Atmospec group project

Final report

Abstract

The Atmospec application has been developed as a tool for use by atmospheric chemists, allowing the calculation of photo absorption cross-sections in a single simplified workflow. The application provides a more accessible path for researchers unfamiliar with computational chemistry to leverage advances in quantum mechanical calculations. This project adds a graphing widget and photolysis rate constant calculation to the existing functionality within the app. The prototype function, as described in the progress report, was integrated into the atmospec application successfully and tested to ensure accuracy.

Introduction

Volatile organic compounds (VOCs), defined by their high vapor pressure and low water solubility, often have extremely short lifetimes due to their reactivity, making them difficult to study in a lab setting. VOCs can be extremely dangerous to both public health and the atmospheric environment even at low concentrations [1,2], meaning that it is of upmost importance to monitor and model the levels of such compounds to better understand their overall effect. Currently, atmospheric modelling algorithms, such as the master chemical model, only consider the ground state reactivity of VOCs resulting in discrepancies between modelled and observed concentrations [3,4]. In the atmosphere primary VOCs, such as acetone, will react to form transient secondary volatile organic compounds, many of which contain chromophoric groups, with the potential to absorb light. Absorbed light will excite the molecules to higher energy levels allowing photodissociation, the breakdown of a molecule due to light, to occur.

As discussed in the progress report [5], the kinetics of photolysis reactions can be quantified using the photolysis rate constant (J), see equation 1, where J is calculated as the integral over wavelength (λ / nm) of differential j-components [6]. J-components are calculated as the product of the absorption cross-section (a measure of how likely the molecule is to absorb light), the actinic flux (intensity of light), and the quantum yield of the photodissociation pathway. The transient nature of volatile organic compounds makes the empirical measurement of absorption cross-sections and quantum yield difficult, resulting in an information gap in databases such as the Mainz spectral database for transient VOCs [7]. Computational chemistry provides a method for calculating the photo absorption cross-sections from molecular structures by Nuclear Ensemble Approach [8,9], which has been implemented in the Atmospec app using ORCA for quantum mechanical calculations.

$$J = \int_{\lambda_{min}}^{\lambda_{max}} \sigma(\lambda)\phi(\lambda)F(\lambda)d\lambda$$

Equation 1

Atmospec produces a photo absorption cross-section from a SMILES string, simplifying the process of submitting a quantum mechanical calculation on a given structure within a single workflow. This abstraction of a complex computational process makes advances in computational chemistry accessible to researchers less familiar with such methods, giving them access to quantum mechanical tools in their respective areas of research. Atmospec is built on the AiiDalab ecosystem, a cloud-based environment for the development and distribution of scientific workflows. All the modules used are held in a docker container, streamlining installation. Figure 1 shows an absorption cross section calculated in Atmospec from the SMILES string for C_3H_4O, while this is already a remarkable achievement; Discussions with atmospheric chemists have resulted in the request for addition functionality, namely the calculation of photolysis rate constants. This project aims to add such functionality to the Atmospec app in the form of a plotting widget located below the photo absorption cross-section where the photolysis rate constant for the given reaction pathway is displayed alongside a graph of differential j-components.

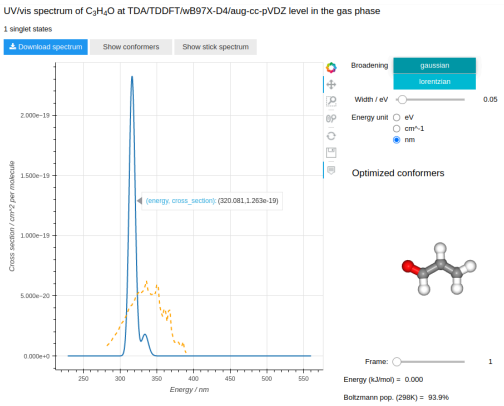

Figure 1: calculated photo absorption cross-section of C_3H_4O

Method & results

The group was subdivided into roles, helping us distribute tasks appropriately among team members to most effectively complete the project. Marco was voted chair, Fay – deputy leader, and Emily as the presenter. Since atmospec currently only supports Linux or MacOS, Will and I took on the role of interacting directly with the Atmospec code. Weekly meetings with our supervisors were arranged as a way of checking in and getting feedback on ideas, as well as meeting in smaller groups when necessary. An initial workflow for the project was agreed with the group and summarized in a flow chart, see figure 2, the flowchart is split into three main sections; coding, reporting results, and presentation preparation with group members focusing on different parts. My contribution to the project was mainly in the integration of our photolysis rate constant calculations and plotting function into the existing atmospec code

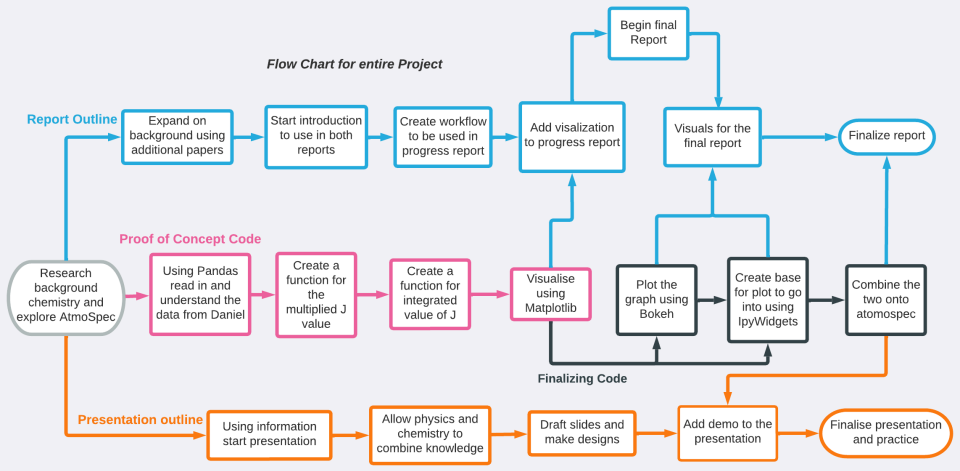

Figure 2: Workflow chart for the project

Prototype functions for the calculation of J and plotting of differential j values were developed using NumPy and Matplotlib. Subsequently, this prototype was adapted to use Bokeh for plotting, since the Atmospec code uses Bokeh, therefore using matplotlib would add an additional dependency. A separate window, PhotolysisPlotWidget, was created in the spectrum analysis window, an initial plot was created using the prototype functions and sample cross-section data for pinacolone. Interactive widgets to accommodate changes to the actinic flux level or quantum yield were added and connected to the graph so that changes both within the PhotolysisPlotWidget and in the spectrum widget are reflected in the graph. Figure 3 shows a basic flowchart of data with respect to the plotting widget, starting with the spectrum widget class where the cross-section is initially calculated and plotted. The integration is done using wavelengths in nm; therefore, some logic was added to ensure that the data passed to the plotting widget is in nm, regardless of the units selected by the user.

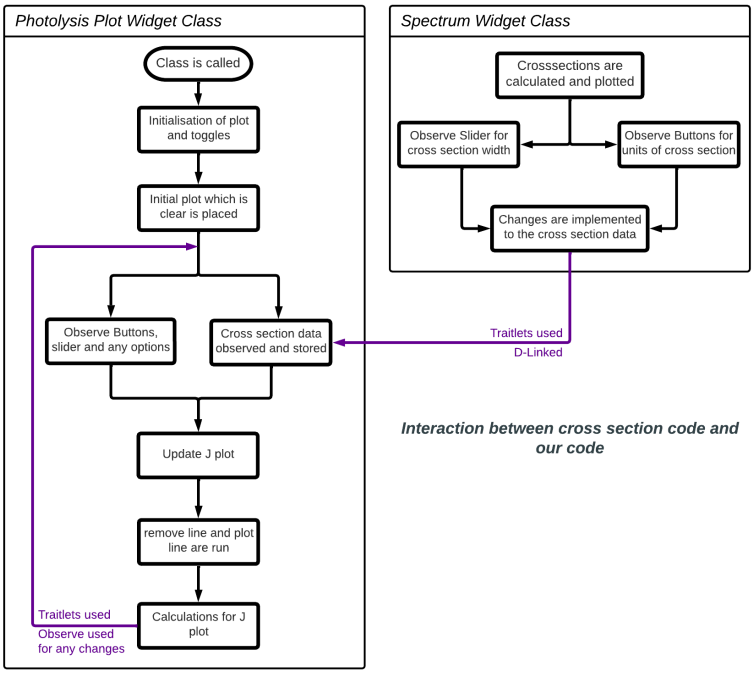

Figure 3: Data flow between spectrum.py and spectrum_analysis.py

Traitlets were used as a means of making dynamic updates to the plots, they are variables that can track changes that occur to themselves, holding the current state and the previous state of the variable. This is extremely useful for handling changes made on the fly using the observer decorator functions to automate the handling of updates to buttons and sliders. Once traitlets had been implemented this allowed us to start working with data straight from the in-app calculations and refining our plotting. Here the weekly meetings were of great benefit, providing an iterative development cycle with prompt feedback and suggestions for improvements. Once the product was developed, a pull request was made to the original ISPG GitHub and a more thorough review of the code began, here the development shifted from large changes in the code to smaller formatting changes and more concentrated optimization of functions.

Figure 4 shows an example of the j-components plot for C_3H_4O at TDA/TDDFT/wB97X-D4/aug-cc-pVDZ level in the gas phase, with a logarithmic axis for the actinic flux level, showcasing the final product of this group project.

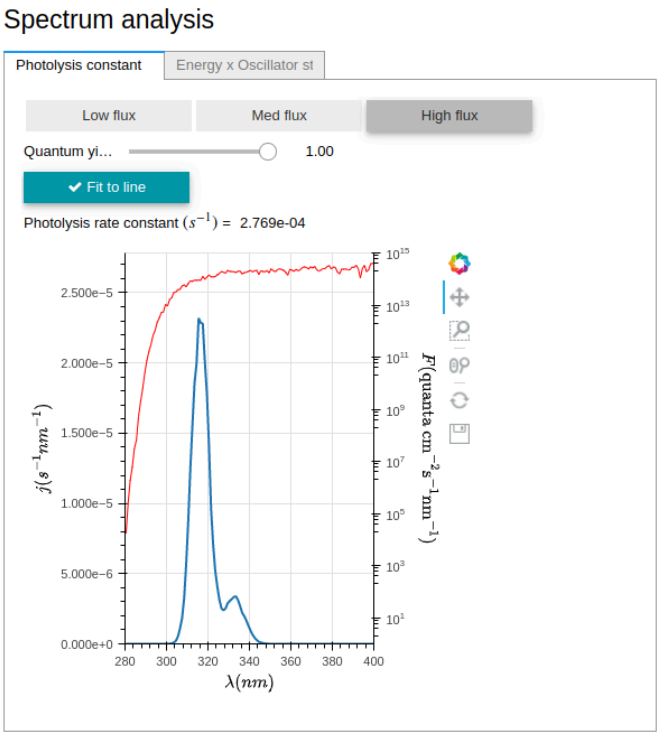

Figure 4: Example plot of j-components for C_3H_4O

While the data for actinic flux was evenly sampled with difference in wavelength of 1 between datapoints, the cross-section data was not, calling for the need to interpolate to set the data to the same wavelength for integration. NumPy functions were evaluated, ensuring that they are performing as expected, beginning with a visual evaluation of the NumPy interpolation function to satisfy that no information has been lost in the interpolation. The interpolation function was applied successfully, with minimal data loss, see Figure 5, showing the photo absorption cross-section before and after interpolation. The NumPy convolve smoothing function was also evaluated for the best smoothing, see Figure 6, the window size was compared to the change in photolysis rate constant, finding that the optimal window size with no loss in data (ie. 0% change from original value), with significantly a smoothed appearance. Smoothing is an important step, as without smoothing we are showing the user accuracy that we cannot promise with this calculation due to noise in the data.

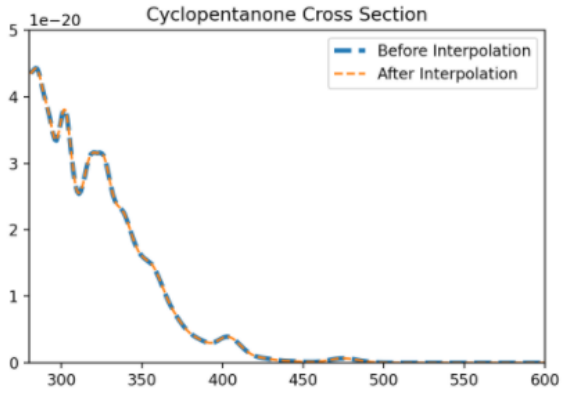

Figure 5: Checking for data loss in interpolation

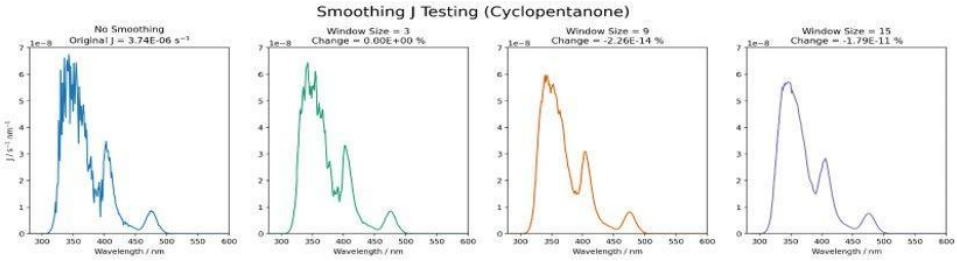

Figure 6: Testing the effect of np.convolve on smoothness

Our functions use the NumPy ‘trapz’ method for calculating the numerical integration using the trapezoidal rule. The function takes values of J, assuming a dx of 1, calculating the integral by the trapezium rule. The result was tested against values calculated by hand to confirm their validity, furthermore the condensation of data was investigated, see Figure 6. It was found that decreasing the percentage of data used for the integration greatly increased the error on the calculation, therefore it was decided to use all data for the integration in favour of accuracy over run speed. The time taken for the calculation (10^{-5} s) is so little that this is a fair trade-off, with little effect on runtime.

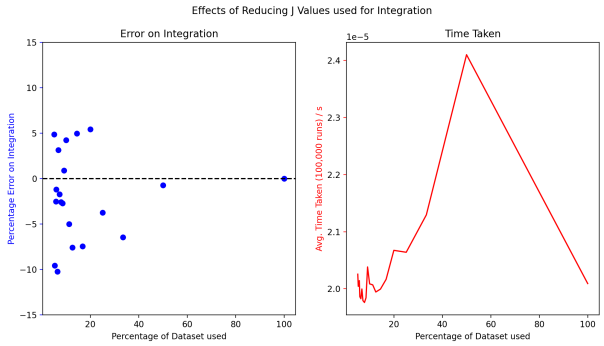

Figure 7: Integration testing over 100,000 iterations of np.trapz

Finally, the accuracy of our product was tested by comparison to the photolysis rate constant for I_2 calculated using empirical data. Our computationally calculated value yielded a value of 0.178 s^{-1}

compared to the reported value of 0.12 ± 0.03 s^{-1} [10]. While there is a difference between the two values, this can be explained by a different sampling resolution in the published cross-section data, additionally the flux data used was not published, so standard actinic flux data was used, potentially accounting for the difference.

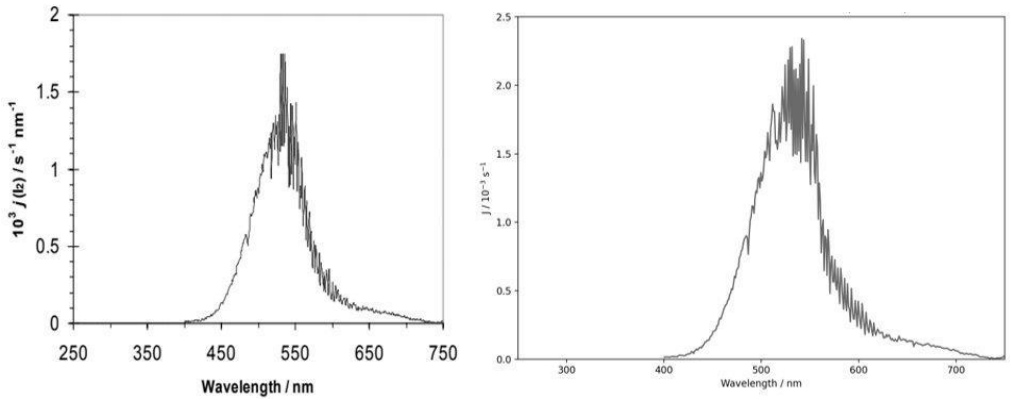

Figure 8: Testing against I_2 data (left) empirical cross-section (right) calculated cross-section

Figure 9 shows the planned and actual project timelines for the whole project; broadly speaking, this plan was successfully implemented. Any deviations from the timeline were agreed upon as a group, ensuring that help is provided to teammates when required to ensure a timely completion of the product.

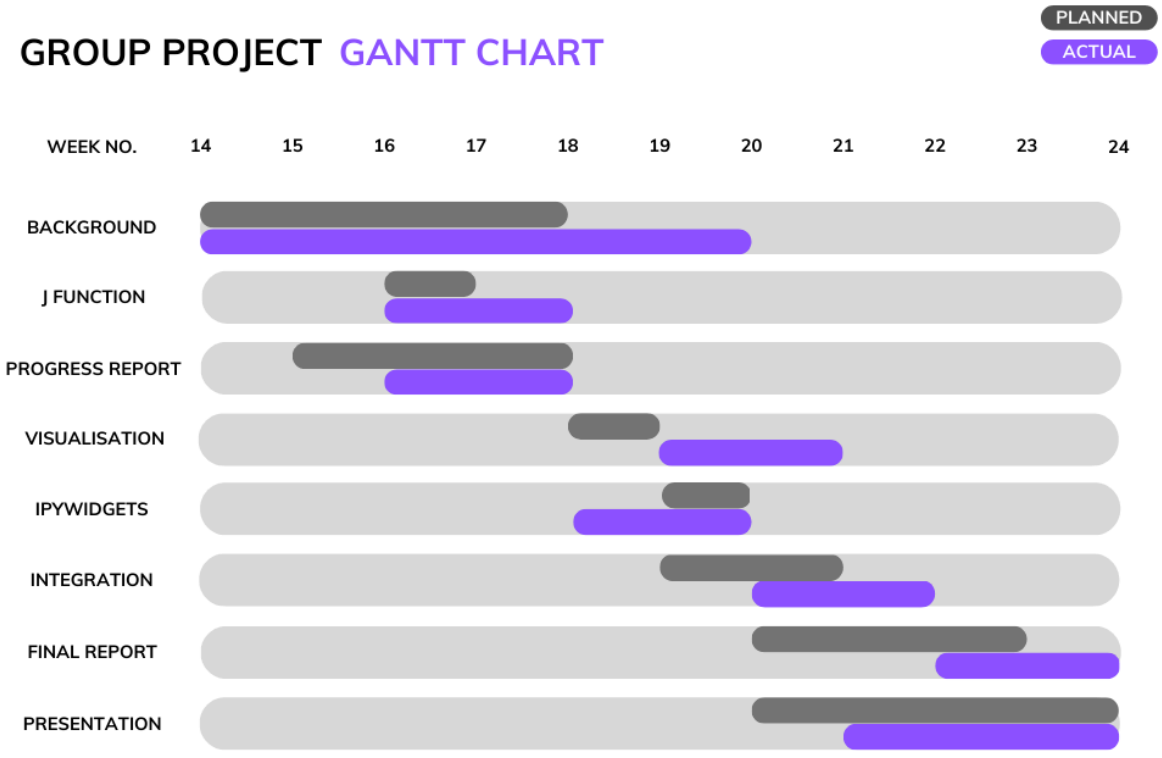

Figure 9: Chart visualizing progress over time

Conclusion

The overall aims of the project, to provide a widget for the calculation and visualization of the photolysis rate constant and differential rate components, have been achieved successfully. The output graph responds to changes made by the user in the user interface, updating dynamically upon a change in inputs. The timeline, as presented in the progress report, has been broadly followed with minor deviations where appropriate, notably the integration took longer than expected to start as we were analyzing the existing code and grasping the uses of different functions and modules within the code. The product has been presented to a group of peers, and a pull request has been submitted to the original ISPG atmospec GitHub to merge the additions we have made.

References

1. Li, A.J., Pal, V.K. and Kannan, K. (2021) "A review of environmental occurrence, toxicity, biotransformation and biomonitoring of volatile organic compounds," *Environmental Chemistry and Ecotoxicology*, 3, pp. 91–116.
2. David, E. and Niculescu, V.-C. (2021) "Volatile Organic Compounds (vocs) as environmental pollutants: Occurrence and mitigation using nanomaterials," *International Journal of Environmental Research and Public Health*, 18(24), p. 13147.
3. Peeters, J., Nguyen, T.L. and Vereecken, L. (2009) "Hox radical regeneration in the oxidation of isoprene," *Physical Chemistry Chemical Physics*, 11(28), p. 5935.
4. https://mcm.york.ac.uk/home.htt
5. Nomerotski, K. (2023) "Atmospec progress report"
6. King, M. *Calculating photolysis rates and estimating photolysis lifetimes*, *Calculating photolysis rates and estimating photolysis lifetimes*. RSC.
7. Keller-Rudek, H., Moortgat, G. K., Sander, R., and Sörensen, R.: *The MPI-Mainz UV/VIS spectral atlas of gaseous molecules of atmospheric interest,* Earth Syst. Sci. Data, 5, 365–373, (2013),
8. Prlj, A. *et al.* (2020) "On the theoretical determination of photolysis properties for atmospheric volatile organic compounds," *The Journal of Physical Chemistry Letters*, 11(14), pp. 5418–5425.
9. Prlj, A. *et al.* (2021) "Calculating photoabsorption cross-sections for atmospheric volatile organic compounds," *ACS Earth and Space Chemistry*, 6(1), pp. 207–217.
10. Saiz-Lopez, A. *et al.* (2004) "Absolute absorption cross-section and photolysis rate of I2," *Atmospheric Chemistry and Physics*, 4(5), pp. 1443–1450.