The Journal of Open Source Software

# rojak: A Python library and tool for aviation turbulence diagnostics

**Hui Ling Wong** [1,2], **Rafael Palacios** [1], and **Edward Gryspeerdt** [2]

1 Department of Aeronautics, Imperial College London, United Kingdom 2 Department of Physics, Imperial College London, United Kingdom

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Editor:** James Gaboardi

**Reviewers:**

- @xoolive
- @danikam

**Submitted:** 21 July 2025
**Published:** unpublished



## Summary

Aviation turbulence is atmospheric turbulence occurring at length scales large enough (approximately 100m to 1km) to affect an aircraft (Sharman, 2016). According to the National Transport Safety Board (NTSB), turbulence experienced whilst onboard an aircraft was the leading cause of accidents from 2009 to 2018 (NTSB, 2021). Clear air turbulence (CAT) is a form of aviation turbulence which cannot be detected by the onboard weather radar. Thus, pilots are unable to preemptively avoid such regions. In order to mitigate this safety risk, CAT diagnostics are used to forecast turbulent regions such that pilots are able to tactically avoid them.

`rojak` is a parallelised python library and command-line tool for using meteorological data to forecast CAT and evaluating the effectiveness of CAT diagnostics against turbulence observations. Currently, it supports,

1. Computing turbulence diagnostics on meteorological data from European Centre for Medium-Range Weather Forecasts's (ECMWF) ERA5 reanalysis on pressure levels (Hersbach, 2023). Moreover, it is easily extendable to support other types of meteorological data.
2. Retrieving and processing turbulence observations from Aircraft Meteorological Data Relay (AMDAR) data archived at National Oceanic and Atmospheric Administration (NOAA)(*NCEP Meteorological Assimilation Data Ingest System (MADIS)*, 2024) and AMDAR data collected via the Met Office MetDB system (Met Office, 2008)
3. Computing 27 different turbulence diagnostics, such as the three-dimensional frontogenesis equation (Bluestein, 1993), turbulence index 1 and 2 (Ellrod & Knapp, 1992), negative vorticity advection (Sharman et al., 2006), and Brown's Richardson tendency equation (Brown, 1973).
4. Converting turbulence diagnostic values into the eddy dissipation rate (EDR) — the International Civil Aviation Organization's (ICAO) official metric for reporting turbulence (*Meteorological Service for International Air Navigation*, 2010)

These features not only allow users to perform operational forecasting of CAT but also to interrogate the intensification in frequency and severity of CAT due to climate change (Kim et al., 2023; Storer et al., 2017; Williams, 2017), such as by analysing the climatological distribution of the probability of encountering turbulence at different severities (e.g. light turbulence or moderate-or-greater turbulence) for each turbulence diagnostic. These applications involve high-volume datasets, ranging from tens to hundreds of gigabytes, necessitating the use of parallelisation to preserve computational tractability and efficiency, while substantially reducing execution time. As such, `rojak` leverages `Dask` to process larger-than-memory data and to run in a distributed manner (Dask Development Team, 2016).

The name of the package, `rojak`, is inspired by its wide range turbulence diagnostics and its applications. While *rojak* refers to a type of salad, it is also a colloquial term in Malaysia and

Singapore for an eclectic mix, reflecting the diverse functionality of the package.

## Statement of need

Numerous studies have investigated the influence of the climate on CAT (e.g. Kim et al., 2023; Storer et al., 2017; Williams, 2017) and the application of turbulence diagnostics for operational forecasting (e.g. Gill, 2014; Gill & Buchanan, 2014; Sharman et al., 2006; Sharman & Pearson, 2017) in products like the Federal Aviation Authority's (FAA) Graphical Turbulence Guidance and the International Civil Aviation Organization's (ICAO) World Area Forecast System. However, to the best of the author's knowledge, none of these studies have made their code publicly available. This work presents the first open-source package for aviation turbulence analysis. Given the inherent complexity of CAT diagnostics and the variability in how these could be implemented, `rojak` serves as a first iteration of a standardised implementation of these CAT diagnostics, providing a basis for future enhancements and refinements by the broader research community. Moreover, the parallelised nature of `rojak` and its architecture, which keeps it open to extensions, positions it as as an indispensable resource to bridging this gap.

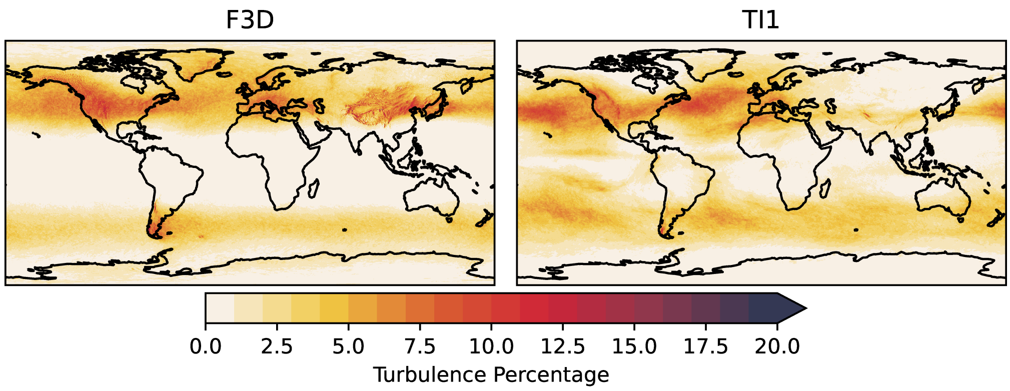


**Figure 1:** Probability of encountering light turbulence during the months December, January, February from 2018 to 2024 at 200 hPa for the three-dimensional frontogenesis (F3D) and turbulence index 1 (TI1) diagnostics

Figure 1 demonstrates the application of `rojak` for characterising CAT's response to climate change. Depicted in the figure is the global climatological distribution of the probability of encountering light turbulence for the boreal winter months (i.e., December, January and February) from 2018 to 2024 at 200 hPa based on the two turbulence diagnostics — the three-dimensional frontogenesis equation and turbulence index 1. This was computed using ERA5 data at 6-hourly intervals with three pressure levels (175 hPa, 200 hPa and 225 hPa) for the aforementioned time period. This required processing 85GB of ERA5 data. The methodology employed by `rojak` for determining the presence of turbulence and the equations for the various turbulence diagnostics is derived from the existing aviation meteorology literature on turbulence. In this instance, it is an implementation of the methodology described in Williams & Storer (2022).

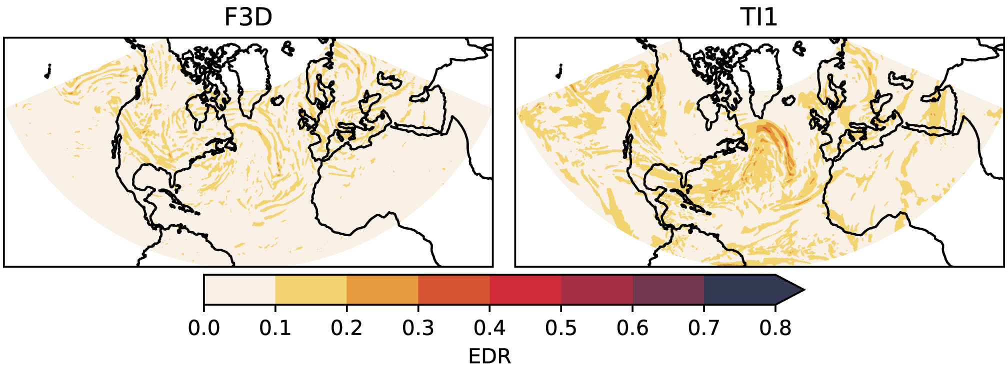


**Figure 2:** 6-hour forecast of eddy dissipation rate (EDR) at 200 hPa for the three-dimensional frontogenesis (F3D) and turbulence index 1 (TI1) on the 1st of December 2024 at 00:00

Similarly, Figure 2 demonstrates the application of `rojak` for operational turbulence forecasting. By employing the methodology described in Sharman & Pearson (2017), `rojak` is able to convert three-dimensional frontogenesis and turbulence index 1 diagnostics into EDR. Figure 2 shows the 6-hour turbulence forecast on the 1st of December 2024 at 00:00 GMT which can be used for flight trajectory planning to avoid turbulent regions. The full range of features, including the details and references, are contained within the documentation of `rojak`.

In the context of operational forecasting (e.g. methods detailed in Gill, 2014; Sharman et al., 2006; and Pearson & Sharman, 2017), the comparison of turbulence diagnostics computed from meteorological data against observational data is a fundamental component. The statistical nature of using an ensemble of turbulence diagnostics which has an optimal balance of a low false positive and false negative rate mandates it. The architecture of `rojak` enables it to seamlessly integrate various sources of meteorological and observational data, with their interactions managed through a central mediator. `rojak` also contains a command-line interface tool which can be launched to perform a variety of aviation turbulence analyses and to retrieve the meteorological and observation data from the various data providers.

In terms of the calculations performed upon the meteorological data, `MetPy` (May et al., 2016) has the greatest similarity to `rojak`. However, it does not natively support `Dask` (Manser, 2020). Given the size of the datasets to be processed, this presented a significant issue. Moreover, `MeyPy` does not implement the calculations required by the turbulence diagnostics.



# Acknowledgements

We gratefully acknowledge the Brahmal Vasudevan Institute for Sustainable Aviation at Imperial College London for funding this research and to the maintainers of the various scientific Python packages which `rojak` depends upon, such as `NumPy` (Harris et al., 2020), `xarray` (Hoyer & Hamman, 2017), `Pandas` (The pandas development team, 2025), `GeoPandas` (Jordahl et al., 2020), and `SciPy` (Virtanen et al., 2020).

# References

Bluestein, H. B. (1993). *Synoptic-dynamic Meteorology in Midlatitudes: Observations and theory of weather systems: Vols. Observations and theory of weather systems*. Taylor & Francis. ISBN: 978-0-19-506268-7

Brown, R. (1973). New indices to locate clear-air turbulence. *Meteorological Magazine (UK)*, *102*, 347–361.

Dask Development Team. (2016). *Dask: Library for dynamic task scheduling*. http://dask.pydata.org

Ellrod, G. P., & Knapp, D. I. (1992). An Objective Clear-Air Turbulence Forecasting Technique: Verification and Operational Use. *Weather and Forecasting*, *7*(1), 150–165. https://doi.org/10.1175/1520-0434(1992)007%3C0150:AOCATF%3E2.0.CO;2

Gill, P. G. (2014). Objective verification of World Area Forecast Centre clear air turbulence forecasts. *Meteorological Applications*, *21*(1), 3–11. https://doi.org/10.1002/met.1288

Gill, P. G., & Buchanan, P. (2014). An ensemble based turbulence forecasting system. *Meteorological Applications*, *21*(1), 12–19. https://doi.org/10.1002/met.1373

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hersbach, B., H. (2023). *ERA5 hourly data on pressure levels from 1940 to present. Copernicus climate change service (C3S) climate data store (CDS).*

Hoyer, S., & Hamman, J. (2017). Xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, *5*(1). https://doi.org/10.5334/jors.148

Jordahl, K., Bossche, J. V. den, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *Geopandas/geopandas: v0.8.1* (Version v0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Kim, S.-H., Kim, J.-H., Chun, H.-Y., & Sharman, R. D. (2023). Global response of upper-level aviation turbulence from various sources to climate change. *Npj Climate and Atmospheric Science*, *6*(1), 1–12. https://doi.org/10.1038/s41612-023-00421-3

Manser, R. (2020). Support for Dask Arrays · Issue #1479 · Unidata/MetPy. In *GitHub*. https://github.com/Unidata/MetPy/issues/1479.

May, R., Arms, S., Marsh, P., Bruning, E., Leeman, J., Bruick, Z., & Camron, M. D. (2016). *MetPy: A Python Package for Meteorological Data*. UCAR/NCAR - Unidata. https://doi.org/10.5065/D6WW7G29

Met Office. (2008). *AMDAR (aircraft meteorological data relay) reports collected by the met office MetDB system*. https://catalogue.ceda.ac.uk/uuid/33f44351f9ceb09c495b8cef74860726; NCAS British Atmospheric Data Centre (NCAS BADC).

*Meteorological service for international air navigation: Annex 3 to the convention on international civil aviation*. (2010). http://digitallibrary.un.org/record/448084

*NCEP Meteorological Assimilation Data Ingest System (MADIS)*. (2024). https://amdar.noaa.gov/.

NTSB. (2021). *Preventing Turbulence-Related Injuries in Air Carrier Operations Conducted Under Title 14 Code of Federal Regulations Part 121* (Safety Research Report SS-21-01). National Transport Safety Board (NTSB).

Pearson, J. M., & Sharman, R. D. (2017). Prediction of Energy Dissipation Rates for Aviation Turbulence. Part II: Nowcasting Convective and Nonconvective Turbulence. *Journal of Applied Meteorology and Climatology*, *56*(2), 339–351. https://doi.org/10.1175/JAMC-D-16-0312.1

Sharman, R. D. (2016). Nature of Aviation Turbulence. In *Aviation Turbulence: Processes, Detection, Prediction* (pp. 3–30). Springer International Publishing AG. https://doi.org/

10.1007/978-3-319-23630-8_1

Sharman, R. D., & Pearson, J. M. (2017). Prediction of Energy Dissipation Rates for Aviation Turbulence. Part I: Forecasting Nonconvective Turbulence. *Journal of Applied Meteorology and Climatology*, *56*(2), 317–337. https://doi.org/10.1175/JAMC-D-16-0205.1

Sharman, R. D., Tebaldi, C., Wiener, G., & Wolff, J. (2006). An Integrated Approach to Mid- and Upper-Level Turbulence Forecasting. *Weather and Forecasting*, *21*(3), 268–287. https://doi.org/10.1175/WAF924.1

Storer, L. N., Williams, P. D., & Joshi, M. M. (2017). Global Response of Clear-Air Turbulence to Climate Change. *Geophysical Research Letters*, *44*(19), 9976–9984. https://doi.org/10.1002/2017GL074618

The pandas development team. (2025). *Pandas-dev/pandas: pandas* (Version v2.3.0). Zenodo. https://doi.org/10.5281/zenodo.15597513

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Williams, P. D. (2017). Increased light, moderate, and severe clear-air turbulence in response to climate change. *Advances in Atmospheric Sciences*, *34*(5), 576–586. https://doi.org/10.1007/s00376-017-6268-2

Williams, P. D., & Storer, L. N. (2022). Can a climate model successfully diagnose clear-air turbulence and its response to climate change? *Quarterly Journal of the Royal Meteorological Society*, *148*(744), 1424–1438. https://doi.org/10.1002/qj.4270