The Journal of Open Source Software

taurex-emcee: a TauREx 3.1 plugin for the emcee sampler

Andrea Bocchieri [1], **Quentin Changeat** [2], **Lorenzo V. Mugnai** [3,4,5], and **Enzo Pascale** [1]

1 Department of Physics, La Sapienza Università di Roma, Piazzale Aldo Moro 2, Roma, 00185, Italy **2** European Space Agency (ESA), ESA Office, Space Telescope Science Institute (STScI), Baltimore, MD, 21218, USA **3** School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff, CF24 3AA, UK **4** Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK **5** INAF, Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, Palermo, I-90134, Italy

DOI: N/A

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: 01 January 1970

Summary

taurex-emcee is a plugin for the TauREx 3.1 atmospheric retrieval framework (A. F. Al-Refaie et al., 2021) that extends the choice of sampling methods available to the user. The plugin provides an interface to the emcee sampler (Foreman-Mackey et al., 2013), a popular affine-invariant ensemble sampler widely used in the astronomy community. Running the sampler to convergence is automated through the autoemcee package, which also supports parallelization with MPI. Thus, the taurex-emcee plugin allows users to easily launch parallelized retrievals of atmospheric spectra with emcee. This enables reliable, efficient, and fast retrievals, especially when coupled with TauREx's GPU-accelerated forward models (A. Al-Refaie et al., 2020).

taurex-emcee is released under the BSD 3-Clause license and is available on GitHub. The plugin can be installed from the source code or from PyPI, so it can be installed as `pip install taurex-emcee`. The documentation is available on readthedocs, including a quick-start guide, a tutorial, a description of the software functionalities, and guidelines for developers. The documentation is continuously updated and is versioned to match the software releases.

Benchmark

We benchmarked the taurex-emcee plugin against the MultiNest sampler (F. Feroz et al., 2009; Farhan Feroz et al., 2019), natively implemented in TauREx 3, on a synthetic transmission spectrum of the hot Jupiter HD 209458b. The aim being first to assess the computational time of the two samplers on a controlled case, second to assess the consistency of the results from the retrievals. The synthetic spectrum and the retrievals were performed on a single node of the Sapienza University of Rome Melodie server, equipped with one NVIDIA A40 GPU. The high-resolution input spectrum was generated assuming stellar and planetary parameters of HD 209458b from Edwards et al. (2019), reported in Table 1, and a gaseous atmosphere with hydrogen and helium at a ratio H_2/He = 0.172 and an isothermal temperature profile.

Table 1: Summary of the selected target's properties

R_p [R_J]	M_p [M_J]	T_p [K]	P [d]	R_s [$R_\odot$]	Mag K	T_s [K]	M_s [$M_\odot$]
1.35	0.71	1613	3.52	1.18	6.31	6,086	1.18

The atmosphere is simulated with five molecular species as trace gases: H_2O (100 ppm), CH_4

(10 ppm), CO (1 ppm), CO_2 (0.1 ppm), and NH_3 (0.01 ppm). Molecular abundances are assumed constant with altitude. Cross sections at a resolution of 15,000 are used for all species, as given in Table 2. Collision-induced absorption (CIA) with H_2–H_2 and H_2–He and Rayleigh scattering are included in the calculation. Finally, 100 pressure layers are used to sample the atmosphere, with a minimum pressure of 1 Pa and a maximum pressure of 10^6 Pa.

Table 2: Cross sections and CIA used in the simulations

Opacity	Reference(s)
H_2-H_2	Abel et al. (2011), Fletcher et al. (2018)
H_2-He	Abel et al. (2012)
H_2O	Polyansky et al. (2018)
NH_3	Coles et al. (2019)
CO	Li et al. (2015)
CO_2	Rothman et al. (2010)
CH_4	Yurchenko et al. (2017)

The spectrum…

The free parameters…

Table 3: Runtime of the retrievals with `emcee` and `MultiNest` samplers.

# Fit params	Emcee [s]	MultiNest [s]
4	16,465	5,423
5	6,550	7,505
6	15,247	10,093
7	14,930	12,847
8	30,151	17,097

Figures can be included like this:

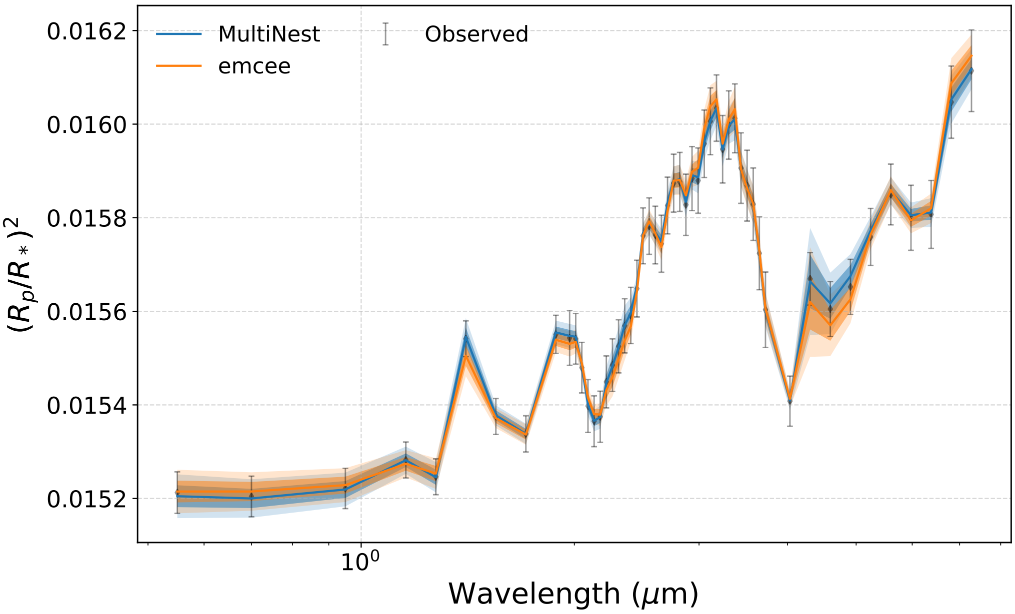

Figure 1: Caption for spectrum figure.

and referenced from text using Figure 1.

Figure sizes can be customized by adding an optional second parameter:

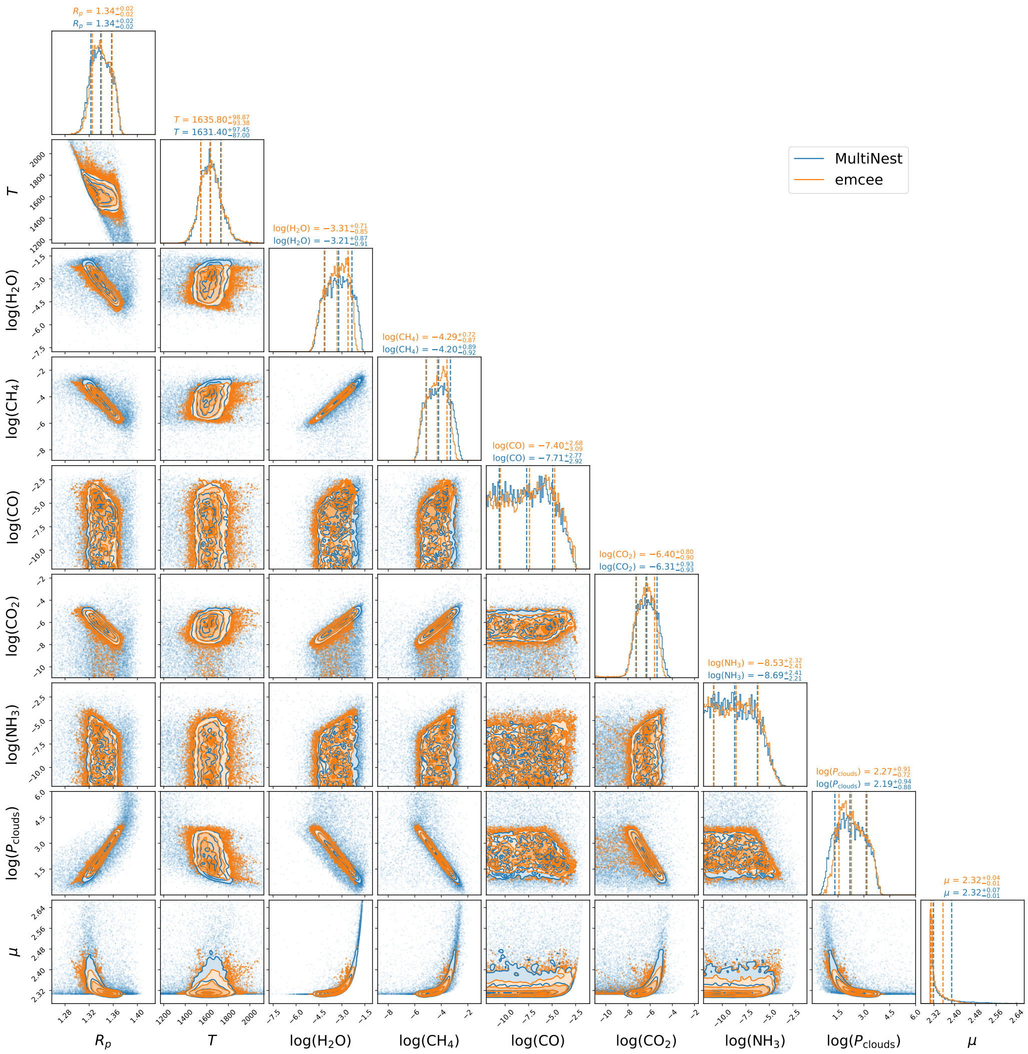

Figure 2: Caption for posteriors figure.

and referenced from text using Figure 2.

Statement of need

Optimized sampling methods are a key component of any retrieval code. Nested samplers (F. Feroz et al., 2009; Farhan Feroz et al., 2019) are a powerful and robust sampling method, successfully applied to the retrieval of exoplanet atmospheric spectra (Barstow et al., 2020; Bocchieri et al., 2023; Changeat et al., 2020). TauREx 3.1 natively implements a suite of nested samplers, including the MultiNest sampler, or makes them available as plugins, such as the UltraNest sampler. The primary target of nested samplers is the efficient calculation of the Bayesian `evidence`, whilst the inference of the `posterior` is a by-product. This is regarded as a key advantage of nested samplers, as the `evidence` can be readily used for model selection. However, the `evidence` is not always required, and the interpretation of the `posterior` from nested samplers necessitates some care. Additionally, algorithmic assumptions of nested samplers may require to tailor the priors to explore the parameter space thoroughly.

Where the inference of the Bayesian `posterior` is the primary target, a well-established alternative to nested samplers are a family of Markov chain Monte Carlo methods known as affine-invariant ensemble samplers (Goodman & Weare, 2010). The implementation in emcee (Foreman-Mackey et al., 2013) is a popular choice in the astronomy community, as it takes care of the heavy lifting of the sampling process, is well documented, and is straightforward to utilize. To date, the `emcee` sampler is not natively implemented in the TauREx 3.1 retrieval framework, nor elsewhere in other retrieval codes, to the knowledge of the authors. To fill this gap, we developed the `taurex-emcee` plugin, which interfaces the `emcee` sampler to TauREx. Key advantages of `taurex-emcee` are that it affords a more straightforward interpretation of the `posterior` and is more robust to the choice of priors. However, the current implementation is not intended to explore multimodal `posteriors`, for which nested samplers will inevitably be more efficient. Moreover, we caveat that it may require substantial computational time to sample high-dimensional parameter spaces with `emcee`, which can be mitigated by coupling `taurex-emcee` with TauREx's GPU-accelerated forward models (A. Al-Refaie et al., 2020).

Acknowledgements

This work was supported by the Italian Space Agency (ASI) with `Ariel` grant n. 2021.5.HH.0.

References

Abel, M., Frommhold, L., Li, X., & Hunt, K. L. C. (2011). Collision-induced absorption by H2 pairs: From hundreds to thousands of kelvin. *The Journal of Physical Chemistry A*, *115*(25), 6805–6812. https://doi.org/10.1021/jp109441f

Abel, M., Frommhold, L., Li, X., & Hunt, K. L. C. (2012). Infrared absorption by collisional H2-he complexes at temperatures up to 9000 k and frequencies from 0 to 20,000 cm(-1). *The Journal of Chemical Physics*, *136*(4), 044319. https://doi.org/10.1063/1.3676405

Al-Refaie, A. F., Changeat, Q., Waldmann, I. P., & al., et. (2021). TauREx 3: A Fast, Dynamic, and Extendable Framework for Retrievals. *917*(1), 37. https://doi.org/10.3847/1538-4357/ac0252

Al-Refaie, A., Changeat, Q., Venot, O., & al., et. (2020). TauREx 3.1 - Extending atmospheric retrieval with plugins. *European Planetary Science Congress*, EPSC2020–669. https://doi.org/10.5194/epsc2020-669

Barstow, J. K., Changeat, Q., Garland, R., & al., et. (2020). A comparison of exoplanet spectroscopic retrieval tools. *493*(4), 4884–4909. https://doi.org/10.1093/mnras/staa548

Bocchieri, A., Mugnai, L. V., Pascale, E., & al., et. (2023). Detecting molecules in Ariel low resolution transmission spectra. *Experimental Astronomy*, *56*(2-3), 605–644. https://doi.org/10.1007/s10686-023-09911-x

Changeat, Q., Al-Refaie, A., Mugnai, L. V., & al., et. (2020). Alfnoor: A Retrieval Simulation of the Ariel Target List. *160*(2), 80. https://doi.org/10.3847/1538-3881/ab9a53

Coles, P. A., Yurchenko, S. N., & Tennyson, J. (2019). ExoMol molecular line lists - XXXV. A rotation-vibration line list for hot ammonia. *490*(4), 4638–4647. https://doi.org/10.1093/mnras/stz2778

Edwards, B., Mugnai, L., Tinetti, G., & al., et. (2019). An Updated Study of Potential Targets for Ariel. *157*(6), 242. https://doi.org/10.3847/1538-3881/ab1cb9

Feroz, F., Hobson, M. P., & Bridges, M. (2009). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *398*(4), 1601–1614. https:

//doi.org/10.1111/j.1365-2966.2009.14548.x

Feroz, Farhan, Hobson, M. P., Cameron, E., & al., et. (2019). Importance Nested Sampling and the MultiNest Algorithm. *The Open Journal of Astrophysics*, *2*(1), 10. https://doi.org/10.21105/astro.1306.2144

Fletcher, L. N., Gustafsson, M., & Orton, G. S. (2018). Hydrogen Dimers in Giant-planet Infrared Spectra. *235*(1), 24. https://doi.org/10.3847/1538-4365/aaa07a

Foreman-Mackey, D., Hogg, D. W., Lang, D., & al., et. (2013). emcee: The MCMC Hammer. *125*(925), 306. https://doi.org/10.1086/670067

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Li, G., Gordon, I. E., Rothman, L. S., & al., et. (2015). Rovibrational Line Lists for Nine Isotopologues of the CO Molecule in the X $^1\Sigma^+$ Ground Electronic State. *216*(1), 15. https://doi.org/10.1088/0067-0049/216/1/15

Polyansky, O. L., Kyuberis, A. A., Zobov, N. F., & al., et. (2018). ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *480*(2), 2597–2608. https://doi.org/10.1093/mnras/sty1877

Rothman, L. S., Gordon, I. E., Barber, R. J., Dothe, H., Gamache, R. R., Goldman, A., Perevalov, V. I., Tashkun, S. A., & Tennyson, J. (2010). HITEMP, the high-temperature molecular spectroscopic database. *111*, 2139–2150. https://doi.org/10.1016/j.jqsrt.2010.05.001

Yurchenko, S. N., Amundsen, D. S., Tennyson, J., & al., et. (2017). A hybrid line list for CH_4 and hot methane continuum. *605*, A95. https://doi.org/10.1051/0004-6361/201731026