions [14], hard-collisions [15] or their combination [16]. Relaxing these assumptions has led to numerical solutions where the effects of velocity changes are parameterized by the frequency of velocity-changing collisions and either the absorber-to-perturber mass ratio [17] in the case of the billiard-ball (BB) model or a velocity memory parameter [18] for the Keilson-Storer (KS) model. The KS approach was found to be in good agreement with Raman spectroscopy measurements in pure H<sub>2</sub> [19] and observations of isolated lines of H<sub>2</sub>O [20]. Numerical-based *ab initio* classical molecular dynamics simulations (CMDS) [21] have been used to compute the speed dependence of internal-state-changing collisions for air- and N<sub>2</sub>-broadened H<sub>2</sub>O [22]. For the case of self-

\* Corresponding author.

E-mail address: [erin.adkins@nist.gov](mailto:erin.adkins@nist.gov) (E.M. Adkins).

and Ar-broadened H<sub>2</sub>, calculations of the frequency of velocity-changing collisions and speed-changing collisions [23] using the BB model were shown to be consistent with CMDS. Combining the KS and CMD models has also enabled numerical calculations of the combined effects of Dicke narrowing and speed-dependent narrowing, as well as the physical correlation between these velocity- and state-changing collisions [22]. Recently, there has also been substantial progress in applying *ab initio* quantum-scattering methods to calculate line shape parameters such as pressure broadening and shifting, speed dependence and Dicke narrowing for the cases of absorption by H<sub>2</sub> [24] and by HD [25] both perturbed by He, illustrating the potential for theory to complement and guide high-precision line shape measurements for more complicated molecular systems.

In practice, adjustable parameters of profiles chosen to represent line shapes are obtained by fits to measured spectra as well as to theoretical calculations [26,27]. These models are fit for purpose and provide a useful alternative to time-consuming calculations associated with BB- or CMDS-based approaches that preclude routine data analysis. The most commonly used profile is the Voigt profile (VP) which follows from convolution of the Lorentz (LP) and Doppler (DP) profiles [28]. The LP describes homogeneous line broadening in terms of the collisional broadening,  $\Gamma_0$ , of the excited state and line shifting term,  $\Delta_0$ , both of which are proportional to pressure. The DP accounts for inhomogeneous Doppler broadening by the rate,  $\Gamma_D$ , equal to the ratio of the thermally averaged absorber velocity and wavelength of the radiation. The DP and LP represent limiting cases of the VP which occur at low and high extremes of gas pressure, respectively. These line broadening mechanisms usually dominate observed spectra leading to traditional selection of the VP for data analysis— especially in cases where the experimental signal-to-noise ratio (SNR) is too low to reveal the effects of weaker contributions to the line shape.

For experiments with sufficiently high SNR, fitting the VP to measured spectra tends to yield characteristic W-shaped residuals. These results can be attributed to the fact that the VP underpredicts the peak height and overestimates absorption in the line wings when line narrowing is present. Such systematic deviations from the VP in observed line shapes arise from the neglect of line narrowing involving velocity-changing collisions and/or speed-dependence in the rate of internal-state-changing collisions [29]. Therefore, to prevent biases in measured line parameters and to reach precision and accuracy targets in remote sensing applications [27,30], the use of advanced line profiles may be required to account for higher-order line shape effects.

A sampling of beyond-the-Voigt profiles includes those that account for Dicke narrowing, speed-dependence of the line width, and correlation effects. Dicke narrowing tends to reduce the line width because it lowers the average speed of the active molecule and the effective Doppler width [12,28,31]. In deriving models of this line-narrowing mechanism, distinct profiles for the soft- or hard-collision cases are usually considered. Both cases involve the parameter  $\nu_{VC}$ , referred to as the frequency of velocity-changing collisions. The soft-collision approximation results in the Galatry profile (GP) [29,32]. This model assumes that the collective effects of many individual velocity-changing collisions are required to be significant and lead to Brownian motion of the absorber, so that  $\nu_{VC}$  can be calculated in terms of molecular diffusion. For the hard-collision case, the VP generalizes to the Nelkin-Ghatak profile (NGP) or equivalently the Rautian profile (RP) and the pre- and post-collision velocities of the absorber are assumed to be independent so all velocities follow a Maxwellian distribution [32]. The speed-dependent Voigt profile (SDVP) uses two additional parameters,  $\Gamma_2$  and  $\Delta_2$ , to quantify the speed-dependence of the collisional width and shift, respectively. These terms are usually assumed to vary quadratically with absorber speed [29] and can be

used to describe speed-dependent line narrowing and broadening. Although the GP, NGP, and SDVP each account for a single narrowing mechanism, both Dicke narrowing and speed-dependence effects can be considered by using the speed-dependent Galatry profile (SDGP) or the speed-dependent Nelkin-Ghatak profile (SDNGP).

While there are numerous line profiles beyond the VP (many of which remain unmentioned here) that account for higher-order line shape effects, there has been strong motivation to adopt a standardized profile with parameters having the appropriate physical attributions. A profile of this type is needed for inclusion in spectroscopic databases and to facilitate comparisons in the literature, as well as to foster consistent analyses among practitioners [29]. To this end, in 2014 an IUPAC task group (Project No. 2011–022–2–100, “Intensities and line shapes in high-resolution spectra of water isotopologues from experiment and theory”) recommended a standardized profile that captures three physical mechanisms which are beyond those embodied in the VP [28]. These effects include Dicke narrowing, speed-dependence of the collisional width and shift, and correlation between velocity-changing and internal-state-changing collisions through the correlation parameter,  $\eta$ . The recommendation by the IUPAC task group was to use the Hartmann-Tran profile (HTP) [28], which generalizes the quadratic SDNGP to include correlation between velocity-changing and internal-state changing collisions, which may be required when  $\Gamma_0$  and  $\nu_{VC}$  are of comparable magnitude [12,26]. In addition to accounting for higher-order effects, the HTP was chosen because its limiting cases are other commonly used profiles such as the VP, NGP, SDVP, and SDNGP discussed above.

The HTP is a function of the frequency detuning about vacuum line center frequency,  $\omega_0$ , given by  $\omega - \omega_0$  and can be written in terms of seven adjustable parameters as  $f_{HTP}(\omega - \omega_0; \Gamma_D, \Gamma_0, \Gamma_2, \Delta_0, \Delta_2, \nu_{VC}, \eta)$ . The additional beyond-the-Voigt terms are the collisional narrowing frequency,  $\nu_{VC}$ , the speed-dependent broadening parameter,  $\Gamma_2$ , the speed-dependent shifting parameter,  $\Delta_2$ , and correlation parameter,  $\eta$  [28]. We note that  $\Gamma_0, \Gamma_2, \Delta_0, \Delta_2$ , and  $\nu_{VC}$  have dimensions of inverse time, are temperature dependent and proportional to pressure, whereas  $\eta$  is dimensionless, ranges from zero to unity, and is usually assumed to be independent of pressure. The pressure normalized parameters ( $\gamma_0, \gamma_2, \delta_0, \delta_2$ , and  $\nu_{VC}$ ) are used for archival purposes and are defined as the lower-case counterparts of the pressure proportional quantities. Additionally, the two speed-dependent parameters are commonly referred to by the dimensionless ratios of the speed-dependent and collisional parameters,  $a_w = \frac{\Gamma_2}{\Gamma_0}$ ,  $a_s = \frac{\Delta_2}{\Delta_0}$ .

The HTP has been shown to reproduce measurements of isolated lines for various molecules to within 0.1%, reducing biases in peak area measurements by more than an order of magnitude [33]. Additionally, despite being based on a total of six parameters beyond the Doppler width, the HTP can be computed in terms of the VP [29], allowing for fast line-by-line (LBL) parameterization critical to implementation in radiative forcing models [28]. Table 1 summarizes the parameters used in each limiting case of the HTP.

The aforementioned IUPAC group recommended that during spectral analysis the HTP be used over a wide range of pressures, with data exhibiting sufficiently high SNR and with a multi-spectrum fitting approach [28,30]. Multi-spectrum fitting involves simultaneously fitting multiple spectra and imposing physically based constraints (*i.e.* pressure, temperature, sample composition) on the line shape. Using this fitting approach over a suitable range of  $\Gamma_0/\Gamma_D$  helps to disentangle parameters that have similar effects (*i.e.* Dicke narrowing and speed-dependent broadening) [28,29]. To this end, Wcislo and coworkers [27,34] also recommend a modification to the HTP, called the  $\beta$ -HTP, which provides a calculable physical constraint on the pressure dependence of the rate of velocity-changing collisions and thereby reduces bias caused by nonlinear correlation between  $\nu_{VC}$  and  $\Gamma_2$  when analyzing spectra

**Table 1**  
HTP and HTP limiting-case line profiles.

Line Profile	Parameters	Limit of HTP
Voigt (VP)	$\Gamma_D, \Gamma_0, \Delta_0$	$\Gamma_2 = \Delta_2 = \nu_{vc} = \eta = 0$
Nelkin-Ghatak (NGP)	$\Gamma_D, \Gamma_0, \Delta_0, \nu_{vc}$	$\Gamma_2 = \Delta_2 = \eta = 0$
Speed-dependent Voigt (SDVP)	$\Gamma_D, \Gamma_0, \Delta_0, \Gamma_2, \Delta_2$ $a_w = \frac{\Gamma_2}{\Gamma_0}, a_s = \frac{\Delta_2}{\Delta_0}$	$\nu_{vc} = \eta = 0$
Speed-dependent Nelkin-Ghatak (SDNGP)	$\Gamma_D, \Gamma_0, \Delta_0, \Gamma_2, \Delta_2, \nu_{vc}$ $a_w = \frac{\Gamma_2}{\Gamma_0}, a_s = \frac{\Delta_2}{\Delta_0}$	$\eta = 0$
Hartmann-Tran (HTP)	$\Gamma_D, \Gamma_0, \Delta_0, \Gamma_2, \Delta_2, \nu_{vc}, \eta$ $a_w = \frac{\Gamma_2}{\Gamma_0}, a_s = \frac{\Delta_2}{\Delta_0}$	

of finite SNR. Fleurbaey et al. [35] demonstrated that use of the  $\beta$ -HTP can increase the fitted  $\nu_{vc}$  by as much as 50% compared to the value obtained using the HTP. As more users adopt the HTP and as advanced line parameters are added to databases such as HITRAN, it is important to understand the limitations that data quality has on the use of this and other profiles [1]. Specific concerns have been raised about numerical correlation between parameters [36], such that simply minimizing residuals does not necessarily result in physically meaningful values or measurement uncertainties [35].

In the remainder of this article, we explore the uncertainty with which parameters of the HTP can be obtained from constrained multi-spectrum fits to synthetic data that are subject to random, uncorrelated noise in both the spectrum detuning and absorption axes. This noise leads to quantifiable statistical uncertainties in the fitted parameters. We assume that the observed spectrum of transmission through a medium,  $\mathfrak{S}(\nu)$ , follows the Beer-Lambert law given by  $\mathfrak{S}(\nu) = e^{-\alpha(\omega)L}$ , where  $\alpha(\omega)$  is the frequency-dependent absorption coefficient and  $L$  is the absorption pathlength. We have also assumed that the optical depth  $OD = \alpha(\omega)L$  of the sample is much less than unity. This case is the optically thin limit where both transmission,  $\mathfrak{S}$ , and absorption,  $(1 - \mathfrak{S} \approx \alpha L)$ , spectra vary linearly with the absorption coefficient so that the respective spectral shapes are independent of pathlength. Although the choice of small OD is a good approximation for the single-pass cavity loss required in cavity ring-down spectroscopy (CRDS) and cavity mode-width spectroscopy (CMWS), many other transmission methods such as Fourier-transform spectroscopy, dual-comb spectroscopy, cavity-enhanced absorption spectroscopy, etc. usually do not operate in this linear domain. Nevertheless, the approach presented here can be generalized to optically thick cases using the full form of the Beer-Lambert law, which introduces a non-linear dependence of the experimentally modeled line profiles on OD. In the supplementary information, we present an example calculation illustrating how the OD can affect the statistics of fitted profile parameters.

We simulated spectra with noise in the spectrum detuning and absorption axes and evaluate the results spanning a range of relevant cases such as sampling density and line shape parameter values. The article proceeds in three sections. First, we quantify how noise-limited measurement precision of frequency and absorption affect uncertainties in fitted parameters and numerical correlations between these quantities when using the full HTP. Next, we analyze the degree of equivalence between the HTP and its limiting-case line profiles in terms of the peak areas as well as line shape parameters. Finally, we introduce a new open-source software package named Multi-spectrum Analysis Tool for Spectroscopy (MATS) [37], which can be used to fit the HTP and its derived profiles to experimental or simulated spectra to explore the limits of the HTP under actual experimental or user-defined conditions.

## 2. Methodology

We developed a Python-based multi-spectrum simulation and fitting program named MATS to provide a generic platform for constrained, multi-spectrum fitting of the HTP to experimental and

synthetic line shape data. MATS [37] allows for the simulation and/or fitting of absorption spectra based on an input spectroscopic line list comprising multiple molecules, isotopologues, collisional partners, and spectroscopic transitions. The present section describes the mechanics of a precursor to MATS, which allowed for the use of the unit area HTP profile with normalized parameter units and lower-computational overhead for single-line simulation and fitting.

The precursor program is broken up into simulation and fitting components, where isolated lines are simulated using the HTP over a range of “pressures” and subsequently analyzed by multi-spectrum fitting where pressure-dependent line profile parameters were constrained to be linearly proportional to pressure. The inputs to the simulation include designated HTP parameters, the spectrum sampling density, and relative noise in the frequency detuning ( $x$ -axis), and absorption ( $y$ -axis) axes. For a given set of input parameters, seven spectra were simulated over the normalized linewidth range  $\Gamma_0/\Gamma_D = 0.1$  to 100. This range was used in validation of the HTP [38] and satisfies the criterion of a broad pressure interval over which the dominant broadening mechanism goes from being nearly Doppler-limited to pressure-broadened limited. In the development of the present program, we used the Python-based HITRAN Application Programming Interface (HAPI) [39] HTP definition based on the mathematical expressions given in Ngo *et al.* [29,38]. The ratio  $\Gamma_0/\Gamma_D$  is used as the pressure proxy in this study, where the pressure-normalized collisional broadening term,  $\gamma_0$ , and the Doppler width are unity. The spectra are centered at zero and each spectrum has a frequency range of  $\pm 10$  halfwidths (calculated using the Voigt profile Lorentzian halfwidth,  $\Gamma_0$ ). The spectra were simulated and modeled at discrete frequencies where the sampling density is defined as the number of points across the Doppler half-width at half-maximum.

To simulate experimental spectra, independent sets of Gaussian white noise with specified standard deviations along the frequency and absorption axes are added to the HTP, providing a consistent magnitude noise level on both  $x$ - and  $y$ -axes. For reference purposes, the noise-free HTP at the simulation condition will be referred to as the “true” spectrum. The standard deviation of the frequency noise ( $x$ -axis) is specified as a fraction of the Doppler width,  $\sigma_x$ , whereas that of the absorption noise ( $y$ -axis) is defined as a fraction of the maximum absorption for each spectrum,  $\sigma_y$ . The difference between the noisy spectrum and the true spectrum is entirely dependent on the  $\sigma_x$  and  $\sigma_y$  magnitudes, such that the combined spectrum noise for the two axes,  $\sigma_{xy}$ , can be calculated as the standard deviation of this difference. The quality-of-fit (QF) figure-of-merit is the ratio of the maximum absorption in a noisy spectrum to the standard deviation of the fit residuals. This quantity is reported with respect to fits of the HTP to the noisy spectrum (with all parameters floated and initial guesses at the simulation condition), as well as for partially constrained fits corresponding to less-general line profiles. The reciprocal of QF at the simulation condition approaches  $\sigma_{xy}$  when the fit converges to the true value. Fits of less-general profiles to the simulated spectrum result in lower QF values that are limited by a combination of systematic effects and noise. Although the QF metric is useful in comparison of fitting various line profiles to the same set of observed spec-

**Table 2**  
Summary of Parameter space covered in Simulations.

Parameter	Simulated Values	Fit Constraints	Line profile Constraints
$\sigma_y$ (%)	0, 0.001, 0.0002, 0.01, 0.02, 0.1, 0.2, 1		
$\sigma_x$ (%)	0, 0.001, 0.01, 0.1, 1, 10		
Spectral Density (pts/ HWHM $\Gamma_D$ )	2, 5, 10, 20		
$\omega_0$	0	0	
Area	1	spectrum	
$\Gamma_0/\Gamma_D$	0.1, 0.5, 1, 5, 10, 50, 100		
$\Gamma_D$	1		
$\gamma_0$	1		
$a_w$	0*, 0.01, 0.05, 0.1, 0.25, 0.5	0 – 1000	0: VP and NGP
$\delta_0$	0.01, 0.05, 0.1, 0.25		
$a_s$	0, 0.01, 0.05, 0.1, 0.25		0: VP and NGP
$\nu_{VC}$	0*, 0.01, 0.05, 0.1, 0.25, 0.5		0: VP and SDVP
$\eta$	0, 0.05, 0.1, 0.25, 0.5	0 – 0.95	0:VP, NGP, SDVP, and SDNGP

\* Spectra with  $\sigma_x = \sigma_y = 0$  and the full range of parameter combinations with  $a_w$  and/or  $\nu_{VC}$  equal to 0 were also included to explore the equivalence between speed-dependent and Dicke narrowing parameters in the NGP, SDVP, and SDNGP line shapes discussed in Supplementary Materials.

tra, because QF is related to the number of spectral points and the spectral window, it is not the ideal metric for spectrum intercomparisons, especially when comparing results from different experiments. Additionally because, we consider only simulated data and use the same functional form for the true spectrum and the spectrum fitting function, our results yield no information about systematic deviations between the HTP and experimental line shapes. To quantify such deviations would require fits of the HTP to spectroscopic measurements and is outside the scope of this article.

We used the LMFIT non-linear least-squares minimization and curve-fitting Python tool for multi-spectrum least squares fitting to the simulated spectra [40]. With LMFIT, the line profile parameters ( $\gamma_0$ ,  $\gamma_2$ ,  $\delta_0$ ,  $\delta_2$ ,  $\nu_{VC}$ ,  $\eta$ ) were constrained to be equal across the different pressure spectra with additional parameter bounds imposed to aid in fit convergence. The peak area was floated independently for each spectrum to allow additional flexibility and to reduce biases in fitted area. Reported peak areas were based on averages across all pressures. Initial guesses for all parameters were set to the simulated values as an idealized case. Additionally, numerical studies revealed that the fitting protocol behaved robustly when initial guesses were within an order-of-magnitude of the true values and parameter convergence was independent of these values. For a given set of simulated spectra, not only the HTP, but also each of the limiting HTP line profiles was fit to the simulated spectra. Each fit originated with a vector of simulated (true) parameter values,  $\psi_s$ , comprising,  $\psi_{s,1}$ ,  $\psi_{s,2}$ ... and resulted in a vector of fitted parameter values,  $\psi_f$ , where the elements  $\psi_{f,1}$ ,  $\psi_{f,2}$ , ... correspond to fitted values for  $\gamma_0$ ,  $a_w$ , ... For each fitted parameter  $\psi_f$ , LMFIT reports a standard error,  $u_{se}(\psi_f)$  with which we define the relative uncertainty,  $u_r(\psi_f) = u_{se}(\psi_f)/\psi_f$ . Also, we replicated simulations and fits of spectra for a given set of conditions with unique noise vectors of equivalent magnitude, yielding a  $\psi_f$  vector comprising mean values of the replicate fitted parameters. We then calculated the magnitude of the relative difference for each fitted parameter and condition as,  $\varepsilon(\psi_f) = |(\psi_f - \psi_s)/\psi_s|$ , which we refer to as the fractional error. The  $u_r$  values were averaged over replicate values in the same way. We explore comparisons of these uncertainty metrics below.

Table 2 summarizes the range of environmental and line profile parameters considered here, along with the fit and constraints applied to each parameter during fitting. We chose the investigated ranges of each line profile parameter based on those used in the Tran *et al.* work that benchmarked the numerical precision of the HTP model [38]. We extended the ranges of the pressure-shift terms to cover a larger interval of  $\Delta_0/\Gamma_0$  and used the same range for  $a_s$ . Finally, we did not float parameters that were simulated to be zero, and we set the maximum value of  $\eta$  to 0.5 to limit

the number of cases. Because the ratio of  $\Gamma_0/\Gamma_D$  is the pressure proxy in this study and  $\gamma_0$  and  $\Gamma_D$  are unity, the  $\Gamma_0$ -normalized line profile parameters are equivalent to the pressure-normalized parameterization. This is unique to the structure of this study. Note that the same symbol is used for the pressure-normalized and pressure-proportional parameterization of  $\nu_{VC}$ . The remainder of this work refers to the pressure-normalized value of this quantity, unless otherwise specified.

To provide some orientation regarding where the properties of an example molecule would fall within this Doppler-width-normalized set of parameters, we compare the parameter ranges to those of a representative  $^{12}\text{C}^{16}\text{O}_2$  line centered at  $6359.967\text{ cm}^{-1}$  (ASCENDS - Active Sensing of  $\text{CO}_2$  Emissions over Nights, Days, and Seasons mission line 30,012←00001 R16e [41]). This line has been measured with various frequency-stabilized cavity ring-down spectroscopy techniques in our laboratory in the last several years [33,42–46], which resulted in a line list generated from fits of the SDNGP to experimental data. For this line, the range of  $\Gamma_0/\Gamma_D$  used in the simulations corresponds to spectra collected over the pressure range 0.8 kPa to 800 kPa, which is certainly a larger interval than would be used in most experiments. The corresponding ratio  $\Delta_0/\Gamma_0$  is -0.07, given by Ref. [46], which is opposite in sign to the shift terms studied here but falls within the magnitude of the range studied. The values of  $a_w$ ,  $a_s$ , and  $\nu_{VC}/\Gamma_0$  for this line were measured to be 0.0884, 0.055, and 0.042, respectively, which all fall within the parameter ranges used in this study. Here, the correlation parameter was not fit, but was calculated to be 0.2 [46]. The range of spectrum sampling densities corresponds to intervals between about 9 MHz and 90 MHz. The noise levels in the frequency axis correspond to root-mean-square deviations ranging from 2 kHz to 20 MHz, in addition to the noise-free case. The noise levels in the absorption axis correspond to SNRs of between  $10^2$  and  $10^5$ . As discussed below, the combined spectrum noise for the two axes,  $\sigma_{xy}$ , is sensitive to noise in both the x- and y-axes, specifically when the absorption noise is relatively low and the frequency axis noise is high.

Although the parameter space used in this study does not apply to all experimental cases, it is intended to cover a wide-enough range to provide general insight into the statistical behavior of parameter retrieval from spectra acquired on samples with small OD and having finite SNR. Note that the MATS software package discussed later in this paper provides the capability for users to consider specific cases involving certain molecular species, overlapping lines, arbitrary experimental conditions, etc.

To sample the full range of parameters considered, the simulation and fit procedure is conducted for each combination of line profile parameter and environmental parameter over the  $\Gamma_0/\Gamma_D$

range. Three replicates per instance were implemented. Results generated on replicated instances were averaged to yield about  $3 \times 10^5$  unique parameter combinations that were successfully analyzed (defined by convergence of the multi-spectrum fit) for each of the five considered line profiles. The relative uncertainty and fractional difference for each parameter were calculated as defined above. The pressure-normalized peak area was calculated as the weighted average over all pressures contributing to the multispectrum fit followed by averaging across replicate samples.

### 3. Results

#### 3.1. HTP parameter retrieval in noise-limited synthetic spectra

In advanced line shape studies of experimental spectra, in the absence of multiple realizations of a given measured parameter, the Type-A evaluation uncertainty in that parameter is often ascribed to the uncertainty reported by the least-squares fitting algorithm. In many fitting programs including the one presented here, such fit uncertainties correspond to the reported standard error in the fitted parameter. This fit metric is calculated from the diagonal elements of the covariance matrix, which are a measure of the parameter variance. Although the off-diagonal elements relate to the numerical correlation between floated parameters in the fit, they are not used in calculation of the standard error. Nevertheless, the fit standard error is an appropriate measure of the uncertainty when fitted parameters are not numerically correlated and when variations are normally distributed. While calculation of confidence intervals through Monte-Carlo analysis or other means can be used for a more robust measure of the statistical uncertainty, these methods can be computationally prohibitive to use in routine fitting procedures.

In Fig. 1 we compare the relative fractional error ( $\varepsilon$ ,  $y$ -axis), between the  $\gamma_{0f}$  (fitted) and  $\gamma_{0s}$  (simulated) values of the collisional broadening term, with the corresponding relative uncertainties ( $u_r(\gamma_{0f})$ ,  $x$ -axis) based on the standard errors from the least-squares fitting algorithm. Results are plotted for a range of simulated conditions and only parsed by  $\sigma_{xy}$  accounting for the observed distribution of values. The  $x$ -axis on this plot represents information that can be obtained when the true value is unknown. In contrast the fractional error values along the  $y$ -axis are not provided by the least-squares fits and require *a priori* knowledge of  $\gamma_{0s}$ . In measurements without Type-B or systematic uncertainty and where the reported Type-A uncertainty is appropriately quantified, the relative uncertainty should approximately bound the relative fractional error. The solid and dashed black lines indicate where there is a 1:1 and 10:1 ratio, respectively, between  $\varepsilon$  and  $u_r(\gamma_{0f})$ . Inspection of the data reveals that the former tends to exceed the latter leading to a median value of 3.7 for the ratio  $\varepsilon/u_r(\gamma_{0f})$ . We ascribe this disparity to numerical correlations between the fitted parameters which are not accounted for using the fit standard error. Replicating this analysis for the other line shape parameters, except for the area, the median ratio of  $\varepsilon/u_r$  was between 4.2 and 5.4. The area median  $\varepsilon/u_r$  ratio was 0.41, however this term was not constrained to the multi-spectrum dependence. Additionally, while the area might be numerically correlated with other parameters, it is the only parameter parameterizing the peak area. Overall, this result illustrates that the fit standard error can significantly underestimate the measurement uncertainty, specifically for the non-area line shape parameters.

The underestimation of uncertainties in fitted profile parameters based on standard error analyses makes it difficult to compare literature values. Further, these discrepancies will propagate through the uncertainty analysis of the resulting line list. As an example, this issue can be critical to remote sensing operations, which have spectroscopic uncertainty goals for retrieval of quanti-

ties like species concentrations and column-integrated abundance. The contour plots in Fig. 2 provide a general estimate of the true uncertainty in the various HTP line profile parameters as a function of noise in the absorption and frequency axes when simulating and fitting spectra with the full HTP. Because these plots average across all parameters, the variation caused by unique combinations of parameter values is not captured. Nevertheless, the MATS-based analysis can be used to further explore the uncertainty for a specific set of conditions, which will be highlighted below.

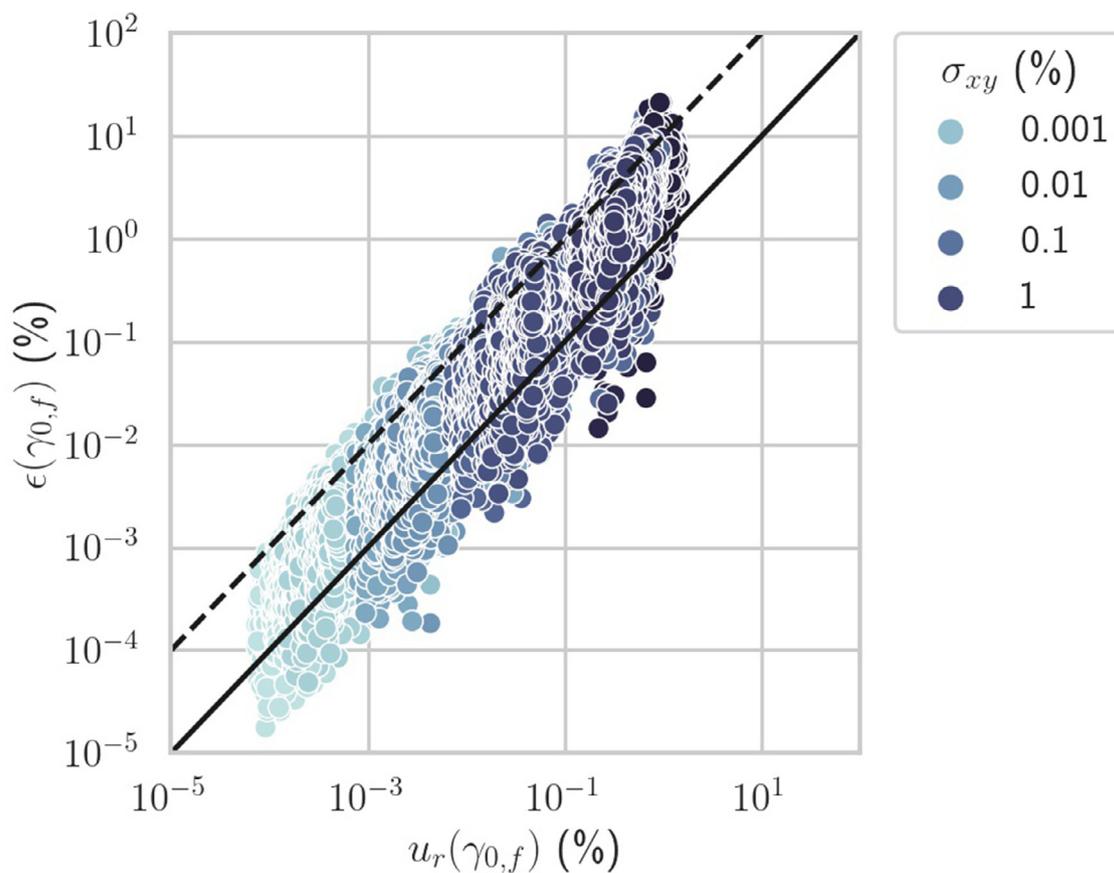
Fig. 2 provides general guidance that is useful for experimental design or as a check of whether a reported fit uncertainty seems to be reasonable. In 2016, Connor *et al.* published an uncertainty quantification of column-averaged dry-air mole fraction of  $\text{CO}_2$  as measured by the NASA Orbiting Carbon Observatory (OCO-2) mission, which included a table of the average fractional uncertainties of the line profile parameters used in their version 7 level-2 algorithm [47]. Although this analysis corresponds to line profiles having fewer parameters than the HTP, Fig. 2 can provide guidance on bounds in frequency and absorption noise that would be necessary to retrieve HTP parameters with specified uncertainties.

The relative uncertainties in line intensity for the OCO-2 bands have reported values between 0.3% and 0.4%, which are driven by uncertainties in the measured peak areas and sample composition. Treating the HTP as a reference profile and proxy for a true line shape, our simulations and multi-spectrum fit analysis show that reproducing the area to the 0.1% level does not require particularly low noise in the frequency or absorption axes. Specifically, we calculate that a frequency noise of less than 2% and SNR of 300 is enough to reproduce the peak area to within 0.1%. However, it is worth noting that while area might have a low relative bias, the use of a more-complicated line profile than is necessary can lead to overfitting. This approach can result in the retrieval of line profile parameters lacking physical meaning and the fitting of non-spectroscopic features, as well as causing convergence issues while fitting. We will discuss problems of this sort in more detail below.

The work by Connor *et al.* also mentions that the reported collisional broadening terms have relative uncertainties of between 0.15% and 0.2%, the pressure shifts have uncertainties between 1.5% and 3%, and the speed-dependent broadening have uncertainties of about 10% [47]. Fig. 2 indicates that to reproduce these uncertainty levels with the full HTP, an experiment would need to have a SNR on the order of  $10^3$  (0.1% absorption noise) and a frequency axis uncertainty that is nominally 0.1% of the Doppler width. The indicated contours can be used to predict how further improvements to the  $x$  or  $y$  noise levels would be expected to improve the parameter uncertainty.

In the OCO-2 version 7 level-2 algorithm, the speed-dependent shift, Dicke narrowing, and correlation parameters were not used. The contour plot for the speed-dependent shift shown in Fig. 2c shows that to measure this term to better than 10%, frequency measurements with uncertainties of less than 0.01% of the Doppler width and SNR values on the order of  $3 \times 10^4$  would be necessary. These figures highlight the experimental challenges, and suggest that frequency noise is a parameter that should not be routinely omitted from uncertainty analyses [48]. Supplementary materials provide additional discussion on the combined impacts of frequency and absorption noise in spectra.

While the Dicke narrowing and speed-dependence terms both account for narrowing mechanisms, the former term exhibits relative errors that are an order of magnitude larger than the latter term, despite being simulated over the same range of values. Also, the absorption-frequency contours of relative fractional error for the Dicke narrowing and correlation parameters are similar. Of the seven fitted parameters considered, for a given level of measurement precision along both axes,  $a_s$  is subject to the greatest relative error.



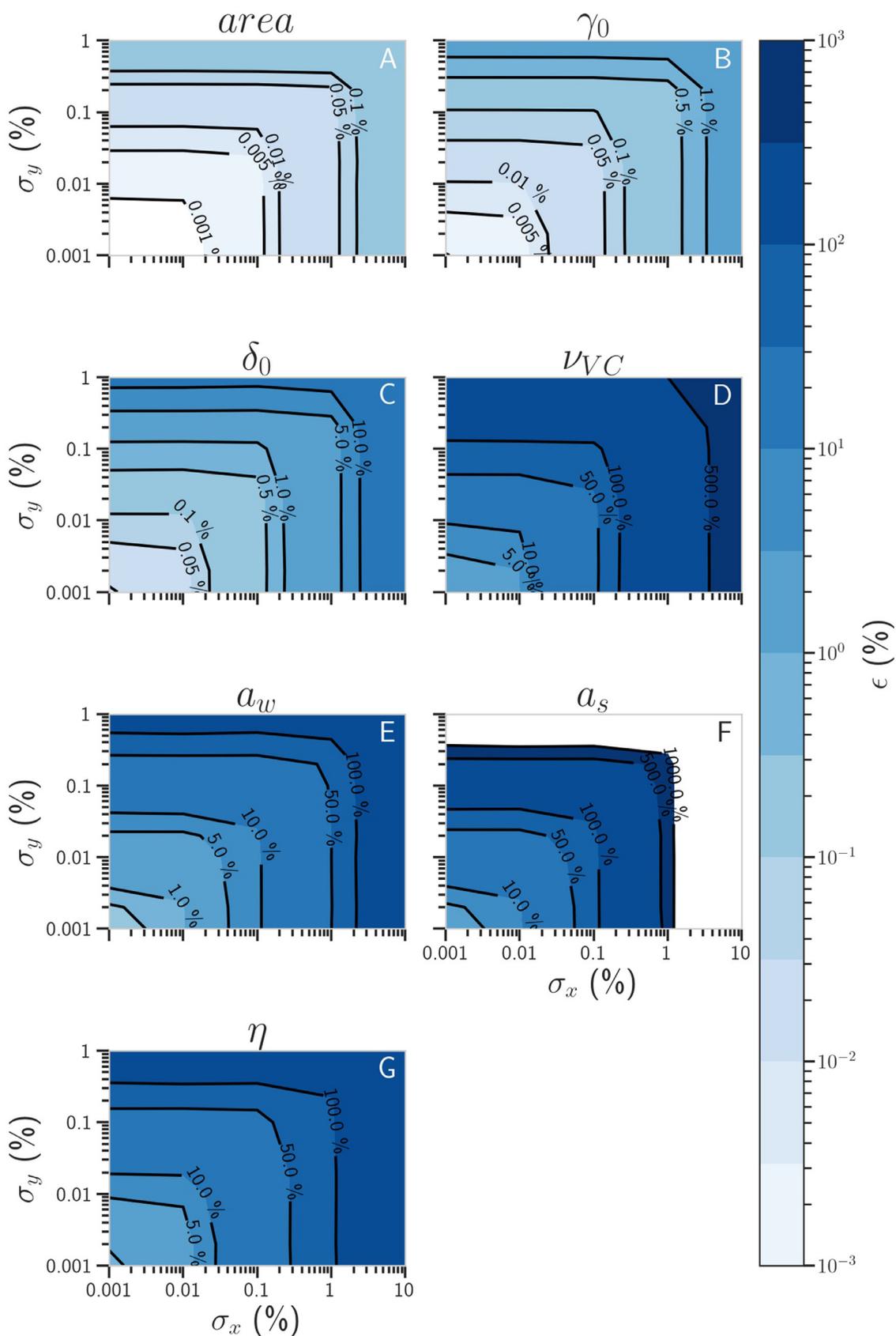
**Fig. 1.** Statistical variation in fit-retrieved values of the collisional width. The magnitude of the relative fractional error  $\epsilon(\gamma_{0,f})$  from replicate fits is plotted against the relative uncertainty,  $u_r(\gamma_{0,f})$  based on the fit-reported standard error. These calculations correspond to fits of the HTP over the full range of simulation conditions given in Table 2. Darker colors in the symbols correspond to increased levels of noise in the spectrum axes. The solid- and dashed-black lines represent ratios of 1:1 and 10:1 between  $\epsilon$  and  $u_r$ , respectively.

In Fig. 3 plot A, the average fitted Dicke narrowing values are plotted versus the corresponding simulated value, with plot B depicting the analogous graph for the correlation parameter. In both plots the error bars correspond to the standard deviation of the fitted values over the corresponding simulated parameter space. In the introduction of the HTP by Ngo et al., the complex-valued, velocity-dependent Dicke narrowing frequency,  $\tilde{\nu}_{VC}(\nu)$  (where  $\nu$  is velocity) at specific pressure is given as a function of  $\nu_{VC}$  and  $\eta$  (both real-valued and independent of  $\nu$ ) and the velocity-dependent quadratic speed dependence and shifting terms,  $\Gamma(\nu) - i\Delta(\nu)$ , as  $\tilde{\nu}_{VC}(\nu) = \nu_{VC} - \eta[\Gamma(\nu) - i\Delta(\nu)]$ . With this formalism, the HTP correlation term  $\eta$  occurs within the velocity integral.

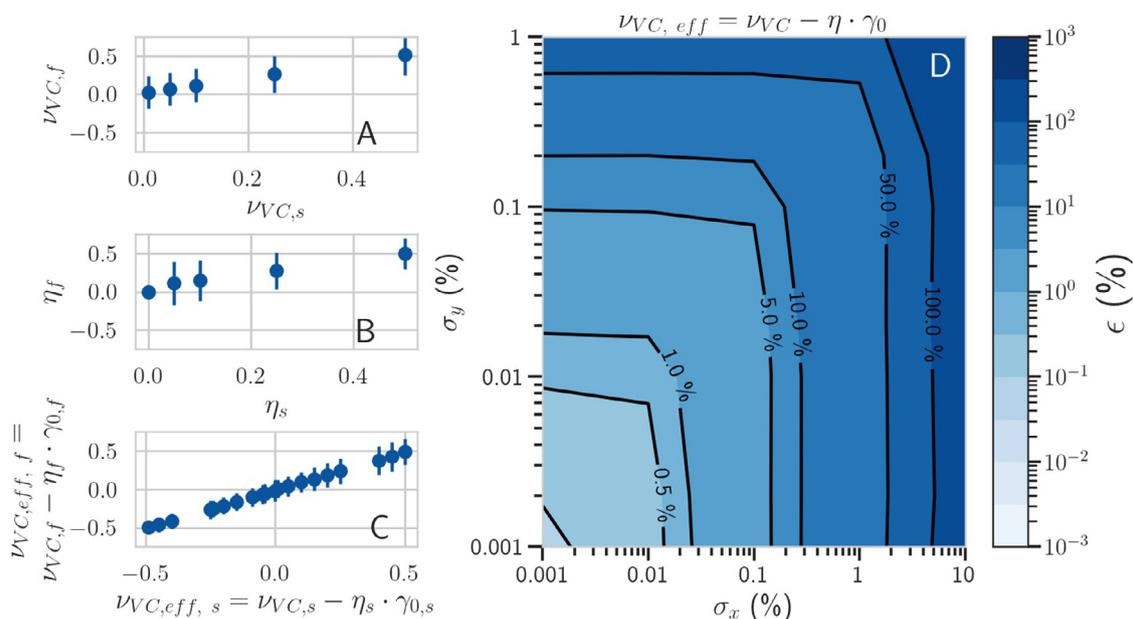
Importantly, the HTP definition of  $\eta$  needs to be distinguished from an analogous parameter used in the partially correlated quadratic speed-dependent Nelkhin-Ghatak profile (pCqSDNGP) described by Lisak et al. and earlier references [49]. Apart from its definition of a correlation parameter (here denoted as  $\eta'$ ), this profile has the same physics and parameter definitions as the HTP. The difference between the mathematical representation of these two profiles is that  $\eta'$  multiplies the velocity-averaged broadening and shifting coefficients, whereas  $\eta$  of the HTP multiplies the corresponding velocity-dependent integrand values as noted above. Consequently, for the pCqSDNGP a complex-valued, Dicke narrowing frequency,  $\tilde{\nu}'_{VC}$ , that is velocity-independent can be written at a specific pressure in terms of the fitted line shape parameters,  $\tilde{\nu}'_{VC} = \nu_{VC} - \eta'(\Gamma_0 - i\Delta_0)$ . Recent work by Stolarczyk et al. provides a detailed discussion about the consequences of choosing the velocity-dependent versus the velocity-averaged definition of the

correlation parameter [50]. Notably, although the HTP and pCqSDNGP model the same correlation effect using a single parameter, their differing definitions means that all other parameters being equal, setting  $\eta$  equal to  $\eta'$  will not give the same line profile, nor will fitting either profile to the other one give identical values of  $\eta$  and  $\eta'$ . Unfortunately, there is no known analytical transformation between the two quantities, making comparison of measurements that report these two quantities difficult to interpret.

While not strictly consistent with the pCqSDNGP, evaluation of the analogous HTP-based quantity,  $\nu_{VC,eff} = (\nu_{VC} - \eta\gamma_0)$  can be interpreted as an effective real-valued Dicke narrowing coefficient that provides valuable insight into the numerical correlation between  $\nu_{VC}$  and  $\eta$  [51]. When comparing plots a-c in Fig. 3, we see that while the individual values of the  $\nu_{VC}$  and  $\eta$  show substantial variation, the fitted values of  $\nu_{VC,eff}$  are highly correlated and nearly equal to the simulated quantity and exhibit far less variation. The contour plots of Fig. 3d also show an order-of-magnitude decrease in the uncertainties in  $\nu_{VC,eff}$  compared to the  $\nu_{VC}$  and  $\eta$ . Simultaneous fitting of both  $\nu_{VC}$  and  $\eta$  is difficult in most cases because of numerical correlation. This complication suggests that a best practice in fitting would involve constraining one or the other of these two parameters to a physically meaningful value. For  $\nu_{VC}$ , this constraint could be a calculated or measured diffusion coefficient. However, the hard-collision approximation assumed in derivation of the HTP can cause  $\nu_{VC}$  to differ from diffusion-based values [28,36,52] when the observed Dicke narrowing is dominated by soft-collisions. If  $\eta$  is constrained to 0, i.e. by using the SDNGP, seemingly unphysical negative values for fitted values of  $\nu_{VC}$  may happen [51]. This result can occur when



**Fig. 2.** Contour plots showing relative fractional error in HTP line shape parameters as a function of frequency and absorption noise; A - area, B - collisional Broadening, C - pressure shift, D - Dicke narrowing, E - speed-dependent broadening, F - speed-dependent shift, G - correlation parameter. Contours are labeled with half-decade divisions, where the darker-the-color on the contour plot the higher-the-fractional-error in that parameter.



**Fig. 3.** Correlation between the Dicke narrowing and correlation parameter in the HTP. A: Plot of the average Dicke narrowing determined by the fit as a function of the simulated Dicke narrowing value. B: Plot of the average correlation parameter determined by fitting as a function of simulated correlation parameter. C: Plot of the average effective Dicke narrowing determined by fit as a function of the average simulated effective Dicke narrowing value. In A, B and C, the error bars correspond to the standard deviation of the calculations for each condition and case. D - Contour plots showing the fractional error in the effective Dicke narrowing as a function of frequency and absorption noise.

physical correlations between velocity- and state-changing collisions are large. It has been previously noted that if the collisional broadening term is much larger than the diffusion coefficient, then the correlation parameter can be assumed to be negligible [52]. The full factorial nature of the parameter space in this study makes it difficult to discuss sensitivity to  $\eta$  because a range of physically likely and unlikely parameter combinations were explored, i.e. large values of  $\eta$  for  $\nu_{VC}$  that were much smaller than  $\Gamma_0$ . Because of the large numerical correlation between these parameters, the effective Dicke narrowing term will be used in the present analysis.

### 3.2. Equivalence of HTP limiting-case profiles in spectra of finite SNR

The previous section explored the expected errors in line profile parameters where spectra were fit with the HTP and its limiting-case profiles, for spectra simulated by the HTP with both absorption and frequency noise. This is useful for guidance for the generation of complete HTP-based line lists. However, if the purpose of a specific study is to extract line intensities, then a simpler line profile might be able to achieve the same level of uncertainty as the HTP without the risk of overfitting and being limited by numerical convergence issues encountered by fitting a high-order profile to noise-limited spectra.

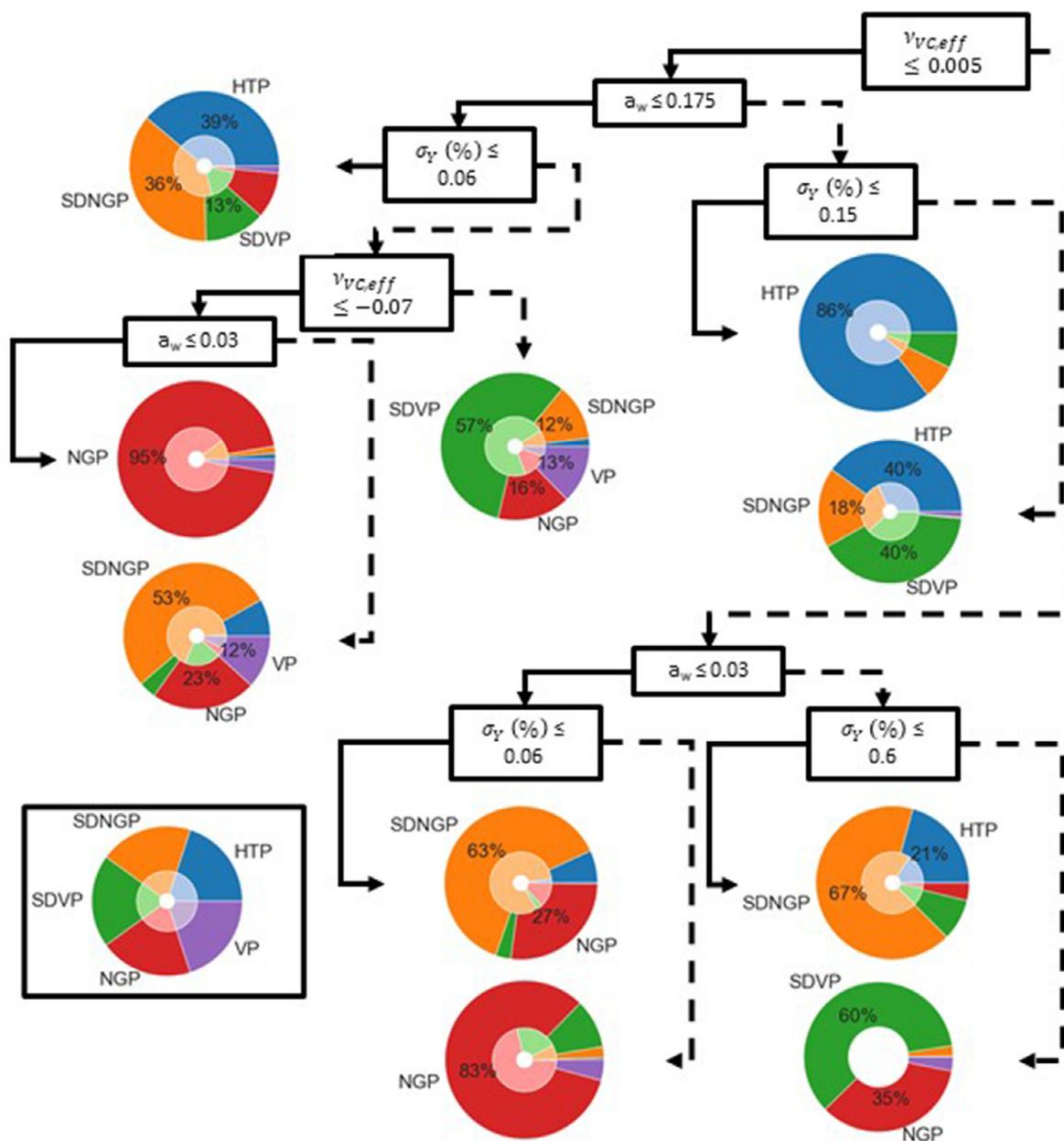
In these simulations, peak area is a proxy for line intensity that would be measured from experimental spectra upon division by the absorber number density. An optimal line profile for retrieving the peak area from a measured spectrum would be the least complicated line shape required to reduce the area bias towards zero and to achieve noise-limited precision when averaged across all pressures. The fractional error in the pressure-averaged areas, as well as the uncertainty in this value are used as the metrics to compare results obtained using the HTP, SDNGP, SDVP, NGP, and VP. Cases where these data were not available for all line profiles were omitted from the analysis.

Line profiles were considered pairwise in terms of complexity beyond the VP: i.e. VP compared to SDVP, SDVP compared

to SDNGP, etc. To classify a line profile as being the preferred profile of each pairwise comparison, the retrieved area bias had to both decrease with an increase in complexity and differ by a statistically significant amount compared to the less complex line profile. Because the SDVP and NGP both account for a single narrowing mechanism past the VP, they were considered in independent tracks (VP  $\rightarrow$  NGP  $\rightarrow$  SDNGP  $\rightarrow$  HTP and VP  $\rightarrow$  SDVP  $\rightarrow$  SDNGP  $\rightarrow$  HTP) and then the resulting optimal line shapes were evaluated. This process labelled each unique combination of parameters with an optimal line profile that minimized the bias in retrieved areas while limiting the line profile complexity.

A decision tree machine-learning algorithm [53] was then used to determine the features that were important in the classification of the optimal line profile. These results were used to generate a flowchart mapping how these features are separated over the parameter space. While more sophisticated machine learning algorithms could provide better classification accuracy, the decision tree is a transparent algorithm that translates easily to an analog visualization of the optimal lineshape across the physical parameter space, which is the most functional output for this study. Additional information about the generation and optimization of the decision tree reported in Fig. 4 is available in the supplementary materials.

Traditionally in machine learning studies, the available data would be separated into training and validation sets to ensure that the model is able to generalize beyond the data used to train the algorithm. The computational nature and full-factorial structure of this study means that separating the simulations outlined in Table 2 into training and validation sets would yield parameter and noise levels at the same discrete magnitudes, which does not capture the continuous nature of these variables. To overcome this issue, a validation set was generated by simulating and fitting spectra, within the parameter and noise bounds of the original dataset, but at different increments. This resulted in over 40,000 unique parameter combinations that were successfully analyzed for each of the five considered line profiles (Supplementary Materials Table 1) to provide a more robust validation data set to test



**Fig. 4.** Decision Tree indicating the least complex HTPlimiting-case line profile capable of reproducing the area simulated with the HTP under different parameter and environmental cases. Solid lines indicate the true cases, which are always on the left, whereas dashed lines indicate the false cases, which are always on the right. The largest section of the pie indicates the optimal line profile in that leaf of the decision tree. The outer ring of the pie corresponds to the training set result, where the inner circle corresponds to the validation set result.

the decision tree. This validation set was only analyzed after the optimization of the decision tree with the training dataset.

The machine learning algorithm indicated that the features best separating the data were the frequency and absorption noise, along with  $a_w$  and  $\nu_{VC,eff}$ . This finding is expected because it accounts for the line width parameters that most strongly affect the peak area and the various noise terms. The highest-level split separated the 10% frequency axis noise instances from the rest of the data, indicating that in this domain the frequency noise is large enough that VP, NGP, and SDVP adequately reproduced peak areas modelled with the HTP.

Because this high-frequency noise limit was an extreme case, we omitted these cases from the data to reevaluate feature importance and to generate a simpler decision tree. When including cases with less than or equal to 1% frequency noise, the absorp-

tion noise,  $a_w$ , and  $\nu_{VC,eff}$  were sufficient to separate the different optimal line profiles. The resulting decision tree is shown in Fig. 4, where the largest section of each pie chart shown in each leaf of the tree indicates the optimal line profile for the corresponding set of conditions. The outer ring of the pie chart indicates the proportion of the optimal line profile cases in the training set, where the inner ring indicates the same information for the validation set. Overall, there is good qualitative agreement between training and validation sets with consistency between the parameter space where there is a clear optimal line profile, where some profiles are deemed sub-optimal, and where a few profiles give similar performance. Fig. 4 provides general guidance on the optimal choice of line profile for modeling peak area, whereas a tool like MATS can provide more precise guidance for specific molecules and exper-

imental conditions, particularly in parameter combinations where the decision tree suggests several similarly performing line profiles.

To navigate through the decision tree, it is necessary to know approximate values for the  $a_w$  and the  $\nu_{VC,eff}$ , where we note that the latter term is a pressure-normalized quantity. Eq. (1) provides a theoretical value for the speed-dependence term based on the collisional broadening temperature exponent ( $n$ ), the mass of the perturber ( $m_p$ ), and the mass of the absorber ( $m_a$ ) [54]. The diffusion coefficient,  $\nu_{diff}$ , can be calculated by Eq. (2), where  $k_B$  is the Boltzmann constant,  $T$  is the temperature,  $D$  is the diffusion coefficient, and  $p$  is the pressure. The calculated diffusion coefficient can be used as an approximation for  $\nu_{VC}$  to provide an estimate of  $\nu_{VC,eff}$  under different assumptions for the value of  $\eta$ .

$$a_w = (1 - n) \cdot \frac{2}{3} \cdot \frac{m_p/m_a}{1 + m_p/m_a} \quad (1)$$

$$\nu_{diff} = \frac{k_B T}{2\pi m_a D p} \quad (2)$$

The top-level separation of the decision tree in Fig. 4 is based on the effective dimensionless Dicke narrowing term being less than 0.005, which can also be interpreted as  $\eta\gamma_0$  being greater than  $\nu_{VC}$ . If this effective Dicke narrowing term is larger than the collisional broadening term times the correlation parameter, then the full HTP is not necessary to reproduce the area of spectra simulated with the full HTP. Within this branch there are further separations between the performance between velocity-changing and speed-dependent narrowing mechanisms based on the magnitude of  $a_w$ . In both cases, as the absorption noise decreases the optimal line profile becomes the SDNGP.

In the case where the dimensionless, effective Dicke narrowing parameter is less than 0.005 (i.e.  $\eta\gamma_0 > \nu_{VC}$ ) then the full HTP is necessary to reproduce the area simulated with HTP under low noise levels. The definition of low noise changes based on the value of  $a_w$ . If the speed-dependent broadening term is  $< 0.175$ , then the HTP is the preferred line profile for SNR greater than about 1650, but if the speed-dependent broadening term is  $\geq 0.175$  the HTP is the preferred line profile for SNR values greater than about 650. For the higher-noise cases, if the speed-dependent broadening is large, then the SDVP is the most appropriate line profile choice. The SDVP is also the optimal profile when the speed-dependent broadening is smaller, but the effective Dicke narrowing is only slightly negative. If the effective Dicke narrowing is significantly negative, for SNR values less than 1650, then small values  $a_w$  indicate preference of the NGP, whereas large speed-dependent parameters indicated lower area biases with the SDNGP.

Fig. 4 provides general guidance in minimizing the bias of retrieved areas balanced with the degree of line profile complexity across a wide range of fitted parameters. However, the pie charts in the leaves of this decision tree indicate that there are still variations within these segments. Additionally, comparison between the training and validation set outcomes provides an additional measure of this ambiguity in the optimal line profile. For example, in the case where the effective Dicke narrowing is less than zero, the speed-dependent broadening is small and the SNR is high, the HTP is the optimal line profile, although the SDNGP is the best option for a significant number of cases. In the validation set, the SDNGP cases make up a larger percentage of the cases in that condition. This means that both Dicke narrowing and speed-dependence need to be considered when analyzing the data in this domain, however depending on the specific molecule and experimental conditions correlation between velocity-changing and internal-state-changing collisions may or may not need to be considered. While this figure can give general guidance, detailed exploration and error analysis would need to be done on specific instances using a tool like MATS, which is introduced in the next section. A clear finding not

shown in Fig. 4 is that the VP was only chosen for the optimal line shape in reproducing the area for cases where the frequency noise was equal to 10%. Also, when the frequency noise was this large either the VP, NGP, or SDVP were adequate to reproduce the area simulated with the HTP.

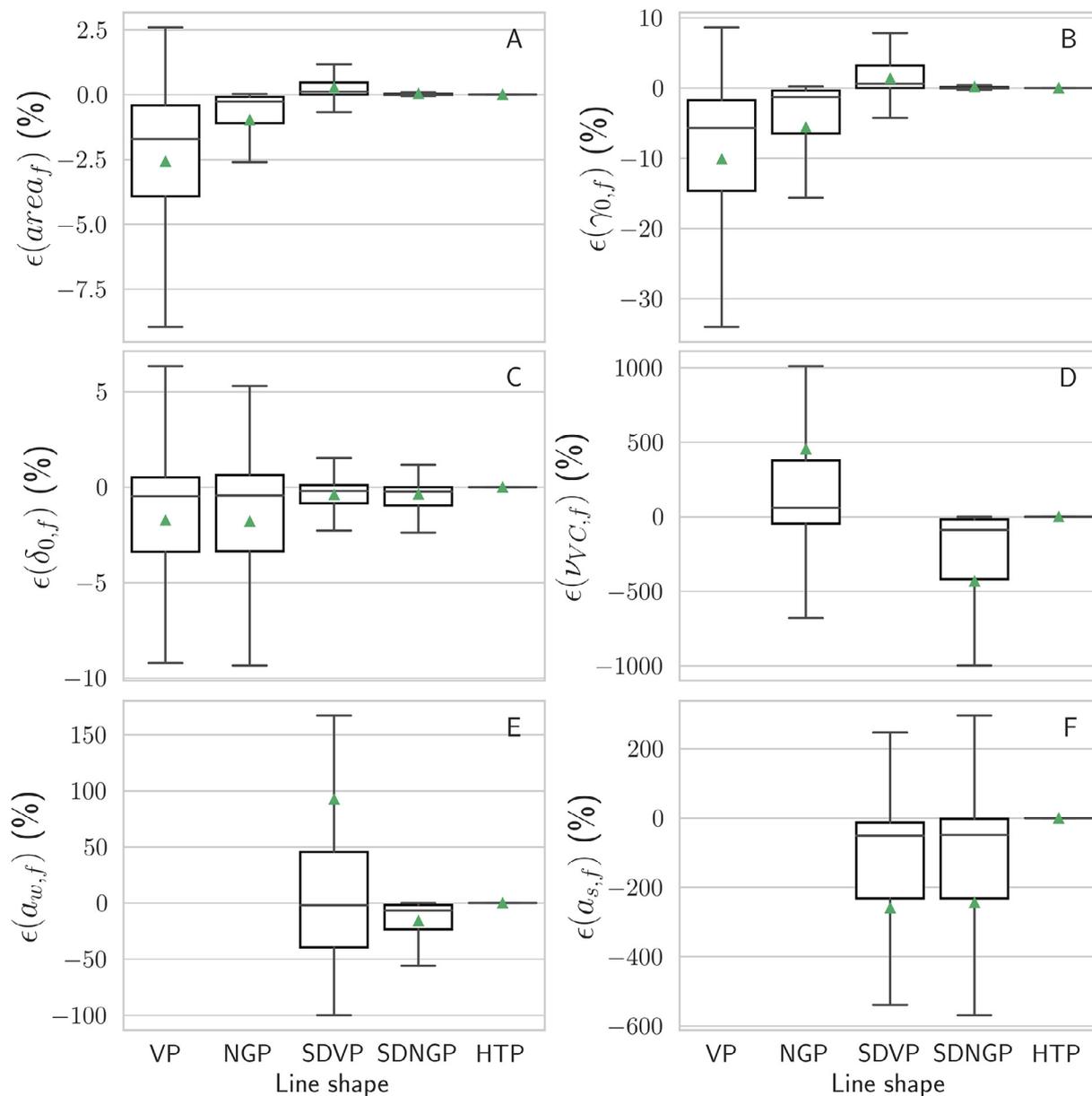
Using the limiting-case profiles derived from the HTP has advantages in terms of balancing the complexity and uncertainty in data products in noise-limited spectra. However, it should be highlighted that the profile might generate an equivalent area, but the difference in parameterization leads to non-equivalent parameter values, despite the same nomenclature and terms being used. Fig. 5 depicts the differences in retrieved parameter values compared to simulated values in noise-free spectra across the range of line shape parameter combinations. This examination provides some guidance on the magnitude of the bias from this non-equivalence between the HTP and limiting HTP profiles.

For all parameters, we see that there is zero fractional error (within the numerical precision of the calculation) in the case where spectra were simulated and fit with the HTP, which should be the case in noise-free spectra. Fig. 5a highlights the systematic differences in retrieved areas across the limiting case profiles, with median differences of -1.71% for the VP, -0.26% for the NGP, 0.11% for the SDVP, and 0.00% for the SDNGP. The magnitudes of these differences are similar to differences reported in the literature [51]. The fact that both the NGP and SDVP impose systematic biases in the area is a key finding and supports our finding that either the SDNGP or HTP profiles are necessary to measure the area at the same level as the HTP at lower absorption noise levels, when the noise is smaller than the systematic difference between profiles.

The trends in the fractional error in  $\Gamma_0$  across the various line profiles are like those of the fitted peak area. The VP significantly underestimates the HTP-simulated  $\gamma_0$  value by a median amount of -5.7%. The NGP also underestimates the  $\gamma_0$  term by a median value of -1.3%, whereas the SDVP overestimates it by 0.6% and the SDNGP has a 0.0% median bias. This behavior is to be expected, because not accounting for any, or only accounting for one, of the narrowing mechanisms leads to compensation in the other width parameters upon fitting.

For  $a_w$  and  $\nu_{VC}$ , the parameter bias between line profiles can only be explored in line profiles where that phenomenon is parameterized. The error in the speed-dependent broadening term using the SDVP has a median value of -1.7% but shows a range of positive and negative values. This contrasts with the SDNGP, which has -6.7% bias, but always reports speed-dependent broadening terms that are less than those simulated with the HTP. This systematic effect occurs because the SDVP does not account for both the Dicke narrowing and correlation terms, whereas the SDNGP only neglects the correlation parameter [51]. The Dicke narrowing also has SDNGP fit values that are always less than the simulated value, with a median value of -89%. However, the NGP fit bias in the Dicke narrowing parameter tends to be skewed larger values than the simulated values by a median value of 76%. These differences are also caused by neglecting the speed-dependence and the correlation parameter. Ngo et al. reported computational studies of CH<sub>4</sub> with multiple collisional partners fit with the HTP and limiting case line profiles. In their fits of the HTP to various calculated line shapes they found that the Dicke narrowing parameter was always larger for models that account only for Dicke narrowing compared to the those that account for both Dicke narrowing and speed-dependence [48].

Investigation of the dependence of the fitted collisional shift term on fitted line profile is interesting because the VP and NGP show a larger bias, with median values of -0.49% and -0.44% respectively. These differences occur because the speed-dependent shift and correlation parameters are not taken into account. The SDVP and SDNGP also show a similar but smaller bias, with me-



**Fig. 5.** Plots of the fractional error,  $\epsilon$ , in HTP parameters from fits with HTP limiting-case line shapes in noise-free spectra across a range of parameters; A – Area, B – collisional broadening, C – pressure shift, D – Dicke narrowing, E – speed-dependent broadening, F – speed-dependent shift. The boxes span the upper and lower quartiles of the data, the median is shown with the solid line, the triangle shows the mean, and the whiskers are one and half times the interquartile range.

dian values of -0.20% and -0.23% respectively, caused by not accounting for the correlation parameter. This result is contrary to a prior literature claim [52] in which it was stated that the fitted collisional shift values were independent of the line shape model. This independence will occur when the speed-dependent shift term is relatively small or when the correlation parameter is too small to introduce substantial asymmetry in the line profile. The bias in speed-dependent shift  $a_s$  can be evaluated for the SDVP and SDNGP cases, with median values of -51% and -48%, respectively. This result was expected because the Dicke narrowing term has no impact on  $a_s$  and therefore the difference between the SDVP and SDNGP would be expected to be small. However, the similar bias with respect to the HTP value is driven by neglect of the correlation parameter. Again this result indicates that the speed-dependent shift will only be independent of the chosen line profile when the correlation parameter is small [52].

This section has highlighted that there are regions where the noise and set of considered parameters yield approximately equivalent areas when fitting limiting-case HTP profiles to spectra simulated with the HTP. However, we have also shown that when fitting these HTP-derived profiles to simulated HTP spectra which are noise-free, there are persistent biases in the area and line profile parameters. It is the balance of these two factors that determines the appropriate line profile for modeling experimental spectra. For example, in fitting a spectrum with numerous lines of various strengths, a consistent line profile should be used in order to avoid the introduction of line-profile-dependent biases, where higher-order line shape parameters for weaker lines could be constrained to appropriate parameter estimates. Multi-spectrum simulation and fitting tools, such as MATS, can be used to explore the uncertainty related to the specific fitting procedure of a given set of experimental spectra.

### 3.3. Multi-spectrum Analysis Toolkit for Spectroscopy (MATS)

To this point we have focused on general guidelines for how evaluating noise in the frequency and absorption axes will affect uncertainties in the fitted parameters when using the HTP as a reference line profile. We have also considered under what set of parameters and noise conditions do the limiting-case profiles derived from the HTP adequately model peak area, and the range of expected systematic differences between fitted parameters obtained from the HTP and limiting-case profiles. Although this information is valuable for a global perspective on the complexities of retrieving physical parameters from least-squares fits, in this section we introduce the Multi-spectrum Analysis Tool for Spectroscopy (MATS) [37]. This new analysis package was developed for experimentalists to aid in data reduction and experimental design and is a platform for fitting the HTP to experimental absorbance data as well as to synthetic, noisy spectra. In its current version MATS is capable of simulating and fitting linear absorbance spectra. MATS uses HAPI [39] for the HTP definition to provide consistency with HITRAN [55]. Complete documentation, along with link to the GitHub repository, is available on the MATS landing page (<https://pages.nist.gov/MATS/>). Here, we provide an overview of the functionality and examples of its uses.

Unlike the simulations reported above, the spectra that are simulated and/or fit in MATS use dimensional units that are consistent with HITRAN [55] as opposed to the corresponding dimensionless quantities employed in the preceding simulations. The MATS program is based on spectrum objects, which are defined not only by their wavenumber and absorption data, but also by meta information regarding spectrum pressure, temperature, baseline characteristics, and sample composition. Each spectrum can be modeled using a combination of spectroscopic parameters, including their temperature and broadener dependences. These relations will be common across all spectra being analyzed in the multi-spectrum fit and spectrum-dependent parameters, such as polynomial baseline terms, baseline etalons, x-axis shift term, sample composition, pressure, and temperature. These spectrum objects can be generated by reading in experimental or simulated data within the MATS framework. The simulated spectrum objects allow for uncertainty to be added to the absorption and frequency axes, as well as to the sample composition, pressure, and temperature. This capability is useful for performing uncertainty analysis in the same framework as primary data analysis by simply switching from experimental to simulated spectra. The MATS spectrum objects are collected into a dataset object, which will use a common line-by-line spectroscopic parameter list for the multi-spectrum fitting.

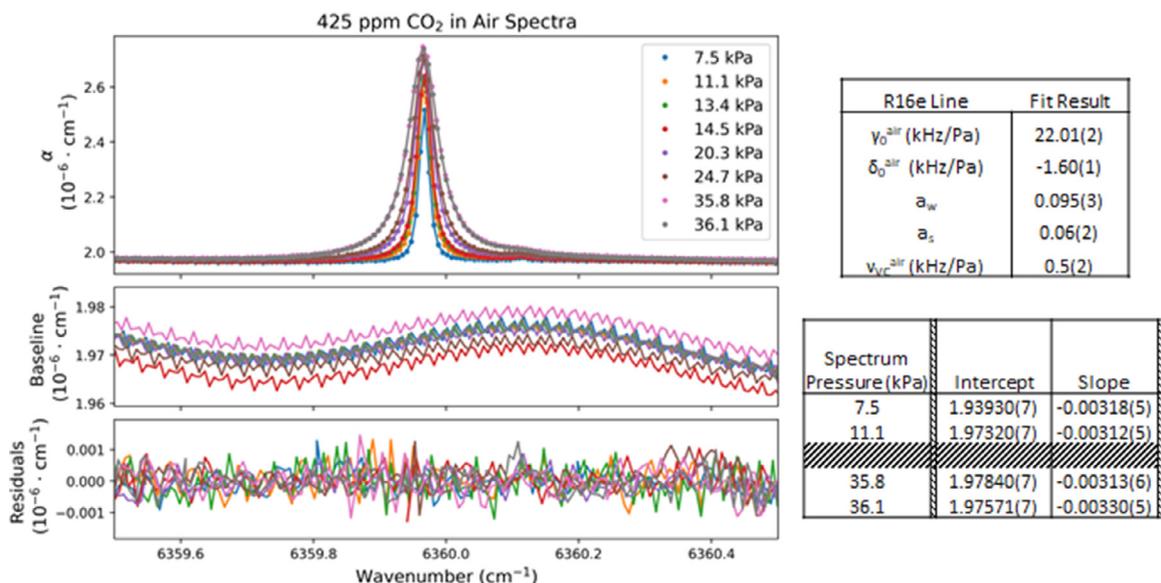
Separate lists for the baseline and spectroscopic parameters are generated specifying which parameters are floated during fitting. These floated parameters are updated with the new fit results, including fit uncertainties, after fitting. By default, although multi-spectrum environmental pressure and temperature constraints are imposed during the fitting step, for each spectroscopic parameter it is possible to remove this constraint and fit a unique parameter value for each spectrum. Additionally, the LMFIT package [40] that is used for fitting in MATS allows for specific constraints to be imposed on the parameters, e.g. constraining the frequency shift to be common across all spectra to account for a calibration error in measured wavelength or setting the advanced line parameters (beyond the VP) to be equivalent for all broadeners. Additionally, the Python framework allows for the functionality available in the MATS package to be readily incorporated into existing data analysis frameworks. These capabilities provide flexibility to test and adapt to experimental and data-driven needs of specific experiments.

As discussed above, the ASCENDS CO<sub>2</sub> transition was used as a representative case to provide context to the Doppler-width-normalized set of parameters. Here we use the same transition to

provide an example of multi-spectrum fits using MATS. The experimental spectra are a subset of the full-band spectra, which were acquired over a broad pressure range on samples containing 425 μmol of CO<sub>2</sub> in air and originally reported in Ref. [42]. The fits used the line shape parameters published in Refs. [42,46] as an initial line list with HITRAN 2016 values used for non-target transitions. In addition to fitting the ASCENDS line parameters, the fits include a linear baseline unique to each spectrum and three weakly perturbing etalon spectra. Each etalon is attributed to a secondary optical cavity of length,  $L_{\text{et}}$ , involving a high-reflectivity ring-down mirror surface and another retro-reflecting external optical element in the experiment, with a fixed etalon period given by  $c/(2L_{\text{et}})$ . This analysis also leverages the flexibility of MATS by removing the multi-spectrum constraint for the line intensity to account for gas sampling variability and by constraining the etalon periods to be consistent across all spectra while also allowing the phases of the two shorter-period etalons to vary over spectrum sub-segments. The results of fitting these spectra with the SD-NGP are shown in Fig. 6 and result in QFs ranging between 1350 – 1900. The analysis script used here is provided as an example with MATS. We note that omission of the two short-period etalons from the fit typically degrades the QF by a factor of three or more.

While the MATS tool is useful for fitting line shape models to experimental data, the ability to simulate spectra is also valuable for experimental design and uncertainty analysis. In the simulations presented above, seven spectra distributed over the normalized pressure range (0.1 – 100)  $\Gamma_0/\Gamma_D$  were used in each multi-spectrum fit, which is a broader range than is typically accessible for most experiments. However, it is necessary to use a broad pressure range to adequately constrain and differentiate between the various higher-order parameters in the HTP and its limiting cases [28]. The MATS tool can be used to simulate spectra at the known frequency and absorption noise of an experiment for arbitrary sets of pressure. This approach enables one to estimate the number of pressure conditions within the experimentally feasible pressure range required to meet the measurement goals. Fig. 2 highlights generally how reduction in the absorption or frequency uncertainty would be expected to reduce the uncertainty of a given parameter. The MATS program enables experimentalists to more specifically determine whether a particular improvement in the experiment would be helpful. The ASCENDS line data fit in Fig. 6 had a reported  $\sigma_x$  of 0.007%, an observed  $\sigma_y$  of approximately 0.06%, and spanned the range 0.93 – 4.5 for  $\Gamma_0/\Gamma_D$ . Synthetic spectra can be generated and analyzed in the same analysis framework to investigate the impact of improving noise levels and adding additional spectra to sample a broader  $\Gamma_0/\Gamma_D$  range. To this end, synthetic spectra are generated at experimental conditions to ensure similar results as the fits to experimental data and to establish a benchmark for comparison to the proposed experimental changes. For this experiment, reducing  $\sigma_y$  to 0.02% and including additional spectra at 0.66, 1.33, 3.33 kPa, to include more measurements in the Doppler regime, are predicted to reduce uncertainties in the spectroscopic parameters by factors of 2.8 to 3.7. The largest improvement would be to the Dicke narrowing parameter, which is aided by not only the higher SNR, but also by the inclusion of additional low-pressure spectra that fall in a domain where the Dicke narrowing mechanism dominates over the speed-dependent narrowing.

In addition to providing useful technical information for experimental design, the MATS program can also be used for uncertainty analysis of parameters derived from measured spectra. Fig. 4 highlights that there are combinations of parameter ranges and uncertainty levels where a simpler line profile can be used to model the area simulated with the HTP, whereas those of Fig. 5 show that in the absence of significant noise there are systematic differ-



**Fig. 6.** Multi-spectrum fits of the ASCENDS CO<sub>2</sub> line using MATS. Top panel plot: Spectra and model for ASCENDS transition. Center panel: Baseline including the three etalons fit during analysis. Bottom panel plot: Fit Residuals for SDNGP. Top table, right: Reports the line shape parameters from SDNGP fits. These values differ from values reported in Refs. [42,46] based on pressure range, constraints used, and inclusion of line-mixing. Additionally, the reported values are to depict MATS outputs and should not be used as reference line shape parameters. Bottom table, right: Portion of the spectrum-specific parameters determined in fits, not shown are the fitted etalon amplitudes, periods, and phases, along with the constrained pressures, temperatures, and sample mole fractions.

ences between the areas fit with various HTP limiting-case profiles. These conclusions are somewhat counterintuitive and require experimentalists to conduct robust uncertainty assessment by quantifying the impact of line shape choice on reported line intensities and/or quantifying sensitivity to parameter correlation or overfitting of the measured line shape parameters.

In a recent study of CO<sub>2</sub> line intensities conducted by our group [56], single-spectrum fits were conducted on line-by-line spectra across the 2.06  $\mu\text{m}$  band of CO<sub>2</sub>. Experimental results were collected and analyzed by two individuals. Although each individual fit their spectra with the SDNGP profile and fixed the speed-dependent broadening parameter to improve fit convergence, different constraints for the  $J$ -dependencies of  $a_w$  were used in each case. To estimate the systematic differences in peak area associated with these different constraints, we treated the HTP as truth and did MATS simulations with and without noise to assess the magnitude of the effect. While there are known theoretical estimates for  $a_w$  and  $\nu_{VC}$ , there are no good estimates for the correlation parameter,  $\eta$ . To account for this uncertainty, noise-free spectra were simulated with  $\eta$  varied uniformly between 0 and 1, and then fit with the SDNGP with fixed  $a_w$  (to both  $a_w$  values used in experimental fits). Those assumed correlation values that caused the modeled QF to drop below the experimental SNR were eliminated as possible values because they would have resulted in systematic residuals in the fits to measured spectra. The remaining model calculations provided both a mean bias in line intensity caused by the choice of line profile, as well as an uncertainty based on the spread in intensities caused by our original assumption that  $\eta = 0$ . The results of this analysis were included in the uncertainty analysis reported in the 2.06  $\mu\text{m}$  CO<sub>2</sub> band line intensity study [56].

The MATS tool can also be used to perform a Monte-Carlo analysis to ensure that the reported fit uncertainties of the measured parameters are of the same magnitude as the standard deviation in retrieved parameter values across multiple simulated fits at conditions matching the experiment. In Fig. 1 we illustrated that the uncertainty of a parameter reported by the fit (based on the assumption of no parameter correlation) can significantly underestimate the uncertainty obtained when parameter correlations are taken into account. Repeating fits on synthetic spectra

with “known” parameter values allows for insight into the extent to which the uncertainty reported by the fitting algorithm underestimates the “true” uncertainty that accounts for numerical correlations between parameters. In a recent study by our group of the 1.27  $\mu\text{m}$  absorption band of O<sub>2</sub> [35], Monte Carlo simulations were conducted for all reported transitions at the frequency axis noise, absorption axis SNR, sampling density, and pressures in the experiment, where we treated the experimental line list as truth. The standard deviation of the ensemble of Monte Carlo simulations compared to the average standard uncertainty reported by the least-squares fits provided a measure of  $\varepsilon/u_r$  as shown in Fig 1 for each fitted parameter and line. The resulting  $\varepsilon/u_r$  values were then used to rescale the experimental standard errors reported by the least-squares fitting algorithm to provide more realistic uncertainties on the line shape parameters that account for numerical correlation between parameters such as Dicke narrowing and speed-dependent narrowing.

#### 4. Conclusions

The IUPAC-recommended Hartmann-Tran profile has emerged as a widely utilized spectroscopic model profile for high resolution spectroscopy with the caveat that its use requires high signal-to-noise ratio data, adequate constraints, and a wide pressure range [28]. Here we used the HTP to generate synthetic spectra across a broad range of parameters to explore how noise affects fitted parameters from this profile and its limiting cases.

In noisy spectra that were both simulated and fit with the HTP, the statistical uncertainty of each fitted parameter was calculated as a function of noise in both the frequency and absorption axes to generically quantify how randomness in these observables affects parameter retrievals. These results indicated relatively large systematic biases in the retrieved Dicke narrowing and correlation parameter terms, which was assigned to numerical correlations between these quantities. This numerical correlation suggests that in practice either the Dicke narrowing parameter or correlation parameter should be constrained to a physically meaningful value.

The HTP has several limiting cases that correspond to other commonly used line profiles. The equivalence between the peak

areas reported by fits with these limiting case profiles was explored by finding the noise and range of parameter combinations, where limiting cases of the HTP were found to be equivalent to the HTP in their ability to fit peak areas. While the single line peak areas might be equivalent in the presence of noise, there are systematic differences in the parameters including the areas, between the HTP and the various limiting profiles in noise-free spectra. These complications highlight the importance of experimentalists and archivists reporting self-consistent line lists and investigating whether the profile choice biases reported line intensities.

This paper also introduces a Multi-spectrum Analysis Tool for Spectroscopy (MATS) [37], which provides experimentalists with the ability to simulate and/or fit spectra to further explore the use of the HTP for specific experiments. Many laboratories that provide spectroscopic reference data typically use custom software [37,57–61] for conducting multi-spectrum HTP fits with known and unknown differences in how the line shape and other parameters are defined. Differences between models used in the fitting and parameterization for spectroscopic databases can lead to incompatibility in line parameters and down-stream algorithms such that they cannot be used to reproduce fits to experimental data. MATS is a free open-source multi-spectrum fitting tool using the standardized advanced line profile definitions and parameterization provided by HAPI [39]. This tool allows for individual experimentalists to analyze spectra and conduct robust uncertainty quantification by enabling development of experiment-specific data analysis software. Additionally, open-source publication of multi-spectrum simulation and fitting software helps in transparency and facilitates interlaboratory collaboration and consistency across multi-spectrum fitting approaches [49,57,60].

This work provides generalized guidance on the use of the HTP and HTP-limiting-case line profiles in the analysis of finite signal-to-noise ratio spectra and introduces an accessible multi-spectrum analysis tool to enable more specific studies. These efforts support the establishment and dissemination of best practices for multi-spectrum fitting, more robust uncertainty analysis, and standardization of advanced line shape definition and parameterization – all of which are critical to increasing the consistency and quality of spectroscopic reference data.

### Declaration of Competing Interest

The authors declare no competing interests.

### CRediT authorship contribution statement

**Erin M. Adkins:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing, Visualization, Software, Formal analysis, Data curation. **Joseph T. Hodges:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – original draft, Writing – review & editing.

### Acknowledgements

We acknowledge funding from the NIST Greenhouse Gas and Climate Science Program and from the National Aeronautics and Space Administration (NASA) [contract [NRA NNH17ZDA001NOCO2](#)]. The authors would like to thank William Krekelberg, Jacob Monroe, and David Sheen of the Chemical Informatics Group at NIST for aiding in the advanced development of MATS. Jacob Monroe and Scott Diddams commented on the manuscript.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jqsrt.2022.108100](#).

### References

- [1] Gordon I, et al. The HITRAN2016 molecular spectroscopic database. *J Quant Spectrosc Radiat Transfer* 2017.
- [2] Jacquinet-Husson N, et al. The 2015 edition of the GEISA spectroscopic database. *J Mol Spectrosc* 2016;327:31–72.
- [3] Endres CP, et al. The cologne database for molecular spectroscopy, CDMS, in the virtual atomic and molecular data centre, VAMDC. *J Mol Spectrosc* 2016;327:95–104.
- [4] Pickett HM, et al. Submillimeter, millimeter, and microwave spectral line catalog. *J Quant Spectrosc Radiat Transfer* 1998;60:883–90.
- [5] Tashkun SA, et al. CSDS-296, high-resolution carbon dioxide spectroscopic databank: an update. *J Quant Spectrosc Radiat Transfer* 2019;228:124–31.
- [6] Tennyson J, et al. The ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J Mol Spectrosc* 2016;327:73–94.
- [7] Huang X, et al. Ames-2016 line lists for 13 isotopologues of CO<sub>2</sub>: updates, consistency, and remaining issues. *J Quant Spectrosc Radiat Transfer* 2017;203:224–41.
- [8] Rey M, et al. TheoReTS – an information system for theoretical spectra based on variational predictions from molecular potential energy and dipole moment surfaces. *J Mol Spectrosc* 2016;327:138–58.
- [9] Kramida, A., et al., NIST atomic spectra database (version 5.8). 2020.
- [10] Maki AG, Wells JS. Wavenumber calibration tables from heterodyne frequency measurements (version 1.3). Editor. Gaithersburg, MD: N.I.o.S.a. Technology; 1998.
- [11] Babikov YL, et al. S&MPO - An information system for ozone spectroscopy on the WEB. *J Quant Spectrosc Radiat Transfer* 2014;145:169–96.
- [12] Hartmann J-M, Boulet C, Robert D. Collisional effects on molecular spectra: laboratory experiments and models, consequences for applications. Elsevier; 2008.
- [13] Ciurylo R. Shapes of pressure- and Doppler-broadened spectral lines in the core and near wings. *Phys Rev A* 1998;58(2):1029–39.
- [14] Galatry L. Théorie de la forme des raies spectrales dans le domaine hertzien. *J Phys* 1963;24(4):265–72.
- [15] Nelkin M, Ghatak A. Simple binary collision model for Van Hove's  $G_s(r,t)$ . *Phys Rev* 1964;135(1A):A4–9.
- [16] Ciurylo R. Shapes of pressure- and Doppler-broadened spectral lines in the core and near wings. *Phys Rev A* 1998;58(2):1029–39.
- [17] Ciurylo R, et al. Solving the line-shape problem with speed-dependent broadening and shifting and with Dicke narrowing. II. Application. *Phys Rev A* 2001:65.
- [18] Keilson J, Storer JE. On brownian motion, boltzmann's equation, and the fokker-planck equation. *Q Appl Math* 1952;10(3):243–53.
- [19] Tran H, et al. An isolated line-shape model based on the Keilson-Storer function for velocity changes. II. Molecular dynamics simulations and the  $Q(1)$  lines for pure H<sub>2</sub>. *J Chem Phys* 2009;131(15):154303.
- [20] Tran H, et al. Velocity effects on the shape of pure H<sub>2</sub>O isolated lines: complementary tests of the partially correlated speed-dependent Keilson-Storer model. *J Chem Phys* 2013;138(3):034302.
- [21] Gamache RR, Lynch R, Neshyba SP. New developments in the theory of pressure-broadening and pressure-shifting of spectral lines of H<sub>2</sub>O: the complex robert-bonamy formalism. *J Quant Spectrosc Radiat Transfer* 1998;59(3-5):319–35.
- [22] Ngo NH, et al. Influence of velocity effects on the shape of N<sub>2</sub> (and air) broadened H<sub>2</sub>O lines revisited with classical molecular dynamics simulations. *J Chem Phys* 2012;137(6):064302.
- [23] Wcislo P, et al. Velocity-changing collisions in pure H<sub>2</sub> and H<sub>2</sub>-Ar mixture. *J Chem Phys* 2014;141(7):074301.
- [24] Wcislo P, et al. The first comprehensive dataset of beyond-Voigt line-shape parameters from ab initio quantum scattering calculations for the HITRAN database: He-perturbed H<sub>2</sub> case study. *J Quant Spectrosc Radiat Transfer* 2021:260.
- [25] Stankiewicz K, et al. Accurate calculations of beyond-Voigt line-shape parameters from first principles for the He-perturbed HD rovibrational lines: A comprehensive dataset in the HITRAN DPL format. *J Quant Spectrosc Radiat Transfer* 2021:276.
- [26] Tran DD, et al. Prediction of high-order line-shape parameters for air-broadened O<sub>2</sub> lines using requantized classical molecular dynamics simulations and comparison with measurements. *J Quant Spectrosc Radiat Transfer* 2019;222-223:108–14.
- [27] Wcislo P, et al. The implementation of non-Voigt line profiles in the HITRAN database: H<sub>2</sub> case study. *J Quant Spectrosc Radiat Transfer* 2016;177:75–91.
- [28] Tennyson J, et al. Recommended isolated-line profile for representing high-resolution spectroscopic transitions (IUPAC Technical Report). *Pure Appl Chem* 2014;86(12):1931–43.
- [29] Ngo N, et al. An isolated line-shape model to go beyond the Voigt profile in spectroscopic databases and radiative transfer codes. *J Quant Spectrosc Radiat Transfer* 2013;129:89–100.
- [30] Lisak D, et al. Application of the Hartmann-Tran profile to analysis of H<sub>2</sub>O spectra. *J Quant Spectrosc Radiat Transfer* 2015;164:221–30.
- [31] Dicke R. The effect of collisions upon the Doppler width of spectral lines. *Phys Rev* 1953;89(2):472.
- [32] Varghese PL, Hanson RK. Collisional narrowing effects on spectral line shapes measured at high resolution. *Appl Opt* 1984;23(14):2376–85.
- [33] Fleisher AJ, et al. Twenty-five-fold reduction in measurement uncertainty for a molecular line intensity. *Phys Rev Lett* 2019;123(4):043001.

- [34] Konefał M, et al. Analytical-function correction to the Hartmann–Tran profile for more reliable representation of the Dicke-narrowed molecular spectra. *J Quant Spectrosc Radiat Transfer* 2020;242.
- [35] Fleurbaey H, et al. High accuracy spectroscopic parameters of the 1.27 $\mu\text{m}$  band of  $\text{O}_2$  measured with comb-referenced, cavity ring-down spectroscopy. *J Quant Spectrosc Radiat Transfer* 2021;270.
- [36] Wcisłó P, et al. The implementation of non-Voigt line profiles in the HITRAN database: H2 case study. *J Quant Spectrosc Radiat Transfer* 2016;177:75–91.
- [37] Adkins, E.M., Multi-spectrum Analysis Tool for Spectroscopy (MATS). 2020.
- [38] Tran H, Ngo N, Hartmann J-M. Efficient computation of some speed-dependent isolated line profiles. *J Quant Spectrosc Radiat Transfer* 2013;129:199–203.
- [39] Kochanov RV, et al. HITRAN Application Programming Interface (HAPI): a comprehensive approach to working with spectroscopic data. *J Quant Spectrosc Radiat Transfer* 2016;177:15–30.
- [40] Newville, M., et al., LMFIT: non-linear least-square minimization and curve-fitting for Python. 2014.
- [41] Kenneth W. Jucks, et al., ASCENDS Science Mission Definition Study. 2015.
- [42] Long D, et al. Frequency-agile, rapid scanning cavity ring-down spectroscopy (FARS-CRDS) measurements of the (30012) $\leftarrow$ (00001) near-infrared carbon dioxide band. *J Quant Spectrosc Radiat Transfer* 2015;161:35–40.
- [43] Long DA, et al. High-accuracy near-infrared carbon dioxide intensity measurements to support remote sensing. *Geophys Res Lett* 2020;47(5) e2019GL086344.
- [44] Reed ZD, et al. Molecular transition frequencies of  $\text{CO}_2$  near 1.6  $\mu\text{m}$  with kHz-level uncertainties. *J Quant Spectrosc Radiat Transfer* 2021 under review.
- [45] Adkins EM, Long DA, Hodges JT. Air-broadening in near-infrared carbon dioxide line shapes: quantifying contributions from  $\text{O}_2$ ,  $\text{N}_2$ , and Ar. *J Quant Spectrosc Radiat Transfer* 2021;270.
- [46] Lin H, et al. Cavity ring-down spectrometer for high-fidelity molecular absorption measurements. *J Quant Spectrosc Radiat Transfer* 2015;161:11–20.
- [47] Connor B, et al. Quantification of uncertainties in OCO-2 measurements of XCO<sub>2</sub>: simulations and linear error analysis. *Atmos Meas Tech* 2016;9(10):5227–38.
- [48] Le T, et al. Isolated line shape of methane with various collision partners. *J Quant Spectrosc Radiat Transfer* 2016;185(Supplement C):27–36.
- [49] Cygan A, Lisak D. Multi-spectrum fitting software for advanced spectral line shapes analysis. *Journal of Physics: Conference Series*. IOP Publishing; 2017.
- [50] Stolarczyk N, et al. Evaluation of different parameterizations of temperature dependences of the line-shape parameters based on ab initio calculations: case study for the HITRAN database. *J Quant Spectrosc Radiat Transfer* 2020;240.
- [51] Larcher, G., et al., Spectral shape parameters of pure  $\text{CO}_2$  transitions near 1.6 $\mu\text{m}$  by tunable diode laser spectroscopy. 2015. 164: p. 82–88.
- [52] Ngo N, et al. Spectral shapes of rovibrational lines of CO broadened by He, Ar, Kr and SF 6: a test case of the Hartmann-Tran profile. *J Quant Spectrosc Radiat Transfer* 2017.
- [53] Pedregosa Fabian, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825–30.
- [54] Ghysels M, et al. A variable-temperature cavity ring-down spectrometer with application to line shape analysis of  $\text{CO}_2$  spectra in the 1600nm region. *Appl Phys B* 2017;123(4).
- [55] Gordon, I.E., et al., The HITRAN2016 molecular spectroscopic database. 2017. 203: p. 3–69.
- [56] Fleurbaey H, et al. Cavity ring-down spectroscopy of  $\text{CO}_2$  near  $\lambda = 2.06 \mu\text{m}$ : accurate transition intensities for the orbiting carbon observatory-2 (OCO-2) "strong band". *J Quant Spectrosc Radiat Transfer* 2020;252:107104.
- [57] Chris Benner D, Malathy Devi CPRV, Smith Mary Ann H, Atkins David. A multi-spectrum nonlinear least squares fitting technique. *J Quant Spectrosc Radiat Transfer* 1995;53(6):705–21.
- [58] Drouin BJ, et al. Multispectrum analysis of the oxygen A-band. *J Quant Spectrosc Radiat Transf* 2017;186:118–38.
- [59] Cygan A, Lisak D. Multi-spectrum fitting software for advanced spectral line shapes analysis. XXIII International Conference on Spectral Line Shapes; 2017. *Journal of Physics: Conference Series*.
- [60] Loos, J., M. Birk, and G. Wagner, New multispectrum fitting software used at DLR for analysis of laboratory Fourier-Transform molecular spectra. 2014.
- [61] Delahaye T, et al. Precise methane absorption measurements in the 1.64 $\mu\text{m}$  spectral region for the MERLIN mission. *J Geophys Res: Atmospheres* 2016 n/a-n/a.