

# taurex-emcee: automated, parallelized atmospheric retrievals with TauREx 3.1 and the emcee sampler

**Andrea Bocchieri**  <sup>1</sup>, **Quentin Changeat**  <sup>2</sup>, **Lorenzo V. Mugnai**  <sup>3,4,5</sup>, and **Enzo Pascale**  <sup>1</sup>

**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

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

taurex-emcee is a plugin for the TauREx 3.1 atmospheric retrieval framework ([Al-Refaie et al., 2020, 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 MPI parallelization. Thus, the taurex-emcee plugin allows users to easily launch parallelized retrievals of atmospheric spectra with `emcee`. This enables reliable, efficient, and fast retrievals.

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 ([Feroz et al., 2009, 2019](#)), natively implemented in TauREx 3, on a synthetic transmission spectrum of a HD-209458 b-like hot Jupiter. 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 (128-core Intel(R) Xeon(R) Platinum 8358 CPU clocked at 2.60 GHz), equipped with one NVIDIA A40 GPU, and using the TauREx 3.1 GPU-accelerated forward model. The software versions used are: TauREx v3.1.4, taurex-cuda v1.0.1, MultiNest v3.10, and `emcee` v3.1.4. The high-resolution forward spectrum was generated assuming stellar and planetary parameters of HD-209458 b from Edwards et al. ([2019](#)), reported in [Table 1](#).

**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 <sub>K</sub>	$T_s$ [K]	$M_s$ [ $M_\odot$ ]
1.35	0.71	1613	3.52	1.18	6.31	6,086	1.18

The stellar spectrum is simulated with PHOENIX stellar models ([Husser et al., 2013](#)), and

the planetary atmosphere is gaseous, with hydrogen and helium at a ratio  $H_2/He = 0.172$ , and has an isothermal temperature profile. We consider five molecular species as atmospheric 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. We utilize cross sections at a resolution of 15,000 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. We also include fully-opaque gray clouds with a cloud-top pressure of 1000 Pa. Finally, 100 pressure layers are used to sample the atmosphere, from 1 Pa to  $10^6$  Pa, uniformly in log-space.

**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)

To simulate the observation of the transmission spectrum, we assume an observation by the Ariel space mission ([Tinetti et al., 2021, 2018](#)) in Tier 2 mode ([Edwards et al., 2019](#)). We utilize radiometric estimates of the total noise on one observation of HD-209458 b obtained with ArielRad ([Mugnai et al., 2020](#)), the Ariel radiometric simulator. ArielRad is based on the generic point source radiometric model ExoRad2 ([Mugnai et al., 2023](#)), adapted with the Ariel payload configuration. The software versions used are: ExoRad2 v2.1.113, ArielRad-payloads v0.0.17, and ArielRad v2.4.26. Furthermore, we fix the data points of the observed spectrum to the expected values, i.e. we do not scatter the spectrum with the noise estimates; this is done to ensure that the results of the retrievals are not affected by the realization of the noise.

In our retrieval benchmarks, we attempt to constrain the abundances of the trace gases, alongside the temperature profile, the planetary radius, and the cloud-top pressure. [Table 3](#) lists each parameter, its units, and the priors (with corresponding scale) used in the retrievals.

**Table 3:** Parameters and priors used in the retrievals.

Parameter	Unit	Prior	Scale
$R_P$	$R_J$	$\pm 10\%$	linear
$T$	K	100; 4000	linear
$H_2O$	VMR	$10^{-12}; 10^{-1}$	log
$CH_4$	VMR	$10^{-12}; 10^{-1}$	log
$CO$	VMR	$10^{-12}; 10^{-1}$	log
$CO_2$	VMR	$10^{-12}; 10^{-1}$	log
$NH_3$	VMR	$10^{-12}; 10^{-1}$	log
$P_{\text{clouds}}$	Pa	1; $10^6$	log

We perform five retrievals each with MultiNest and emcee, with increasing number of molecules included in the free parameters (the other molecules being fixed to their true values), from only  $H_2O$  to all five trace gases. With MultiNest, we set the evidence tolerance to 0.5 and sample the parameter space through 1500 live points. With emcee, we utilize 100 walkers and run two independent chains, each to convergence, where we adopt the default convergence criteria of the [autoemcee](#) package, i.e. the Geweke convergence diagnostic ([Geweke, 1991](#)) is  $z\text{-score} < 2.0$  and the Gelman-Rubin rank diagnostic ([Gelman & Rubin, 1992](#)) is  $\hat{r} < 1.01$ .

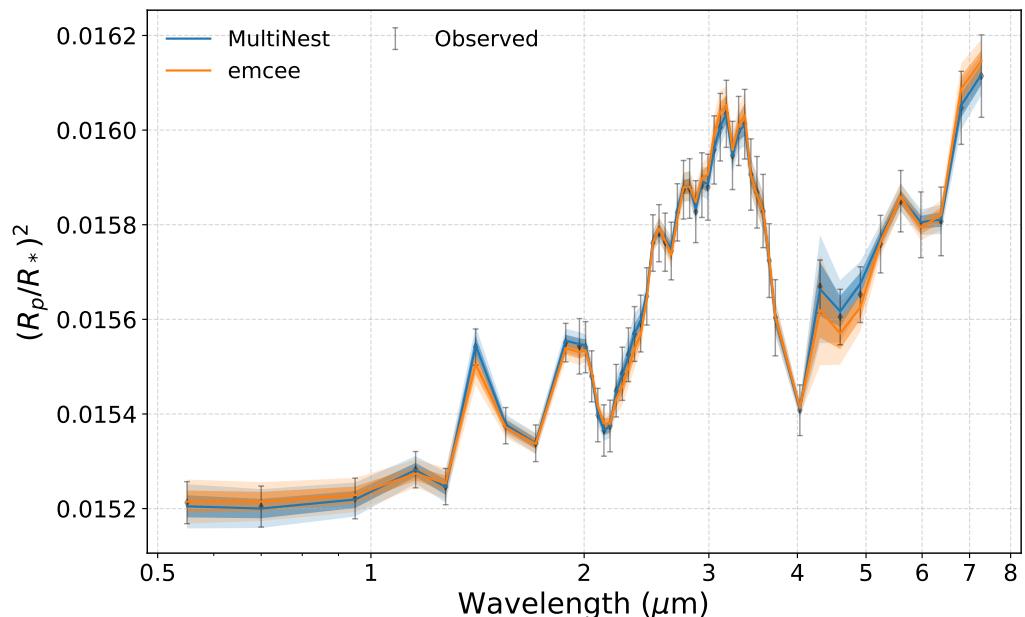
When iterating the chains, we increase the number of steps by multiplying the number of steps of the previous iteration by a growth factor of 2, a parameter that we have enabled as the default value of 10 was deemed too conservative for our tests. As a first comparison, we report in [Table 4](#) the computational time of the retrievals with emcee and MultiNest, alongside the number of samples of each retrieval. Note that, for emcee, the number of samples reported is per chain.

**Table 4:** Sampling times and number of samples of the retrievals with each sampler.

Fitted molecules	emcee [s]	MultiNest [s]	emcee [#]	MultiNest [#]
H <sub>2</sub> O	6,300	5,400	20,000	18,800
H <sub>2</sub> O, CH <sub>4</sub>	6,600	7,500	20,000	22,000
H <sub>2</sub> O, CH <sub>4</sub> , CO	15,200	10,100	40,000	21,900
H <sub>2</sub> O, CH <sub>4</sub> , CO, CO <sub>2</sub>	14,900	12,800	40,000	23,100
H <sub>2</sub> O, CH <sub>4</sub> , CO, CO <sub>2</sub> , NH <sub>3</sub>	30,200	17,100	80,000	23,200

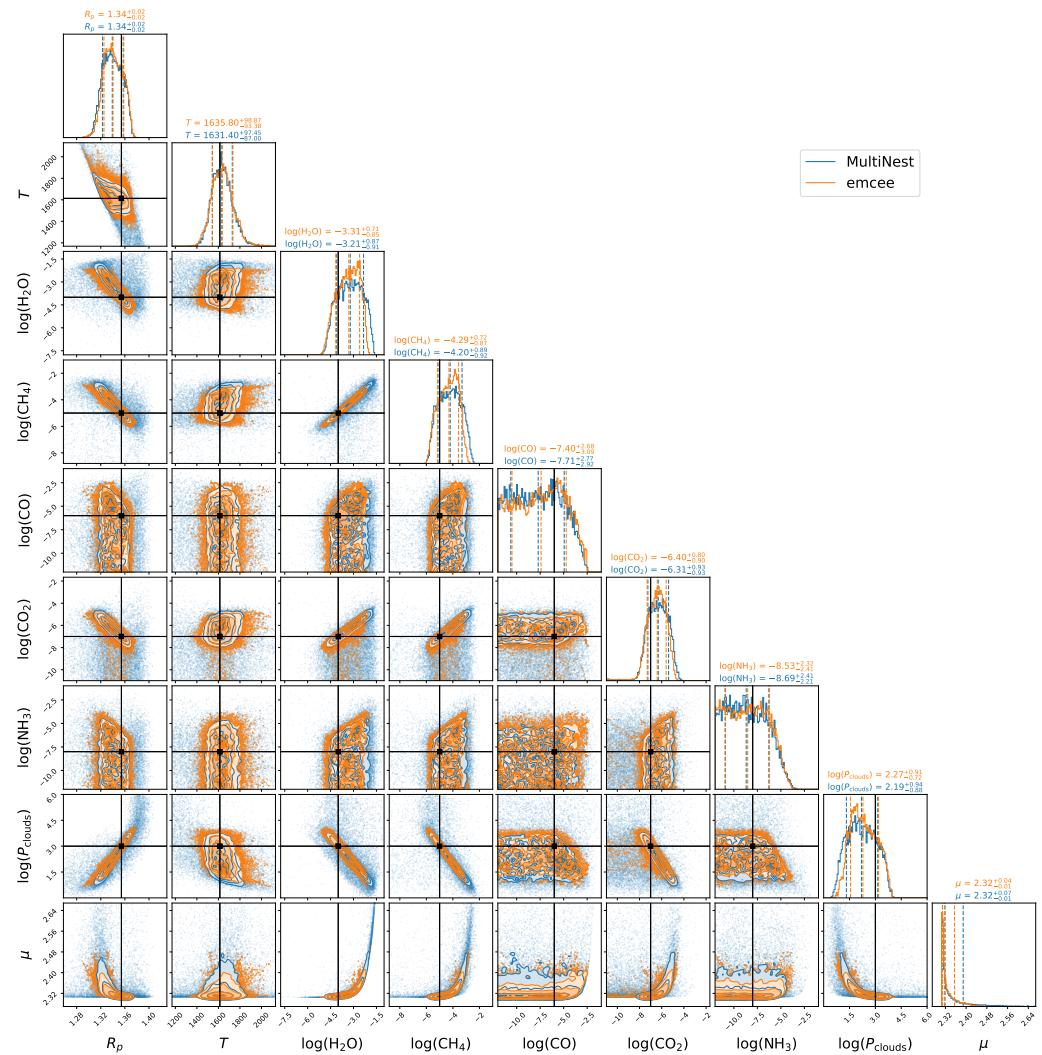
MultiNest is faster than emcee for all retrievals except the one with H<sub>2</sub>O and CH<sub>4</sub>, where the sampling time is around 15 minutes longer. The difference in speed correlates with the number of samples, which is higher for emcee than MultiNest for all retrievals. Additionally, sampling times with MultiNest scale slower with increasing dimensionality of the parameter space.

Next, we compare the results of the retrievals. For simplicity, we only discuss the retrieval with the full stack of trace gases included in the free parameters, as the results of the other retrievals are consistent. [Figure 1](#) shows the retrieved spectrum yielded by the two samplers, alongside the observed spectrum. The retrieved spectrum is shown as the best-fit model and the 1 $\sigma$  and 2 $\sigma$  confidence intervals. The retrieved spectra are consistent with each other and with the observed spectrum within the experimental uncertainties.



**Figure 1:** Best-fit spectra for the HD-209458 b retrievals with MultiNest (blue) and emcee (orange). The synthetic observed spectrum is shown as the black error bars. The 1 $\sigma$  and 2 $\sigma$  confidence intervals are shown as the shaded regions. The error bars of the observed spectrum are generated with ArielRad ([Mugnai et al., 2020](#)) and the spectral grid corresponds to the Ariel Tier 2 mode ([Edwards et al., 2019](#)).

**Figure 2** shows the posteriors of the retrieved parameters, and **Table 5** reports the median and 16% and 84% quantiles of the marginalized posteriors relative to the median.



**Figure 2:** Posterior distributions of the retrieved parameters for the HD-209458 b simulated observations with MultiNest (blue) and emcee (orange). The true values are shown as the black lines. The vertical dashed lines in the histograms on the diagonal show the median and 16% and 84% quantiles.

**Table 5:** Retrieved parameters and their uncertainties. True values are reported for comparison.

Parameter	True value	emcee	MultiNest
$R_p$	1.35	$1.34^{+0.02}_{-0.02}$	$1.34^{+0.02}_{-0.02}$
T	1613	$1635.8^{+98.9}_{-93.4}$	$1631.4^{+97.5}_{-87.0}$
$\log(H_2O)$	-4	$-3.31^{+0.71}_{-0.85}$	$-3.21^{+0.87}_{-0.91}$
$\log(CH_4)$	-5	$-4.29^{+0.72}_{-0.87}$	$-4.20^{+0.89}_{-0.92}$
$\log(CO)$	-6	$< -3.6$	$< -3.8$
$\log(CO_2)$	-7	$-6.40^{+0.80}_{-0.90}$	$-6.31^{+0.93}_{-0.93}$
$\log(NH_3)$	-8	$< -5.2$	$< -5.2$
$\log(P_{clouds})$	3	$2.27^{+0.91}_{-0.72}$	$2.19^{+0.94}_{-0.88}$
$\mu$		$\mu = 2.32^{+0.04}_{-0.04}$	$\mu = 2.32^{+0.04}_{-0.04}$

The similarity of the results from the two samplers is reassuring, and the retrieved parameters are consistent with the true values within  $1\sigma$ . It is worth noting that the median values are, to an extent, biased for both samplers, in essentially the same manner. This result is expected, as the retrieval traces the degeneracies between the parameters and the fitted molecules have strong correlations with each other, as seen in Figure 2. In addition, for some parameters, namely CO and NH<sub>3</sub>, due to the combination of opacities and abundances, the retrievals only recover an upper limit. In these cases, mean, median, and mode are not defined, and the reported values depend on the choice of the prior. Therefore, for CO and NH<sub>3</sub>, Table 5 reports the 95% upper limit, defined as the cumulative of the marginalized posterior at the 95% quantile.

In summary, while the emcee sampler is generally slower than MultiNest, it is a robust and reliable alternative to nested samplers, as shown by the consistency of the results from the two samplers in our benchmark.

## Statement of need

Optimized sampling methods are a key component of any retrieval code. Nested samplers (Buchner, 2023) 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., 2022, 2020; Mugnai et al., 2021; Yassin Jaziri et al., 2024). 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 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 are more efficient but also require extra caution in interpreting the results. 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. To conclude, while taurex-emcee is a new and powerful tool for the exoplanet community, comparison of various statistical inference strategies, such as MultiNest, UltraNest, Variational Inference (Yip et al., 2024), and now emcee, is necessary to ensure un-biased estimates of parameter space in exo-atmospheric studies with HST, JWST, and Ariel.

## 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 H<sub>2</sub> 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 H<sub>2</sub>-he complexes at temperatures up to 9000 k and frequencies from 0 to 20,000 cm<sup>-1</sup>. *The Journal of Chemical Physics*, 136(4), 044319. <https://doi.org/10.1063/1.3676405>
- Al-Refaie, A. F., 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>
- 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>
- 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>
- Buchner, J. (2023). Nested Sampling Methods. *Statistics Surveys*, 17, 169–215. <https://doi.org/10.1214/23-SS144>
- 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>
- Changeat, Q., Edwards, B., Al-Refaie, A. F., & al., et. (2022). Five Key Exoplanet Questions Answered via the Analysis of 25 Hot-Jupiter Atmospheres in Eclipse. 260(1), 3. <https://doi.org/10.3847/1538-4365/ac5cc2>
- 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, F., 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>
- Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7, 457–472. <https://doi.org/10.1214/ss/1177011136>
- Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* (Staff Report No. 148). Federal Reserve Bank of Minneapolis. <https://ideas.repec.org/p/fip/fedmsr/148.html>

- 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>
- Husser, T.-O., Wende-von Berg, S., Dreizler, S., & al., et. (2013). A new extensive library of PHOENIX stellar atmospheres and synthetic spectra. 553, A6. <https://doi.org/10.1051/0004-6361/201219058>
- 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>
- Mugnai, L. V., Bocchieri, A., & Pascale, E. (2023). ExoRad 2.0: The generic point source radiometric model. *The Journal of Open Source Software*, 8(89), 5348. <https://doi.org/10.21105/joss.05348>
- Mugnai, L. V., Modirrousta-Galian, D., Edwards, B., & al., et. (2021). ARES. V. No Evidence For Molecular Absorption in the HST WFC3 Spectrum of GJ 1132 b. 161(6), 284. <https://doi.org/10.3847/1538-3881/abf3c3>
- Mugnai, L. V., Pascale, E., Edwards, B., & al., et. (2020). ArielRad: the Ariel radiometric model. *Experimental Astronomy*, 50(2-3), 303–328. <https://doi.org/10.1007/s10686-020-09676-7>
- 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>
- Tinetti, G., Drossart, P., Eccleston, P., & al., et. (2018). A chemical survey of exoplanets with ARIEL. *Experimental Astronomy*, 46(1), 135–209. <https://doi.org/10.1007/s10686-018-9598-x>
- Tinetti, G., Eccleston, P., Haswell, C., & al., et. (2021). Ariel: Enabling planetary science across light-years. *arXiv e-Prints*, arXiv:2104.04824. <https://doi.org/10.48550/arXiv.2104.04824>
- Yassin Jaziri, A., Pluriel, W., Bocchieri, A., & al., et. (2024). ARES VI: Viability of one-dimensional retrieval models for transmission spectroscopy characterization of exo-atmospheres in the era of JWST and Ariel. *arXiv e-Prints*, arXiv:2401.03809. <https://doi.org/10.48550/arXiv.2401.03809>
- Yip, K. H., Changeat, Q., Al-Refaie, A., & al., et. (2024). To Sample or Not to Sample: Retrieving Exoplanetary Spectra with Variational Inference and Normalizing Flows. 961(1), 30. <https://doi.org/10.3847/1538-4357/ad063f>
- Yurchenko, S. N., Amundsen, D. S., Tennyson, J., & al., et. (2017). A hybrid line list for CH<sub>4</sub> and hot methane continuum. 605, A95. <https://doi.org/10.1051/0004-6361/201731026>