# An Evaluation of Spectroscopic Databases: Content, Structure, Toolsets, Usability, and Documentation

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Joseph William Braker
May 2024

Copyright © 2024 by Joseph William Braker
All rights reserved.

# ACKNOWLEDGEMENTS

For my wife, Sarah Braker, and our daughters, Allison and Gillian. I am eternally grateful for your patience, motivation, and for the sacrifices you have made to allow me to pursue my dreams.

Thank you to Dr. David Plemmons. Without your tenured guidance and the experience gained working with you, my journey would have ended much sooner and incompleted.

Thank you to (perhaps already Dr.) Evan Combes. Your belief and confidence in me is something that drove me forward in times of struggle. You are, without question, among the greatest of those who teach and I hope you follow your dreams and share that gift with the world.

# ABSTRACT

Spectroscopic databases provide a critical service in the greater field of spectroscopy. Many subdisciplines of spectroscopists use the data contained within, whether calculating theoretical models, evaluating the atmosphere of a planet in the furthest reaches of the galaxy, or monitoring the emission levels of a chemical process. The intention of this research is to seek a deeper level of understanding of spectroscopy and the resources available while providing a comparison of the databases, the content within, and their peripheral tools and features. An experimental effort was derived to utilize existing experimental data from multiple sources and determine which data source provides the most precise representation for a given condition.For example, for pressures "P" and temperatures "T":

1. Data source 1 provides the most precise representation between $P_1T_1$ [Condition 1 of Pressure 1 and Temperature 1] and $P_2T_2$ [Condition 2 of Pressure 2 and Temperature 2]
2. Data source 2 is best when pressures and temperatures are higher than $P_2$ and $T_2$ [Condition 2 of Pressure 2 and Temperature 2].
3. Data source 1 provides the lowest total integrated error over the available data and is the best choice if only one set of line lists can be used for a wide range of conditions

These data sources come in the form of spectroscopic databases that contain information on the spectroscopic parameters of atoms and molecules, including various isotopologues of the same molecule. While success was found in learning spectroscopy, the results on comparison of the databases was less than satisfying. In many cases, the same partition function and broadening parameters are utilized in multiple databases leading to identical results at baseline conditions. Additionally, the tool sets available to interpret their data did not allow for use before modification due to the inconsistent formatting between datasets. Comparative data is reported along with suggested next steps.

# TABLE OF CONTENTS

Chapter One: Introductions and General Information..... 1
Chapter Two: Literature Review..... 3
A Brief Introduction to Spectroscopy..... 3
Quantized Energy and Emission Spectrum..... 3
Spectral Line Broadening..... 6
Chapter Three: Databases, Methods, and Tool Sets..... 9
The Representation and Modeling of Spectra in HITRAN..... 9
The Representation and Modeling of Spectra in ExoMol..... 11
Other Requisite Tools..... 12
Other Considered but Unused Toolsets..... 13
Chapter Four: Results and Discussions..... 15
Validating Function of the ExoCross Toolset..... 15
Validating Function of the HITRAN HAPI Toolset..... 17
A Comment on Isotopologue Choice..... 17
Toolset Validation Conclusions..... 17
Variation in Temperature..... 19
Variation in Pressure..... 26
Chapter Five: Conclusions and Recommendations..... 30
Works Cited..... 32
Vita..... 34

# LIST OF FIGURES

Figure 1: HITRAN and ExoMol spectrum showing many emission lines of water..... 5
Figure 2: HITRAN and ExoMol showing broadness of the spectral features of water..... 7
Figure 3: Doppler Broadening Comparison of ExoMol Tools and ExoCross..... 16
Figure 4: Doppler Broadening Comparison of HAPI, ExoMol Tools and ExoCross..... 18
Figure 5: Zoom of Figure 4 on an absorption line at 3734.93 cm-1..... 18
Figure 6: The entire molecular spectrum from 3700 cm-1 to 3800 cm-1 for 1H216O.....20
Figure 7: Percent difference from 3700 cm-1 to 3800 cm-1 for 1H216O..... 20
Figure 8: Percent difference from 3700 cm-1 to 3800 cm-1 for 1H216O as a histogram..... 20
Figure 9: Doppler Broadening of ExoCross for 1H216O at a range of temperatures.....22
Figure 10: Doppler Broadening of ExoCross for 1H216O at a range of temperatures.....22
Figure 11: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 50K.....22
Figure 12: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 100K.....23
Figure 13: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 300K.....23
Figure 14: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 400K.....23
Figure 15: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 500K.....24
Figure 16: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 750K.....24
Figure 17: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 1000K.....24
Figure 18: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 2000K.....25
Figure 19: Doppler Broadening Comparison of HAPI and ExoCross for 1H216O at 3000K.....25
Figure 20: A zoom of Figure 14 to show the weak spectral line..... 25
Figure 21: Pressure Broadening of HITRAN for 1H216O at a range of pressures..... 29
Figure 22: Pressure Broadening of ExoCross for 1H216O at a range of pressures.....29

# Chapter One: Introductions and General Information

In nearly every industry, the examination of optical spectra has been proven to provide new and previously unavailable data. From the evaluation of material characteristics to the monitoring of process steps, the wavelength of light emitted or absorbed by a material provides key indicators for decision making. The science of spectroscopy continues to be critical in the advancement of industries and has become a tool utilized in countless applications. The use of spectroscopic data requires the ability to interpret the information that it provides into meaningful units of measure. This is accomplished with an understanding of which molecules emit and absorb particular wavelengths of light and how the associated spectrum changes as a function of environmental conditions, such as pressure and temperature. Tremendous amounts of measured and calculated data at known conditions has been taken and stored in various data sources that are available for public use.

These data sources come in the form of spectroscopic databases that contain a litany of information on the spectroscopic parameters of atoms and molecules. Data is broken down further by the isotopologues of a molecule, or the configurations of a molecule where the number of neutrons can vary in the atoms within the molecule. For example, this work focuses on the $^{1}H_{2}^{16}O$ isotopologue of $H_2O$ but other isotopologues exist for water including $^{1}H_{2}^{18}O$, $^{1}H^{2}H^{16}O$, and $^{2}H_{2}^{16}O$. All of these represent a molecule of water composed of atoms in different configurations.

Multiple databases exist and each utilizes a different methodology for capturing, calculating, and displaying the relevant data. For example, the HITRAN (Rothman et al.) database uses a *.par file as the primary means of transferring information. Meanwhile, the ExoMol (Tennyson et al.) database uses several files of *.def, *.cp, *.broad, *.trans, *.states, and *.par, where each contains different parameters relevant to the isotopologue and molecule being evaluated.

While the data sources all provide the basic parameters of an absorption line, the ExoMol database contains substantially more information. For example, the majority of the data within HITRAN has its foundation at Earth-like conditions at atmospheric pressure (1 atm) and at a temperature of 296K.

The pressure broadening parameters are often limited to air broadening and self broadening in the standard available files but additional information can be exported from the database through the use of custom file formatting. A subsection of the HITRAN database called HITEMP contains data that was generated while evaluating spectra at temperatures and pressures that are much higher than the standard.

In ExoMol, the data is primarily used to analyze and model exoplanet atmospheres. This allows the range of pressures and temperatures to be available when generating the database to be very diverse and will depend on the source of the measurement. As the atmospheres of exoplanets can have a variety of compositions at a variety of temperatures and pressures, the pressure broadening values for various molecules can be very diverse. As such, substantially more information is often available for a given molecule and the given molecule's specific isotopologue. However, the level of complexity in the ExoMol database makes it much more challenging to utilize and the tool sets provided are either much more limited or have tremendous learning curves.

# Chapter Two: Literature Review

## A Brief Introduction to Spectroscopy

In this section, a brief introduction to spectroscopy will be provided. This topic has been covered in many works across countless publications. In the words of Dr. Ramamurti Shankar (Shankar 1980) "one would think the public needs one more as much as it does, say, the latest version of the Table of Integrals." However, one more is provided here.

## Quantized Energy and Emission Spectrum

In the study of Quantum Mechanics, we learn early that one of the most fundamental principles is that atomic and molecular energy is quantized. We accept that any atom, molecule, or electron contained within may only exist in certain states, referred to as "quantum states". These states represent the amount of energy stored within the molecule and, again, it is not a continuous range of options. (Griffiths and Schroeter). It's a common misconception that atoms are like small solar systems with electrons orbiting the nucleus of protons and neutrons. Electrons in an atom or molecule are more accurately represented as waves, specifically, standing waves that obey an equation called Schödinger's Equation. When a solution to this equation is found, the various "wave functions" for a given particle are determined. Each of these wave functions describes a quantized energy state.

For the field of spectroscopy, this very valuable piece of information must be true for the optical spectrum to possess any meaning. When electrons change energy levels, our old friend from Freshman Physics, Conservation of Energy, must still be abided. To move between states, the difference in the energy the electron currently possesses and the energy the electron needs to possess to occupy the destination state must be addressed. Should the electron move to a higher state, it must absorb energy and to move to a lower state, it must release energy. This can be done through the release or absorption of a photon of the specific energy that is required to make the transition.

The internal energy of a molecule is the combination of three different categories of energy, including Rotational, Vibrational, and Electronic energies. Oftentimes, there are many potential transitions for a molecule to make from its starting Rotation, Vibrational, and Electronic states. Transitions between energy states are rarely due to changes in independent categories of energy. For example, most transitions in the UV or Visible range of the optical spectrum are a combination of all three, called "rovibronic". Transitions in the IR spectrum, called rovibrational, are those with changes in Rotational and Vibrational energies.

With just this information, it's possible to make the hypothesis that a group of molecules making transitions between energy levels are unlikely to make precisely the same transition. There will be a range of both starting states and ending states. To say this more directly, the change in energy for the various molecules or particles is not guaranteed to be the same (Hanson et al.). This leads us to the first equation relating to the optical spectrum and spectroscopy. The energy of the released photon is the total energy from all categories and the photon wavelength is inversely proportional to the wavelength as shown in Equation 1.

$$\Delta E_{Rotational} + \Delta E_{Vibrational} + \Delta E_{Electronic} = \Delta E_{Total} = \frac{hc}{\lambda} \quad \text{(Eq. 1)}$$

Electronic energy transitions are large, vibrational transitions are small, and rotational transitions are very small. Because we correctly hypothesized that not all transitions from molecules will be of equal energy, we can and likely will see more than one spectral line, or spectral feature, in the spectrum of the molecule's transition.

To reinforce a critical point, allow me to repeat that the internal energy of a molecule or particle can only have specific values. Close values in energy leads to close values in transition energies. This "Total Internal Energy" is the sum of energy which can allow for multiple "Total Internal Energy" levels to be relatively close in value. This in turn leads to spectral features that are close together and appear as a series of emission or absorption lines. This is shown in both ExoMol (Polyansky) and HITRAN data (Gordon et al.) in Figure 1.

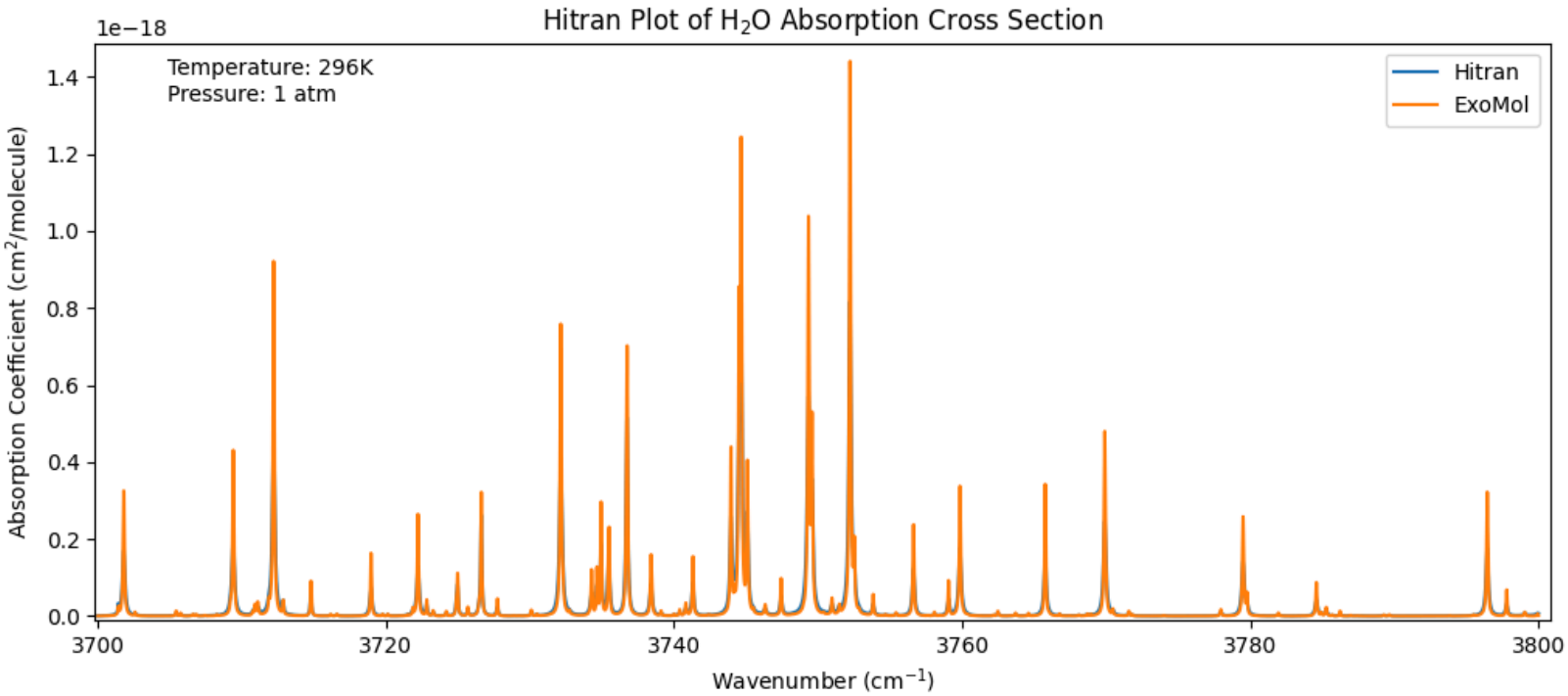



Figure 1: HITRAN and ExoMol spectrum showing many emission lines of water

**Spectral Line Broadening**

In addition to the presence of a series of spectral lines, each of the spectral lines does not correspond to a single wavelength. While the explanations provided in the prior section would imply that only a single differential energy value is allowed for a particular transition between certain energy levels and that this corresponds to a unique and singular wavelength, this is a simplification. In practice, the spectral line for the transition will have some range of wavelengths that are present, as is seen in Figure 2. The width of the range of wavelengths is known as the spectral width and is typically measured in half width at half maximum, abbreviated HWHM. There are many contributors to the width of the spectral line including local and non-local effects. These include (Hanson et al.):

- Natural Broadening - Most people familiar with Heisenberg's Uncertainty Principle are familiar with it in the form of position and momentum, as shown in the left portion of Equation 2. However, the same principle directly relates to the uncertainty in energy and time, as shown in the right half of Equation 2. This is the underlying source of a broadening effect. The amount of time a particle will stay in an excited energy state is called the lifetime. The lifetime of an electron currently in an excited energy state has an uncertainty. The uncertainty in lifetime relates to an uncertainty in the energy released in the photon and, therefore, an uncertainty in its wavelength. This is the source of Natural Broadening

$$\Delta x \Delta p \geq \frac{h}{4\pi} \leq \Delta E \Delta t \qquad \text{(Eq. 2)}$$

- Pressure Broadening - Pressure broadening comes in two main forms, collisional broadening and quasistatic pressure broadening.
    - Collisional Broadening: When particles or molecules collide, this disrupts lifetimes of the particles in their energy state. As discussed in the Natural Broadening, the uncertainty in these lifetimes leads to uncertainty in energy and the collisions increase that uncertainty. This broadens the line further.

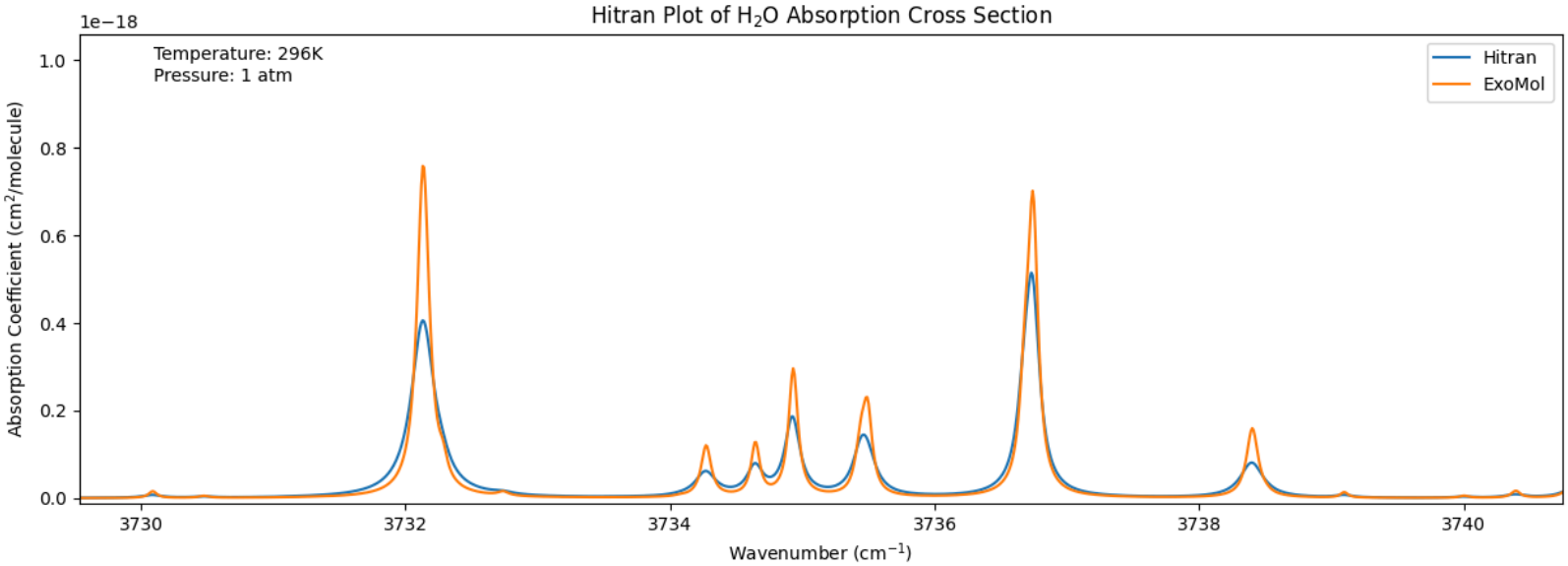



Figure 2: HITRAN and ExoMol showing broadness of the spectral features of water

    - Quasistatic Broadening: When other particles are nearby the molecule being evaluated, their presence disrupts the natural energy levels of a molecule.
- Local Doppler Broadening - The molecules and particles contributing to the spectral lines are not stationary but rather are moving around with various velocities in all directions due to the thermal motion of the molecules. This causes the emitted photon to experience a Doppler shift relative to the measurement instrument. The combined or bulk motion from all the particles will result in a Gaussian shaped spectra.
- Non-Local Doppler Broadening - If the different areas of a light source are moving in different directions, non-local Doppler shifts can create broadening effects as well. For example, a large quickly rotating star would have a different Doppler shift on the edge that's moving toward the measurement device when compared to the other edge that's moving away. It's worth noting that if the spectrum is only sampled on one of these edges, it will appear as a shift rather than broadening.

# Chapter Three: Databases, Methods, and Tool Sets

**The Representation and Modeling of Spectra in HITRAN**

Modeling the spectra of a molecule is done by evaluating the features of the spectrum one at a time. Although the emission or absorption spectrum of a molecule may include hundreds or thousands of spectral lines, they must be evaluated individually and combined to create the total spectrum. Below, the equation for the spectral line intensity used by HITRAN is presented. The equation will be presented as Equation 3 in full before being evaluated in more details:

$$S_{ij} = I_a \frac{A_{ij}}{8\pi c \nu_{ij}^2} \frac{g' e^{-c_2 E''/T} (1 - e^{-c_2 \nu_{ij}/T})}{Q(T)} \quad \text{(Eq. 3)}$$

- $\nu_{ij}$ : the wavenumber of the spectral line transition from $E_i$ to $E_j$.
- $I_a$ : the natural terrestrial isotopic abundances.
- $c_2$ : the Planck constant times the speed of light divided by Boltzman's constant or hc/k.
- E` : the upper state energy.
- E``: the lower state energy.
- g` : the upper state statistical weight or the probability of occupying the upper state.
- T : Temperature (Kelvin).
- $A_{ij}$ : the Einstein-A coefficient of a particular transition shown in Equation 4.

$$A_{ij} = \frac{64\pi^4}{3h} \nu_{ij}^3 \frac{g''}{g'} \Re_{ij} \times 10^{-36} \quad \text{(Eq 4.)}$$

- $\Re_{ij}$ : weighted transition moment squared.
- Q(T) : the Total Internal Partition function shown in Equation 5.

$$Q(T) = \sum_k g_k e^{(-c_2 E_k)/T} \quad \text{(Eq. 5)}$$

By knowing the energy levels of the starting and ending states, we are able to calculate the spectrum for a molecule for a particular transition and generate the Spectral Line Intensity Equations 3, 4, and 5. (Gordon et al.)

Note there are no broadening parameters or pressure shift parameters in these equations. Pressure broadening HWHM is typically denoted using $\gamma_{\text{broadener}}$ and pressure line shift is given by $\delta_{\text{broadener}}$, where "broadener" may be air, He, $H_2$, "self" for self-broadening, or any other broadener. To incorporate pressure and temperature into our solution for the spectrum, the broadened spectral line width must be

calculated separately. For a temperature and pressure change to the linewidth, a Voigt profile acts as the standard modeling shape. This is a combination of a Gaussian and Lorenztian shape. The width of the Lorentian is defined by HITRAN as shown in Equation 6:

$$\gamma(p,\ T)\ = \left(\frac{T_{ref}}{T}\right)^{n_{air}}\left(\gamma_{air}\left(p_{ref}\ ,T_{ref}\right)\left(p\ -\ p_{self}\right)\ +\gamma_{self}\left(p_{ref}\ ,T_{ref}\right)p_{self}\right) \quad \text{(Eq. 6)}$$

Equation 7 shows how this can be be written more generally as:

$$\gamma(p,\ T)\ = \left(\frac{T_{ref}}{T}\right)^{n_{total}} p\left(\sum_{i}^{all\ broadeners}\left(\gamma_{i}\left(p_{ref}\ ,T_{ref}\right)\left(\frac{p_{i}}{p}\right)\right)\right) \quad \text{(Eq. 7)}$$

where all $p_i$ values sum to p. This utilized the same broadening as a function of the partial pressure but allows more broadeners to be utilized than simply self and one additional. It's critical to note here that the pressure broadened HWHM values are determined at a reference temperature and pressure so may be subject to imprecision as conditions vary. Additionally, Local Doppler Broadening HWHM can be calculated using Equation 8:

$$\alpha_{D}(T)\ = \frac{\nu_{ij}}{c}\sqrt{\frac{2N_{A}\ kT\ ln(2)}{M}} \quad \text{(Eq. 8)}$$

where M is molar mass and $N_A$ is Avagadro's Number of 6.022 x $10^{23}$ $mol^{-1}$. Equation 9 shows how to scale $S_{ij}$, the Spectral Line Intensity with temperature and the heavy dependence on the Total Internal Partition Function should be noted.

$$S_{ij} = S_{ij}\left(T_{ref}\right)\cdot\frac{Q(T_{ref})}{Q(T)}\cdot\frac{e^{(-c_2E'')/T}}{e^{(-c_2E'')/T_{ref}}}\cdot\frac{1-e^{(-c_2\nu_{ij})/T}}{1-e^{(-c_2\nu_{ij})/T_{ref}}} \quad \text{(Eq. 9)}$$

When evaluating these functions as a whole, there is a standout point that must be addressed. The Total Internal Partition function, Q(T), and the pressure broadening HWHM values, $\gamma_{broadener}$ , are determined at a reference temperature. While the Internal Partition Function only has a temperature dependency, the broadeners have a dependency on both temperature and pressure. When viewing data in HITRAN, these values are consistently determined at 1 atm of pressure and 296K.

The HITRAN database provides an Application Programming Interface (API) called the "HITRAN Application Programming Interface" (HAPI). This API includes a Python programming library to make the HITRANonline database easily accessible. From getting data directly off the online database to performing the calculations required to generate the spectral lineshape, the library contains functions to perform them all (Kochanov). Investigation into the functions provided in the HAPI Python Library seem to show that the functions and methods are performing the calculations as shown in the equations above. In addition, the library is pulling appropriate Total Internal Partition values (Gamache et al.) from the database.

**The Representation and Modeling of Spectra in ExoMol**

The ExoMol database provides a monumental amount of information with tremendous granularity. However, it proved substantially more difficult to work with than HITRAN due to the incredible amount of data and the complexity of the existing tool sets and documentation. ExoMol has two general types of tools available for evaluating the available data. First, the website has online service tools that will calculate the absorption cross section, but only at zero pressure for Doppler-broadened lines only (Christian et al.). This provides some validation but does not provide useful data for any pressure broadening parameters. Secondarily, an additional tool called ExoCross (Yurchenko) is available. ExoCross is a source code that generates both spectra and thermodynamic properties. It is written in Fortran which allows tremendous speed in calculation, but requires knowing the Fortran language to understand the details of how it works and the steps that are in the calculation. The ExoCross Documentation shows that it does similar calculations to those listed in HITRAN's documentation; however, this was not able to be validated from the source code.

An additional tool exists to ease the interface with ExoCross. PyExoCross is effectively a binding that interfaces between Python and ExoCross. While ExoCross requires command-line calls to a specific file structure and executable arguments that follow, PyExoCross allows the use of Python functions to build the command line call and file within the script, call the compiled Fortran executable

with the specific arguments, then retrieve and interpret the results in a Python compatible way. Unfortunately, the tool appears ill prepared to handle variation and inconsistency in the ExoMol files because it reads the ExoMol definitions file line by line. As such, it expects a particular return type to be on a specific line that is a predefined number of lines below the last line read. This allows for failures in the PyExoCross program because not all of the files are formatted the same way. As an example, the BT2 dataset for $^{1}H_{2}^{16}O$ (Furtenbacher et al.) has a definitions file that defines two pressure broadeners and then defines two distinct file names for those broadeners. However, the Pokazatel (Polyansky et al.) dataset for $^{1}H_{2}^{16}O$ describes four pressure broadeners but only defines two distinct file names to contain them. Likewise, the XABC (Wong et al.) (Qu et al.) dataset for $^{14}N^{16}O$ has a definitions file that describes no pressure broadeners and therefore only doppler broadening can be implemented using ExoCross and the ExoMol Datasets.

**Other Requisite Tools**

The primary portion of programming was done in Python using Python 3.9.6. Anaconda was used as the primary package manager and Pip was used as an installer in support of certain packages. JetBrains Pycharm Professional was used as the Integrated Development Environment. The following Python Libraries were used:

- NumPy is a scientific computing library with a backend written in Fortran. It is a very popular and powerful library showing case studies in the detection of gravitational waves at Laser Interferometer Gravitational-Wave Observatory, known as LIGO, and the first image of a black hole taken at CalTech. Numpy provides the mathematical toolbox that was used to implement calculations from the database.
- MatPlotLib is a toolbox for creating graphs, plots, and other visualizations. It provides a library for quality plots and interactive figures. In this work, it was used to generate plots for data evaluations and the plots in this document.

- The HITRAN database provides an Application Programming Interface (API) called the "HITRAN Application Programming Interface" (HAPI). This API includes a set of Python functions and routines to make the HITRANonline database accessible. For evaluation of the HITRAN database, HAPI was used as the primary tool as there was little need for more thorough or additional functions.
- PyExoCross is a Python library that provides an interface to the ExoCross executable in Python and was described above.

Because processing data using the ExoMol database required use of the ExoCross utility, a Fortran compiler was also required. Jetbrains CLion IDE is designed for use in C and C++ but a plugin is available that supports Fortran source code. This combination was used to compile the Fortran source code for ExoCross into an executable for use with PyExoCross. The hardware used was a 2023 Apple Macbook Pro with Apple M2 Max CPU, 32GB of Memory, and operating on Sonoma 14.2.1 operating system.

**Other Considered but Unused Toolsets**

Another work exists that has created tools to process data from multiple databases. A very thorough opacity calculator known as HELIOS-K (Grimm et al.) provides toolsets for ExoMol, HITRAN, HITEMP, NIST, Kurucz, and VALD3. While this appears to have provided a complete toolset to use, the main purpose of this effort was to learn the fundamentals of spectroscopy. Putting the parameters for evaluation into the metaphorical "black box" and retrieving results seemed to be outside the intention of this effort. HELIOS-K was not used to generate the data in this paper and neither were subsections of the source code. At the start of the Literature Review for this project, the HELIOS-K tools for ExoMol were evaluated.

The source code for HELIOS-K included various tools for using the ExoMol database and while the method in which they work together is unknown, the individual tools were reviewed for potential use in this work. There were two main subprograms that were reviewed. The first decomposes the various

file types from the ExoMol database and generate a *.param file for use in the primary opacity curve generating function. This appears to run a calculation similar to that of ExoCross, however, it must be noted that Python, C++, and Ada are languages in which preexisting experience and knowledge of the syntax make evaluation of source codes quite straightforward for the author. Unfortunately, as mentioned above, ExoCross is written in Fortran and this is a language where there is no experience or familiarity with the syntax. Learning the Fortran language is believed to be outside the scope of this work so the previously mentioned comparison of the calculations in the first portion of HELIOS-K and ExoCross is not being expressed as a factual conclusion.

The other tool from the HELIOS-K source code uses Python libraries called BeautifulSoup and Selenium. These libraries provide access to the internet and were used to access the ExoMol online service tool sets through the use of the Python and captured the data reported in the tool, saved it to a file, and imported it into Python. This tool was not used as it was also restricted to the Doppler only broadening and the data it provided was simpler to that obtained by accessing the ExoMol website directly.

# Chapter Four: Results and Discussions

**Validating Function of the ExoCross Toolset**

The predominant portion of the effort with ExoMol was getting an understanding of the organization of the data available and implementing the toolsets in a way that enables them to function. This required compiling the Fortran source code for ExoCross from within the CLion IDE. The GNU compiler was not implemented in a way that allowed Fortran to compile using the "make" commands that can usually be used. Fortran compatibility came from a third party plugin but allowed compilations once configured appropriately. Once compiled, some basic commands were issued to the executable to ensure proper responses were received and when validated, the focus was switched to PyExoCross.

PyExoCross was more straightforward to install and setup. The Pip package manager handled the installation and dependencies through the Python Console of a virtual environment established by PyCharm when the project was initiated. Once PyExoCross was linked to the ExoCross executable, basic commands were issued to ensure functionality. When functionality was validated, the initial step was to verify spectra were being generated appropriately by comparing the results of PyExoCross to the ExoMol's website online service tools. This online tool only allows for zero pressure Doppler-only broadening. While the temperatures could be set to match, ExoCross would not accept a uniquely zero pressure setting so it was approximated to zero using $10^{-7}$ and a 50/50 split of broadening between He and $H_2$. Figure 3 shows a graphical comparison of the results of the Pokazatel dataset for $^{1}H_2{}^{16}O$ . Further validation was done using different regions of the curve spectrum. Comparative results did not vary substantially at any point. The results are very close and were believed to be within reasonable precision to proceed given the approximation in pressure and unknown pressure broadeners in the online tool.

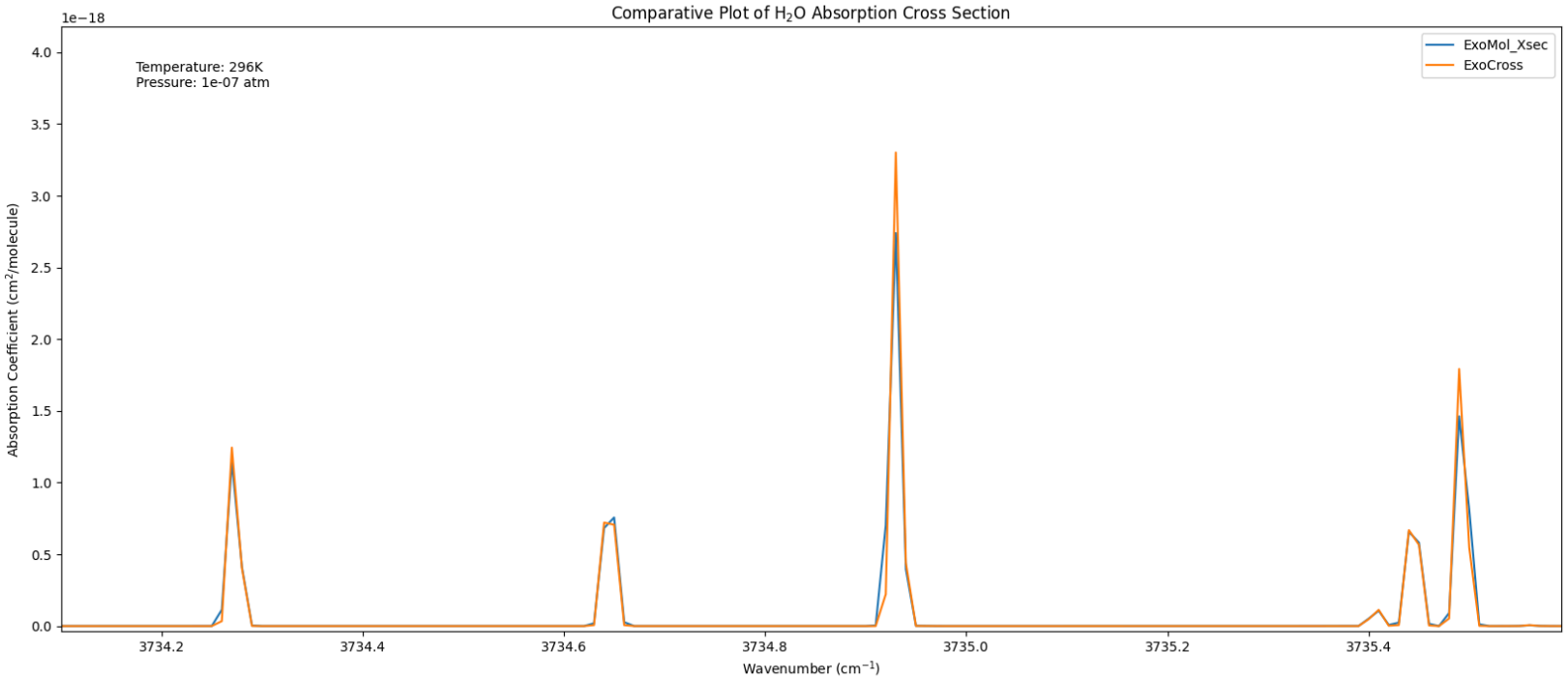



Figure 3: Doppler Broadening Comparison of ExoMol Tools and ExoCross.

**Validating Function of the HITRAN HAPI Toolset**

With the belief that ExoCross was generating valid spectra from the datasets, the HITRAN API, HAPI, was installed with the same ease as other Python libraries using Pip and the virtual environment in PyCharm. With no compilation steps, the HAPI library was able to be called directly from Python. The same process used to validate the ExoMol implementation of ExoCross was also used to validate the HITRAN implementation of HAPI. Both were provided the same pressure of $10^{-7}$ and a temperature of 276K. Figures 4 and 5 show the results from the comparison. The plots of the HAPI implementation of HITRAN and the ExoCross implementation of ExoMol are remarkably similar and vary only by a small percentage of their values while the ExoMol web tool based plot deviated from them substantially.

**A Comment on Isotopologue Choice**

While the original intent was to evaluate the $^{14}N^{16}O$ isotopologue of NO, the ExoMol database provides no broadening parameters for the molecule. The evaluation of this molecule would only be possible for Doppler Broadening which would not exercise the pressure broadening functionality of the tool sets. As such, evaluation continued of the $^{1}H_2{}^{16}O$ isotopologue of water vapor.

**Toolset Validation Conclusions**

At the peak shown in Figure 5, HITRAN shows a peak value of 3.25728976e-18 located at 3734.93$cm^{-1}$ while at the same location, ExoCross shows a peak value of 3.30192634e-18. This represents a difference of 1.35% in the peak values. With this minimal error in the absorption peak values, ExoCross and HITRAN HAPI are believed to be within acceptable precision to one another for Doppler Broadening limited scenarios. The online web service tool from ExoMol appears to be inconsistent with the HITRAN HAPI and ExoCross reconstruction of the molecular spectrum. The precision in their mutual reconstruction was evaluated at multiple locations on the spectrum and showed

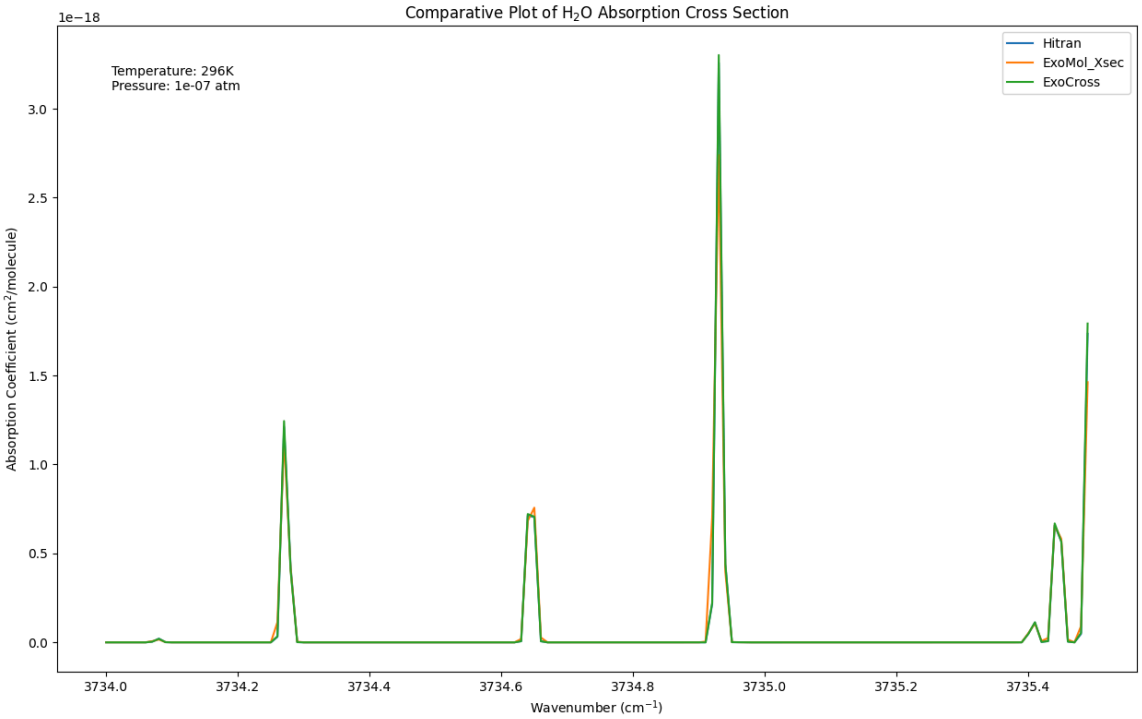



Figure 4: Doppler Broadening Comparison of HAPI, ExoMol Tools and ExoCross.

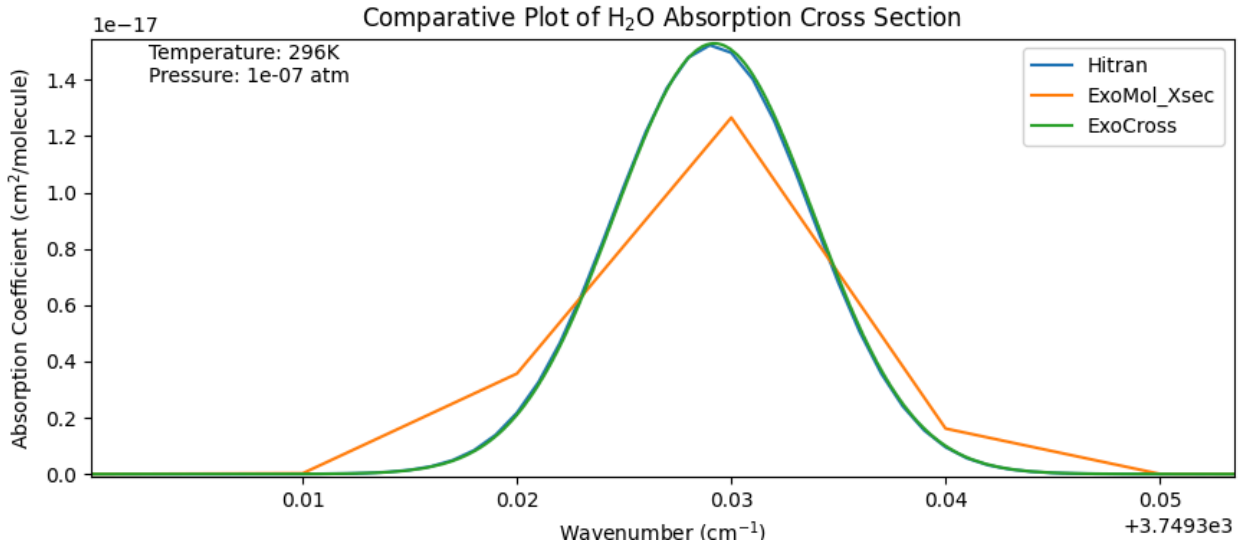



Figure 5: Zoom of Figure 4 on an absorption line at 3734.93 $cm^{-1}$

similar precision across the curve. Figures 6, 7, and 8 show a plot of the entire molecular spectrum from 3700 $cm^{-1}$ to 3800 $cm^{-1}$, a plot of the percent difference between the value of HAPI implementation of HITRAN and the ExoCross implementation of ExoMol as a function of wavenumber, and the same percent difference plotted as a histogram, respectively. The error plots were generated by first building the entire molecular spectrum from 3700 $cm^{-1}$ to 3800 $cm^{-1}$. To avoid large swings in error due to the noise floor, any datapoint containing a value greater than $10^{-20}$ generated from either database's reconstruction was evaluated. The percent difference shown in the plot was calculated using Equation 9.

$$\Delta S = \left| \frac{S_{ExoMol} - S_{HITRAN}}{S_{ExoMol}} \right| \quad \text{(Eq. 9)}$$

From these plots, we see a reasonably small error over the entire width of the generated spectra and peaks in error are related to low absorption coefficients. The high absorption lines show substantially less error. From this, a conclusion was drawn that for this situation of Doppler Broadening only, the molecular spectra generated from the ExoCross implementation of Exomol and the HAPI implementation of HITRAN can be considered representative of one another for this range of the $^{1}H_{2}{}^{16}O$ isotopologue of water vapor.

**Variation in Temperature**

From Equation 3, the absorption coefficient has a clear dependency on the temperature. Additionally, Equation 8 shows the temperature dependence of the local doppler broadening effects. As such, the next step in our comparison of the databases is to reevaluate the changes in both spectral width of the spectral lines along with their change in strength. With the performance of the two spectral generation methods validated for Doppler only broadening at near zero pressure and Earth-like temperatures of 296K, the temperature was then varied and the molecular spectra was calculated at various intervals in a large range from very low to very high temperatures. This increase in temperature is expected to induce an incremental increase in the spectral width of individual features; however the strength of the absorption lines is expected to decrease as the temperature increases.

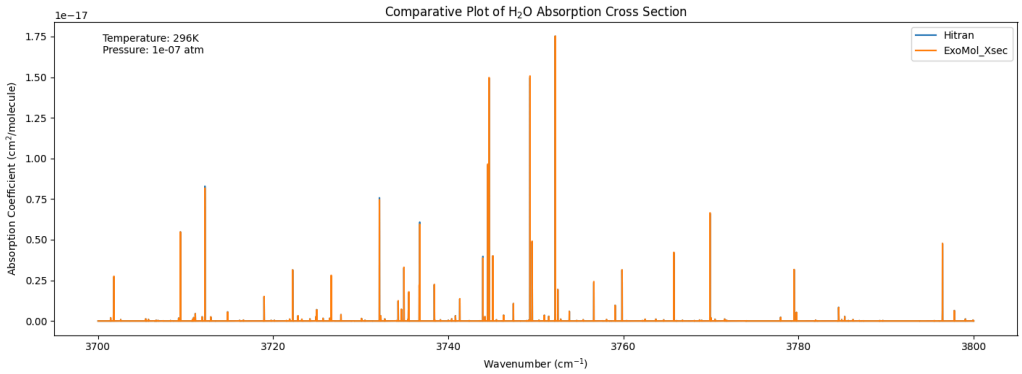



Figure 6: The entire molecular spectrum from 3700 cm$^{-1}$ to 3800 cm$^{-1}$ for $^{1}H_2{}^{16}O$

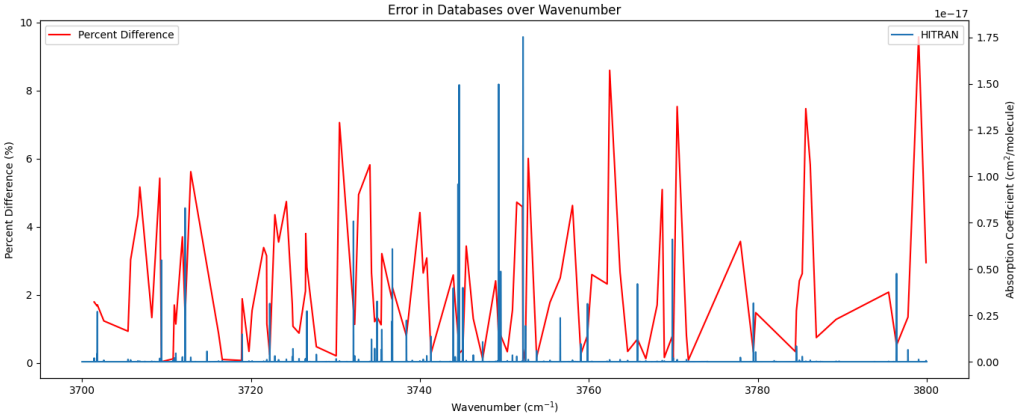



Figure 7: Percent difference from 3700 cm$^{-1}$ to 3800 cm$^{-1}$ for $^{1}H_2{}^{16}O$

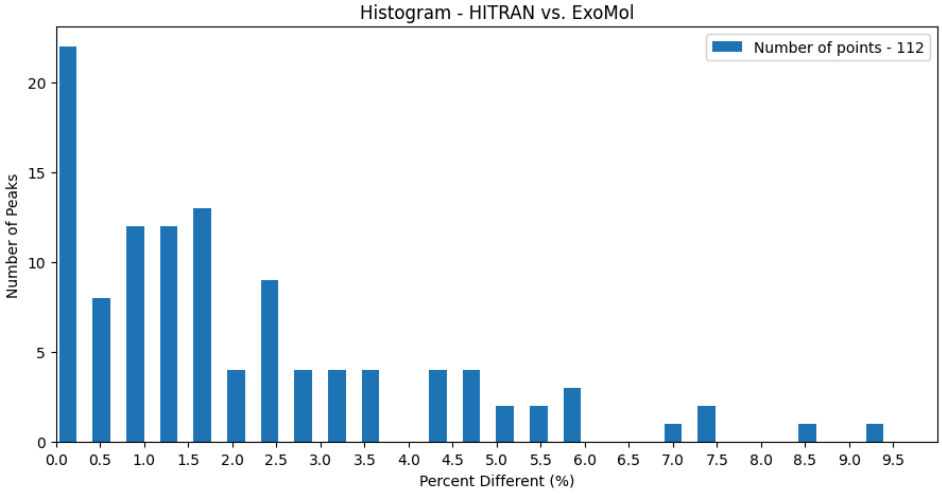



Figure 8: Percent difference from 3700 cm$^{-1}$ to 3800 cm$^{-1}$ for $^{1}H_2{}^{16}O$ as a histogram.

This is explained by the Ideal Gas Law. The Ideal Gas Law defines Equation 10.

$$PV = nk_bN_AT \quad \text{(Eq. 10)}$$

It can be gathered that the pressure (P) of an ideal gas within a defined volume (V) will be the same as the amount of the ideal gas (n) and its temperature (T) when multiplied by the Ideal Gas Constant (R). The Ideal Gas Constant and Temperature can be rewritten to represent the kinetic energy of an average molecule within the gas as Equation 11.

$$\frac{1}{2}mv^2 = \frac{3}{2}k_bT \quad \text{(Eq. 11)}$$

With this direct relation between temperature and the average velocity of any given molecule of the gas (since mass and Boltzman's Constant aren't changing), it is possible to conclude that a higher gas temperature will result in a higher molecular velocity (Wolfson and Pasachoff). With a higher molecular velocity, the center of the spectral line will shift for a given particle and, in a large statistical sample, this will result in a larger variation of the upper and lower bounds of the Doppler shift. With larger bounds of a Doppler shift, the larger sample size won't see a peak shift due to the increase in the variation of doppler shifts. Rather this will result in a more significant broadening Local Doppler Broadening and wider spectral features of the molecular spectra.

The reconstructions that are shown in Figure 9 and Figure 10 show the two lines of the $^{1}H_2{}^{16}O$ isotopologue of water vapor for the databases independently. Evaluating the data from these plots, the results are as expected. The lines broaden with increases in temperature and the strength at the peak of the absorption line decreases. In Figures 11 through 17, the generated spectra are nearly identical and have percent differences of less than 1.5 percent for each of the peak strengths. In Figure 18, a new peak starts to become visible and becomes stronger as the temperature continues to increase in Figure 19. This peak is not observable in the very low temperatures but is first visible in the spectrum at 300K. It can be seen as a very small spectra feature by zooming in very far into the data of Figure 13, which is shown in Figure 20.

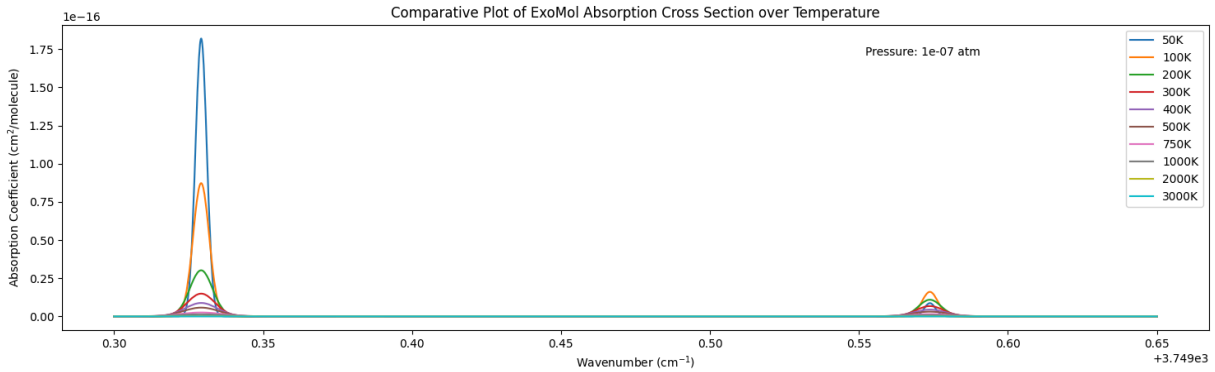



Figure 9: Doppler Broadening of ExoCross for $^{1}H_{2}{}^{16}O$ at a range of temperatures

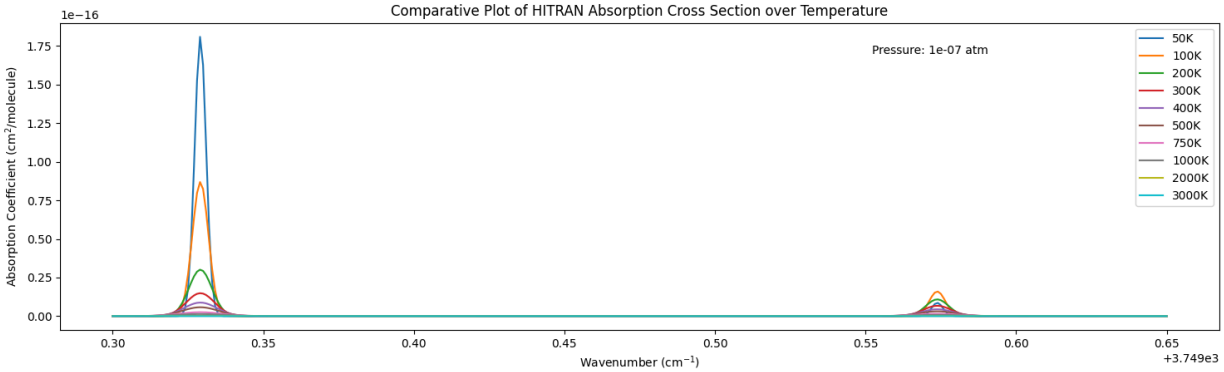



Figure 10: Doppler Broadening of ExoCross for $^{1}H_{2}{}^{16}O$ at a range of temperatures

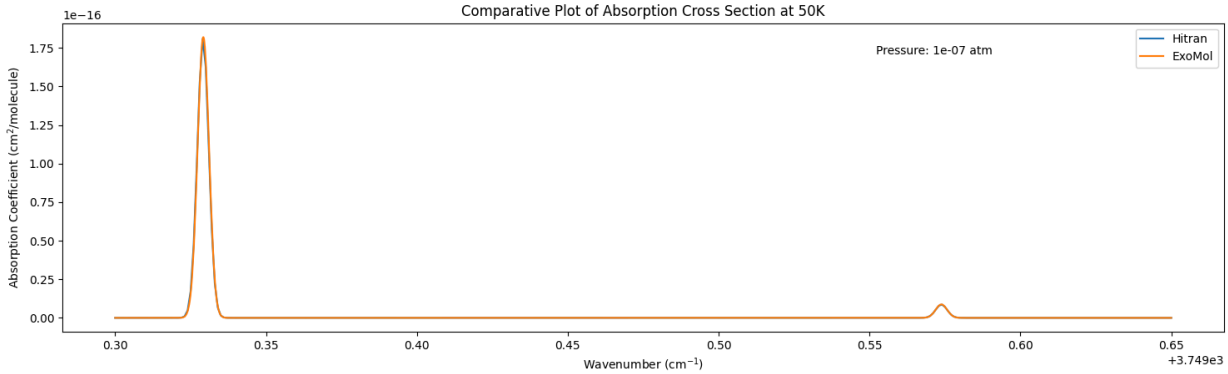



Figure 11: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 50K

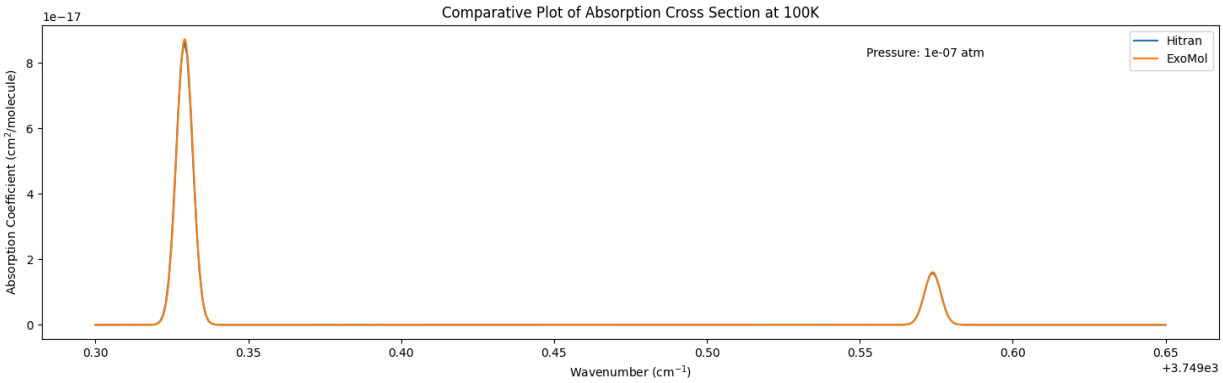



Figure 12: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 100K

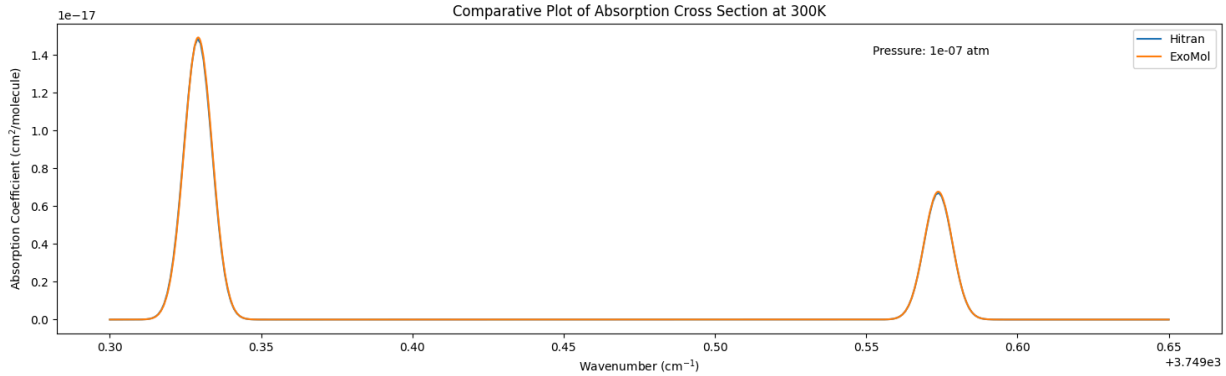



Figure 13: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 300K

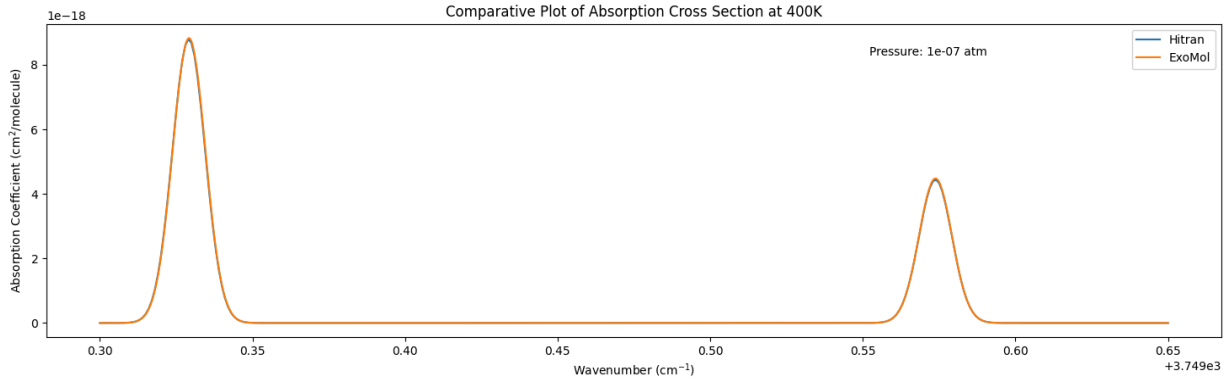



Figure 14: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 400K

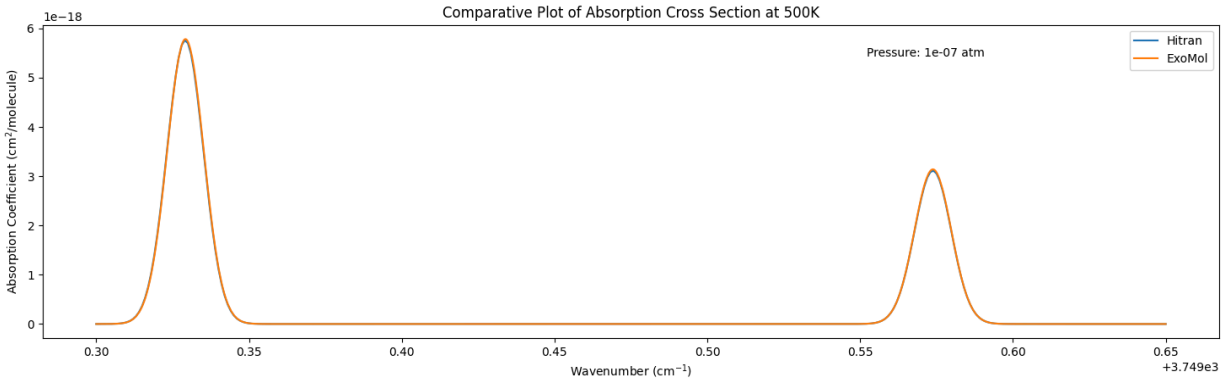



Figure 15: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 500K

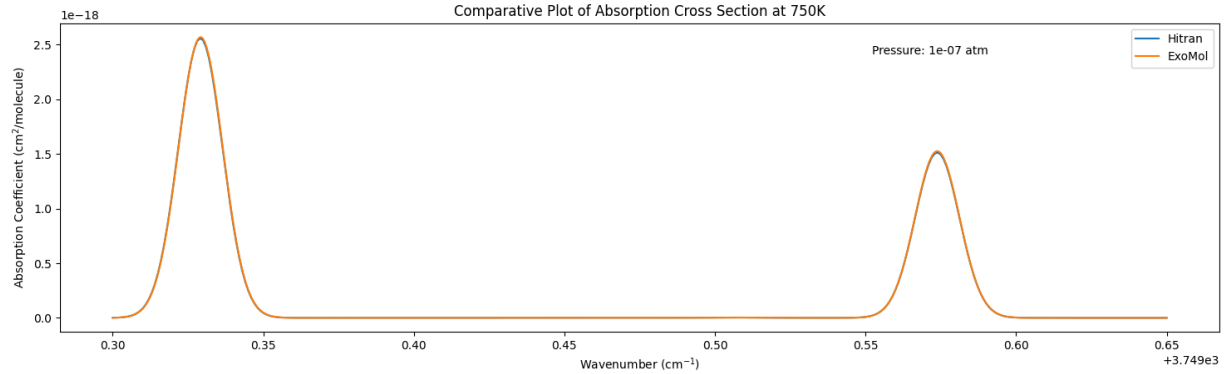



Figure 16: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 750K

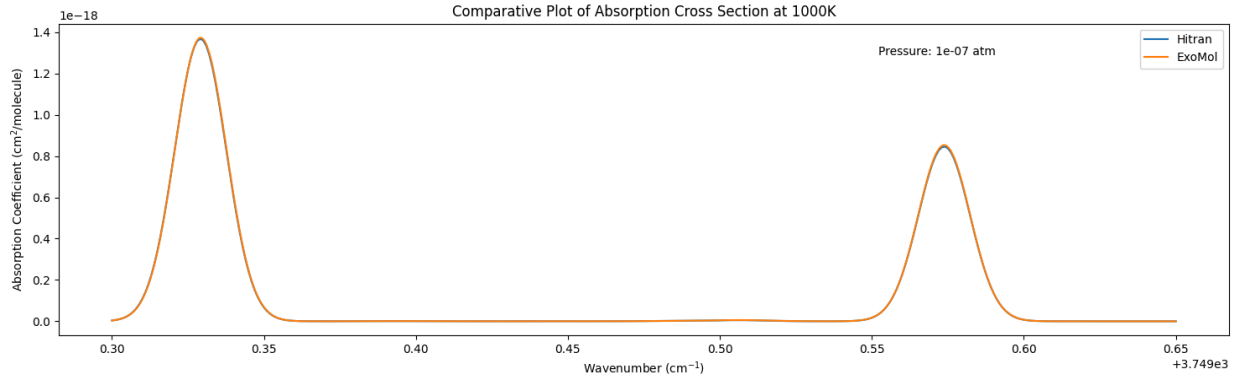



Figure 17: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 1000K

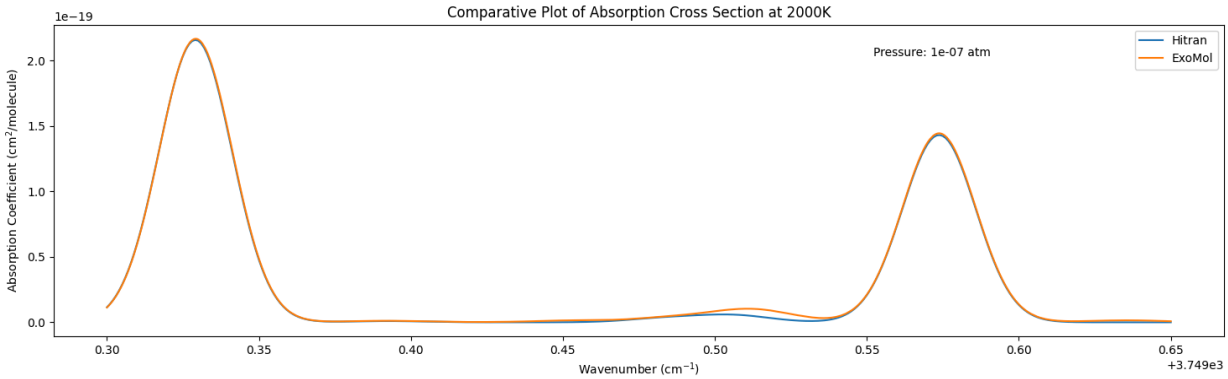



Figure 18: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 2000K

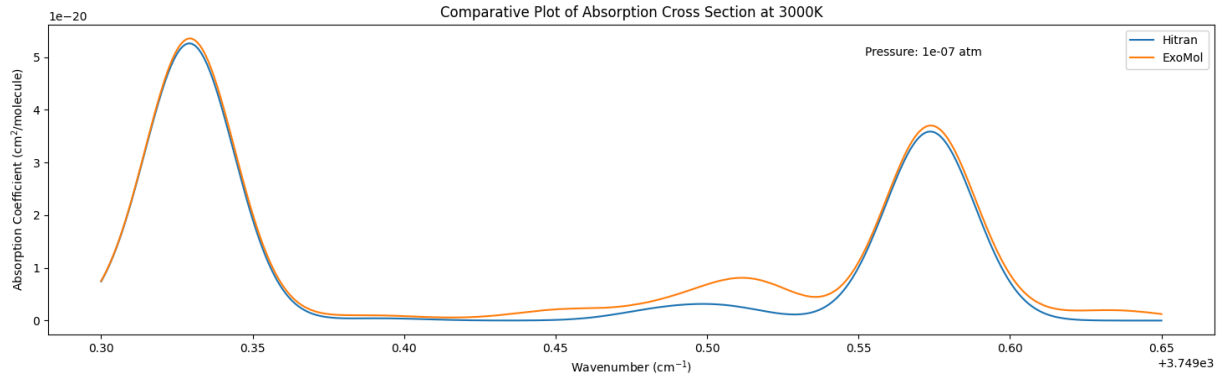



Figure 19: Doppler Broadening Comparison of HAPI and ExoCross for $^{1}H_{2}{}^{16}O$ at 3000K

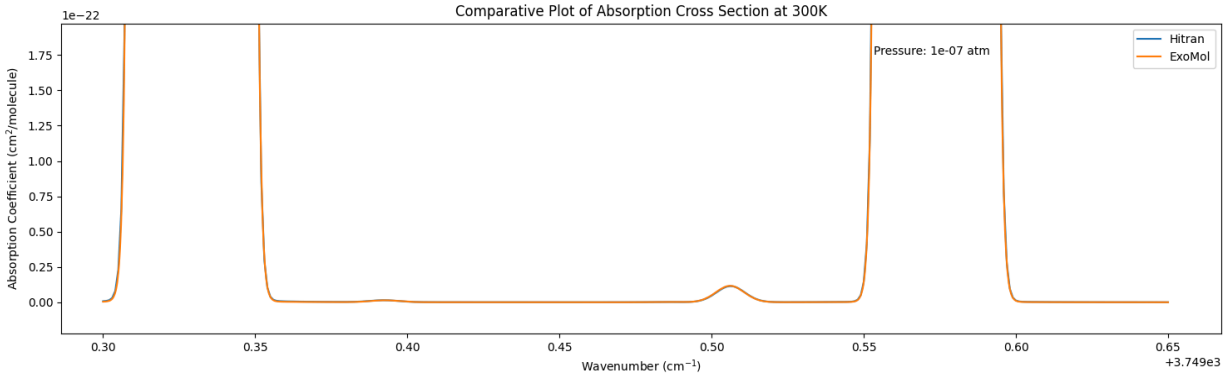



Figure 20: A zoom of Figure 14 to show the weak spectral line

The new line that appears as temperature increases is not an unexpected phenomenon. As the temperature increases and the energy of the individual molecules increases, new energy levels become populated making more transitions between energy levels available. This leads to new wavelengths being available in the same way that was discussed in Chapter 2. However, contrary to the original two lines under investigation, this new third line increases in strength as the temperature increases.

This aligns well with expectations because, the higher the temperature, the higher the likelihood that any particular molecule in a larger sample size will possess the energy required to occupy the higher energy states that are required to perform the transition that generates that wavelength of photon. Perhaps more clearly, in terms of absorption, a higher temperature means that more of the molecules are at a higher energy level. With more molecules occupying the high energy ground states, there are more molecules capable of making the transition that releases the photon that generates this third spectral line. Looking at Equation 3, the spectral line intensity is inversely proportional to the the ground state energy divided by temperature:

$$S_{ij} = I_a \frac{A_{ij}}{8\pi c \nu_{ij}^2} \frac{g'' e^{-c_2 E''/T} (1 - e^{-c_2 \nu_{ij}/T})}{Q(T)} \sim \frac{1}{e^{c_2 E''/T}} \quad \text{(Eq. 12)}$$

Therefore when E'' is much larger than the temperature, the spectral line intensity is very small. Likewise with E'' is much smaller than the temperature, the spectral line intensity is large. This, among many other reasons in a complex system, is why the strength of the original peaks under investigation decreases with increased temperature. The molecules are spread into more energy levels with more transition options.

**Variation in Pressure**

Varying the pressure in the reconstruction of the molecular spectra is the most challenging portion of the comparison. By default, the HAPI functions allow for self broadening and air broadening. Options flags are available in the HITRAN database that allow for additional broadening information to be downloaded and included in the *.par file. Since version 1.1 of HAPI, additional parameters can be included in the function used to query the data, "fetch" or "fetch_by_ids", so that they can be used.

However, once the additional broadeners have been included in the HAPI data that's downloaded, the implementation of these broadeners into the calculation requires manual flagging. An additional input parameter defining the "Diluent" ratio can be called that allows the user to specify the volume mixing ratios of the gasses directly in the absorption coefficient calculation function. It's critical to note that the sum of the parameters in the diluent specification must add to 1. As such these represent the normalized partial pressures that were described in Equation 6 and Equation 7.

The ExoMol definitions file for the Pokazatel dataset of the $^{1}H_2{}^{16}O$ isotopologue of water vapor only includes reference material relating to Helium (He) and Hydrogen ($H_2$). Other files for diluents are available on the ExoMol website but they are marked as being from HITRAN. Only the $H_2$ and He broadeners are unique to the Exomol website for this isotopologue. Review of the readme files and reference of the self broadening and air broadening functions provide no indication that there is any overlap in the Pokazatel dataset for the $^{1}H_2{}^{16}O$ isotopologue of water vapor and the HITRAN database. In fact, a paper on the 2020 update to the HITRAN database (Gordon et al.) makes a direct comment on the Pokazatel database, "*Atmospheric observations from Lampel et al.* (Lampel et al.) *indicate that the "POKAZATEL"* (Polyansky et al.) *line list underestimates the magnitude of absorption features at near-ultraviolet wavelengths (363 nm) by a factor of 2.6, a potential source of error for TEMPO retrievals.*" This quote indicates the Pokazatel dataset being used from ExoMol is considered questionable at low wavelengths due to experimental data.

There appears to be a method to add a broadener to those available from an external file. However, as mentioned before, broadener files for self broadening and air broadening in ExoMol are noted as being from HITRAN. The method to implement a new broadener was not found in the documentation for the PyExoCross. The suggestion that implementation was possible comes from a review of the source code. However, this functionality was unable to be obtained.

It's critical to note that HITRAN allows for a "self" broadener to be included. For example, a ratio of 50% "self", 25% $H_2$ and 25% He would be acceptable. ExoMol does not appear to allow for a self broadener unless one is added separately or included in the definitions

file. As such, only a ratio of $H_2$ and He will be accepted. While this critical point must be noted, HITRAN does not appear to contain the necessary data for $H_2$ and He broadening for the $^{1}H_2{}^{16}O$ isotopologue of water vapor. When this data is pursued through the website or through the HAPI Python functions, the data file shows the location where data is expected from the database and these locations are simply filled with "#". However, there is no indication that this data is unavailable prior to manual examination of the file.

When comparing the molecular absorption spectra between the two databases, the ratio of broadeners is a critical component. As with temperature related Doppler Broadening, the collisions between particles are what causes the broadening effects. The ratio between the molecule under test and other molecules is important because it defines the probability that the collision is with a certain type of molecule. As an example, heavier or lighter molecules deposit different amounts of energy.

In Figure 21 and Figure 22, the changes in the molecular spectra for ExoMol and HITRAN using the broadeners that are available is shown. Comparison data is not presented as the spectra and broadening characteristics are not expected to be the same due to the difference in diluents.

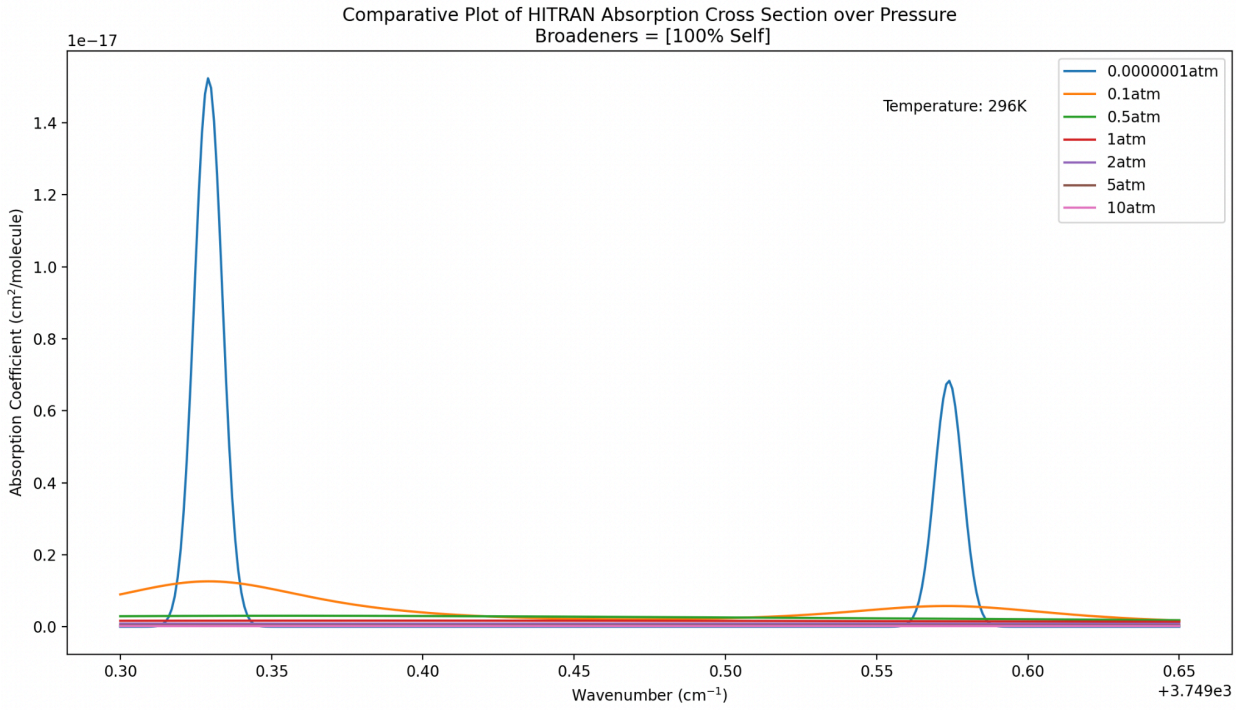



Figure 21: Pressure Broadening of HITRAN for $^{1}H_{2}{}^{16}O$ at a range of pressures

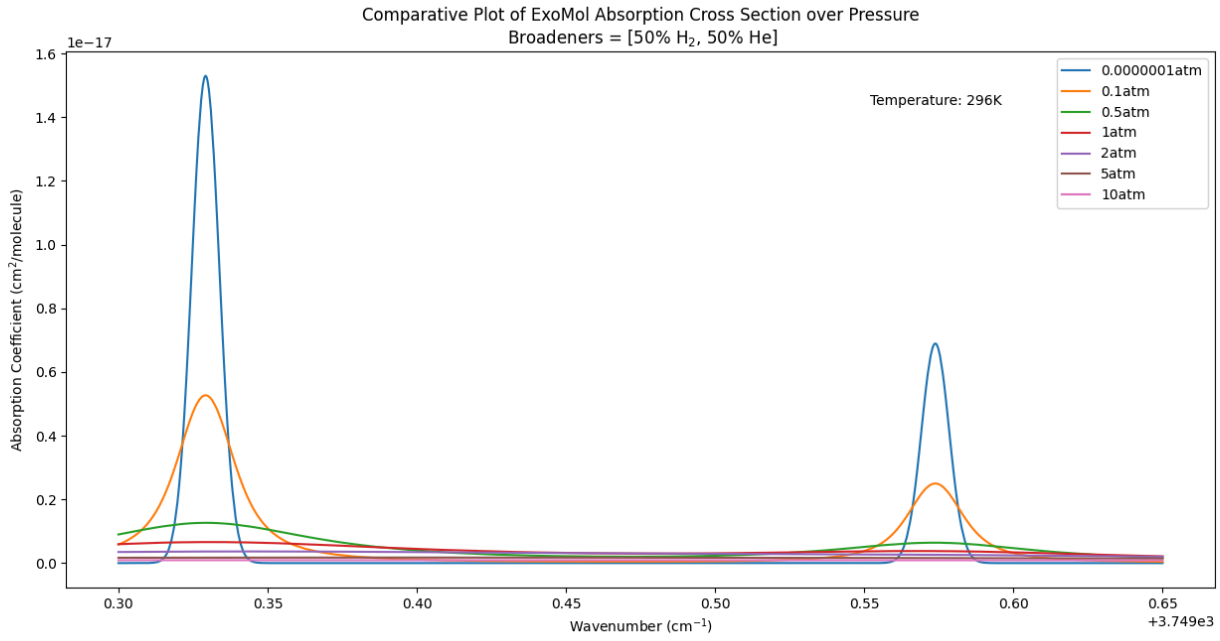



Figure 22: Pressure Broadening of ExoCross for $^{1}H_{2}{}^{16}O$ at a range of pressures

# Chapter Five: Conclusions and Recommendations

Spectroscopy remains a longstanding scientific study that continues to see regular advancement with datasets that are ever evolving and updating. The use of these databases is a regular occurrence in both scientific, academic, and industrial environments. Ease of use and confidence in the answer received are two of the most critical aspects in the quality and performance of a spectroscopic database.

In addition to the pursuit of knowledge, the HITRAN and ExoMol databases were evaluated and utilized to reconstruct spectra of absorption coefficients. It was shown that near the standard HITRAN temperature of 296K, the spectra generated between ExoMol and HITRAN are identical with <2% difference in the peak for any particular spectral feature. Once the temperatures increase far enough that the spectral features generated by high temperature conditions are comparable in strength to the low temperature spectral features, the spectra begin to deviate.

The challenges faced were documented and explained as thoroughly as possible. The standard tool sets that exist to access these databases are vast and each of them could likely be a sufficient topic for a thesis. They are very powerful tools with a very steep learning curve. In this research, some functionality was used to evaluate a single isotopologue though it's clear these tools are capable of substantially more. Several isotopologues of various molecules can be evaluated simultaneously to generate spectroscopic emissions from realistic multi-gas sources where each can be included as an emission source or an absorbance source. Largely in this work, molecules were evaluated by absorption coefficient but this is not the only method to evaluate these molecular spectra.

This work demonstrated HITRAN provides the strongest toolset for processing the molecular spectra. There is the least amount of setup and clear documentation that sets the expectations of the results quite clearly. In some cases there is a lack of reporting on the data that is received from the database when queried as seen when looking at alternative broadeners. This could be rectified by throwing an error or warning that the datatype received was not the type expected rather than processing the data and simply reporting columns of zeros. This created a rather large inconvenience when

computation times are not short. The implementation of HAPI is in Python. Python is a powerful and versatile language but it is interpreted rather than compiled. This dramatically reduces the efficiency of running large datasets leading to long computation times. As such, reporting unexpected results as early as possible is a priority.

ExoMol runs in Fortran and is very fast in processing. However, the tool set was very painful to understand due to the steep learning curve and the source code being unavailable as a reference due to lack of experience in Fortran. Once over the learning curve, the use of PyExoCross made calling the ExoCross toolset of Exomol much simpler to operate though some datasets required modification of the native PyExoCross functions. This is due to inconsistencies in the datasets and is described above in detail.

A logical continuation of this effort, beyond the quest for further knowledge, is the development of more accessible tool sets that are straightforward and usable enough for the novice or student. Many of the functions are deeply written with layered functions, methods, and classes. Accessing a low level function to add some unique data would be an immense exercise. For example, if a custom broadener needed to be added to a calculation, the challenge to accomplish that in the tool sets of the existing databases appears to be monumental. However, as mentioned before, the tool sets include vast functionality. As is often the case in programming, it may be that a simple function exists to accomplish such a goal. Simply finding the available functions can be a challenge and these functionalities are only as usable as the documentation is readable.

# Works Cited


Christian, Hill, et al. “Temperature-dependent molecular absorption cross sections for exoplanets and other atmospheres.” *Icarus*, vol. 226, no. 2, 1677, p. 1673. *ScienceDirect*, https://www.sciencedirect.com/science/article/abs/pii/S0019103512003041.

Furtenbacher, Tibor, et al. “Definitive Ideal-Gas Thermochemical Functions of the H216O Molecule.” *Journal of Physical and Chemical Reference Data*, vol. 45, no. 4, 2016, p. 45. *AIP Publishing*, https://pubs.aip.org/aip/jpr/article-abstract/45/4/043104/959177/Definitive-Ideal-Gas-Thermochemical-Functions-of?redirectedFrom=fulltext.

Gamache, Robert, et al. “TIPS_2021, Total Internal Partition Sums for Molecules on HITRAN2020 database.” *Journal of Quantitative Spectroscopy and Radiative Transfer*, no. TBD(HITRAN2020 Special Issue), 2021. *Zenodo*, https://zenodo.org/records/4708099.

Gordon, I. E., et al. “The HITRAN2020 molecular spectroscopic database.” *Journal of Quantitative Spectroscopy & Radiative Transfer*, vol. 277, 2022, pp. 1-82.

Griffiths, David J., and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. Cambridge University Press, 2018.

Grimm, Simon L., et al. “HELIOS-K 2.0 Opacity Calculator and Open-source Opacity Database for Exoplanetary Atmospheres.” *The Astrophysical Journal Supplement Series*, vol. 253, no. 1, 2021, p. 38, https://ui.adsabs.harvard.edu/abs/2021ApJS..253...30G/abstract.

Hanson, Ronald K., et al. *Spectroscopy and Optical Diagnostics for Gases*. Springer International Publishing, 2015.

Kochanov, Roman V. *HITRAN Application Programming Interface (HAPI) User Guide*. 4.3 ed., Harvard-Smithsonian Center for Astrophysics, 2019.

Lampel, Johannes, et al. “Detection of water vapour absorption around 363nm in measured atmospheric absorption spectra and its effect on DOAS evaluations.” *Atmospheric Chemistry and Physics*, vol. 17, 2017, pp. 1271-1295, https://acp.copernicus.org/articles/17/1271/2017/.

Polyansky, Oleg L., et al. “ExoMol molecular line lists XXX: a complete high-accuracy line list for water.” *Monthly Notices of the Royal Astronomical Society*, vol. 480, no. 2, 2608, p. 2597. *Oxford Academic*, https://academic.oup.com/mnras/article/480/2/2597/5054049.

Qu, Q., et al. “ExoMol molecular line lists – XLII. Rovibronic molecular line list for the low-lying states of NO.” *Monthly Notices of the Royal Astronomical Society*, vol. 504, 2021, pp. 5768-5777, https://doi.org/10.1093/mnras/stab1154.

Rothman, L. A., et al. “The HITRAN 2004 molecular spectroscopic database.” *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 96, no. 2, 2005, pp. 139-204. *ScienceDirect*, https://doi.org/10.1016/j.jqsrt.2004.10.008.

Shankar, Ramamurti. *Principles of Quantum Mechanics*. Springer US, 1980.

Tennyson, Jonathan, et al. “The ExoMol database: Molecular line lists for exoplanet and other hot atmospheres.” *Journal of Molecular Spectroscopy*, vol. 327, 2016, pp. 73-94. *ScienceDirect*, https://doi.org/10.1016/j.jms.2016.05.002.

Wolfson, Richard, and Jay M. Pasachoff. *Physics with Modern Physics for Scientists and Engineers*. Addison-Wesley, 1999.

Wong, A., et al. “ExoMol line list XXI: nitric oxide (NO).” *Monthly Notices of the Royal Astronomical Society*, vol. 470, 2017, pp. 882-897, http://dx.doi.org/10.1093/mnras/stx1211.

Yurchenko, Sergey. “ExoCross 1 Documentation.” *ExoCross 1 Documentation*, 2015, https://exocross.readthedocs.io/en/latest/index.html.

# Vita

Joseph Braker was born in Tullahoma, Tennessee. He went to Rose-Hulman Institute of Technology for undergraduate study where he graduated with a Bachelor's of Science in Optical Engineering in 2013. After graduating, he began work at DILAS Diode Laser, Inc. in Tucson, AZ working as an Optical Engineer designing diode lasers. Joseph was promoted through the company to Product Line Manager and stayed in this role after DILAS was acquired by Coherent, Inc. In 2018, Joseph began work at Quantel USA, a laser manufacturer specializing in diode and solid state lasers, as the Director of Business Development. He worked in this role until moving back to Tullahoma and starting work as a scientist at Arnold Air Force Base in the Advanced Technology group. In parallel to his work, Joseph is pursuing a Masters of Science in Physics from the University of Tennessee. Joseph is married to his wife Sarah and has two daughters, Allison and Gillian.