



## Line-broadening for exoplanetary atmospheres: extension of COLLINE database using machine learning

Sujet de thèse de Doctorat proposé par Jeanna Buldyreva (Pr, UTINAM)

Il a été établi récemment que presque toutes les étoiles de notre Galaxie supportent un système planétaire. Les molécules qui composent l'atmosphère d'une exoplanète déterminent son évolution et peuvent être identifiées à partir de leurs signatures spectrales. A cause des conditions extrêmes (hautes températures, très forts niveaux d'insolation), les caractéristiques spectrales de référence ne peuvent pas être mesurées en laboratoire dans la plupart des cas et demeurent inconnues ; leurs estimations théoriques sont alors nécessaires pour interpréter correctement des spectres enregistrés depuis le sol ou par les missions spatiales. Notamment, les données théoriques de forme de raies spectrales (élargissement et déplacement par pression) sont devenues d'une première urgence suite à la mission ARIEL (Atmospheric Remote-sensing Infrared Exoplanet Large-survey, lancement en 2029) par l'European Space Agency. ARIEL est généralement destinée à étudier des objets chauds, avec une température d'équilibre supérieure à 500K, en analysant leurs spectres dans les longueurs d'ondes InfraRouge-Visible-UltraViolet. Les grandes performances des instruments d'ARIEL rendent indispensable la création de bases de données spectroscopiques « haute résolution » - incluant non seulement les positions et les intensités de raies spectrales, mais aussi leurs largeurs et déplacements.

Un premier pas dans ce sens a été fait par la création d'une nouvelle section *Line Shapes* dans la base de données ExoMol (<https://www.exomol.com/>) localisée à l'*University College London* et soutenue par deux financements de l'*European Research Council* (2011-2016 et 2020-2025). Après avoir travaillé dans l'équipe ExoMol dans le cadre d'une mobilité sortante (décembre 2021 – juillet 2022) sur les développements méthodologiques et production de premières estimations [1,2] inclus dans l'édition 2024 d'ExoMol [3] et dans l'article de revue du groupe de travail *Spectroscopic databases* du consortium ARIEL [4], la porteuse de la présente demande a créé une base de données spécifiquement dédiée aux largeurs de raie collisionnelles des exo-molécules : COLLINE (COLLisional-LINE-broadening; <https://colline.u-bourgogne.fr/>) couvrant 51 molécules actives et 12 gaz-perturbateurs [5].

Cependant, les estimations théoriques remplissant actuellement COLLINE sont des valeurs constantes pour chaque paire moléculaire "exotique", alors que les dépendances en nombres quantiques rotationnels sont généralement bien prononcées et doivent être modélisées correctement (par exemple, *James Webb Space Telescope* avec résolution  $R \sim 1000-3000$  aura des erreurs de 40% dans les cross-sections si l'élargissement collisionnel de raies est ignoré [Hedges & Madhusudhan, MNRAS 458, 1427; 2016]). Comme les calculs traditionnels sont très coûteux en temps CPU, mais surtout comme il n'y a pas de potentiels d'interaction intermoléculaire pour les exo-molécules, des approches alternatives modernes s'avèrent indispensables.

Le présent sujet de thèse vise la mise en place de procédures efficaces pour la production de paramètres d'élargissement, avec un accent particulier sur l'exhaustivité. Notamment, des premières tentatives de techniques *Machine Learning* [6,7] se sont montrées relativement bonnes pour fournir des prédictions valides à 69% quand l'algorithme est entraîné sur des données de l'élargissement par l'air de la base HITRAN (<https://hitran.org/>) [Guest et al., JMS 401, 111901 (2024)]. Dans le cadre de la thèse, après avoir affiné la sélection des *features*, différents modèles (*gradient boosting*, *adaptive boosting*, *random forest*, *voting*,...) [Sarker, SN Comput.Sci. 2, 160 (2021)] seront testés d'abord sur des données HITRAN, puis avec une extension à d'autres perturbateurs, en utilisant le *scikit-learn Python package* [Pedregosa et al., J. Mach. Learn. Res. 12, 2825 (2011)].

La demande de ce type de données ne se limite pas par les besoins de la mission ARIEL, mais est aussi constamment soulevée par les chercheurs impliqués dans les modélisations des atmosphères planétaires "chaudes". La production de ce type de données dans le cadre d'une thèse à Besançon permettra d'enrichir les compétences d'UTINAM dans le domaine d'astrophysique et de les faire valoir au niveau mondial, dans le cadre d'une base de données Open Source.

Sélection de publications sur le sujet :

- [1] **J. Buldyreva**, S.N. Yurchenko, J. Tennyson “Simple semi-classical model of pressure-broadened infrared/microwave linewidths in the temperature range 200–3000 K” *RAS Techniques and Instruments* 1, 43–47 (2022); <https://doi.org/10.1093/rasti/rzac004>
- [2] **Jeanna Buldyreva**, Ryan Brady, Sergey N. Yurchenko, Jonathan Tennyson “Collisional broadening of molecular rovibronic lines for exoplanetary studies”, *JQSRT* 313(122), 108843 (2024); DOI:10.1016/j.jqsrt.2023.108843
- [3] Jonathan Tennyson, Sergei N. Yurchenko, Jingxin Zhanga, Charles A. Bowesman, Ryan P. Brady, **Jeanna Buldyreva**, Katy L. Chubb, Robert R. Gamache, Maire N. Gorman, Elizabeth R. Guest, Christian Hill, Kyriaki Kefala, A. E. Lynas-Gray, Thomas M. Mellor, Laura K. McKemmish, Georgi B. Mitev, Irina I. Mizus, Alec Owens, Zhijian Peng, Armando N. Perri, Marco Pezzella, Oleg L. Polyansky, Qianwei Qu, Mikhail Semenov Oleksiy Smola, Andrei Solokov, Wilfrid Somogyi, Apoorva Upadhyay, Samuel O.M. Wright, Nikolai Zobov “The 2024 release of the ExoMol database: molecular line lists for exoplanet and other hot atmospheres” *JQSRT* 326, 109083 (2024); <https://doi.org/10.1016/j.jqsrt.2024.109083>
- [4] Katy L. Chubb, Séverine Robert, Clara Sousa-Silva, Sergei N. Yurchenko, Nicole F. Allard, Vincent Boudon, **Jeanna Buldyreva**, Benjamin Bulte, Athena Coustenis, Aleksandra Foltynowicz, Iouli E. Gordon, Robert J. Hargreaves, Christiane Helling, Christian Hill, Helgi Rafn Hrodmarsson, Tijs Karman, Helena Lecoq-Molinos, Alessandra Migliorini, Michaël Rey, Cyril Richard, Ibrahim Sadiek, Frédéric Schmidt, Andrei Sokolov, Stefania Stefani, Jonathan Tennyson, Olivia Venot, Sam O. M. Wright, Rosa Arenales-Lope, Joanna K. Barstow, Andrea Bocchieri, Nathalie Carrasco, Dwaipayan Dubey, Oleg Egorov, Antonio García Muñoz, Ehsan (Sam) Gharib-Nezhad, Leonardos Gkouvelis, Fabian Grübel, Patrick Gerard Joseph Irwin, Antonín Knížek, David A. Lewis, Matt G. Lodge, Sushuang Ma, Zita Martins, Karan Molaverdikhani, Giuseppe Morello, Andrei Nikitin, Emilie Panek, Miriam Renge, Giovanna Rinaldi, Jack W. Skinner, Giovanna Tinetti, Tim A. van Kempen, Jingxuan Yang, Tiziano Zingales “Data availability and requirements relevant for the Ariel space mission and other exoplanet atmosphere applications” *RAS Techniques and Instruments* 3(1), 636–690 (2024); <https://doi.org/10.1093/rasti/rzae039>
- [5] **Jeanna Buldyreva**, Kathleen Stehlin, Sergey N. Yurchenko, Elizabeth Guest, Jonathan Tennyson “Semi-classical estimates of pressure-induced linewidths for infrared absorption by hot (exo)planetary atmospheres”, *Astrophysical Journal Supplement Series*, 276(1), 23 (2025); <https://doi.org/10.3847/1538-4365/ad9b19>
- [6] E.R. Guest, J. Tennyson, S.N. Yurchenko, **J. Buldyreva** “Using machine learning to meet the need for pressure-broadening data in exoplanetary atmospheric studies”, 15th HITRAN meeting, Reims, France, 24-26 August 2022.
- [7] E.R. Guest, J. Tennyson, S.N. Yurchenko, **J. Buldyreva** “Using Machine Learning to meet the need for pressure-broadening data in exoplanetary atmospheric studies”, XXVIII Colloquium on High Resolution Molecular Spectroscopy, Dijon, France, 28 August – 1 September 2023.