



University
of Exeter

Open Research Awards 2024

ESE disciplines prize

Liam Berrisford

Title of case study

Operationalizing and Democratizing Ambient Air Pollution Data and Analytics

Introduction to the case study

We have developed an open-access platform that democratizes ambient air pollution data and predictive analytics. Our work leverages open data and reproducible research practices to provide data and predictive analytics information to researchers, policymakers, and the public. Importantly, it fosters a collaborative approach, giving everyone a say in tackling air quality issues. Before the release of this research, predictive analytics were firmly held within the realm of expert stakeholders and are now easily accessible by a layperson audience. Three research papers provide an overview of the high-resolution England dataset (1km² hourly) [1], the global dataset (0.25 degree hourly) [2] and the Python package [3].

Detail the practices you have used to support openness/reproducibility

Publishing Pre-prints: All our findings and methodologies were promptly shared through pre-print platforms, ensuring immediate access to our research outputs [1,2,3].

Open Data and Code: We made all our datasets and analysis code freely available on public repositories (e.g., GitHub and Google Drive as a stopgap [4]). Once the papers have passed through the peer review process, the datasets will be uploaded to Dryad, and the code will be sent to Zenodo to enable others to replicate our studies and build upon our work [7].

I have also tried to engage with communities that don't typically engage with this sort of research, such as during Pint of Science [8] and the British Science Festival [9].

Outline any barriers or challenges you encountered, and how they were handled

The primary barrier was managing the vast amount of data produced, approximately 300GB of air pollution data model output and around 1TB of model input. Without dedicated funding for data hosting, I strategically focused on hosting critical air pollution output data and a subset of model input data to ensure reproducibility. Additionally, I aimed to publish in venues like the Royal Society Open Science, which covers hosting costs. This approach balanced accessibility and practical constraints, enabling the continued sharing of essential research data.

What benefits have you identified from your case?

The openness of our project has started to foster some further work; for example, the Met Office is beginning to work on a machine learning project within the air quality team based on the work of the England model. The Environment Agency has also expressed that they have begun looking at the Python package for use within some of their work. Further, the use of GitHub to host the

package has helped to improve visibility for some individuals starring the project even though we have never interacted directly with the individuals. Other organizations, such as Exeter Science Center, have also expressed interest in the work, alongside using platforms like Pint Of Science to advertise the work.

How do you plan to continue this case moving forward?

Moving forward, I wish to make the work more accessible and interactive. Currently, to engage with the work, you need to know how to use Python and have some programming ability. In the future, I plan to build a website interface for the package that enables users to make their own air pollution predictions and access data via a custom user interface, helping to ensure everyone is able to engage with the research conducted. I hope this will help build a community around the work and enable more individuals to engage with air pollution research.

Conclusions

Our commitment to open and reproducible research has significantly enhanced the impact and reach of our work on ambient air pollution. By fostering openness and collaboration, we aim to drive further advancements in understanding and addressing air quality issues, engaging individuals who have historically been unheard from in discussions surrounding air pollution.

Links to any materials/resources relevant to the case

[1] ArXiv Link for paper 1: <https://arxiv.org/abs/2401.08735>

[2] ArXiv Link for paper 2: <https://arxiv.org/abs/2402.10248>

[3] ArXiv Link for paper 3: <https://arxiv.org/abs/2403.03664>

[4] Project GitHub Repo (paper 3): <https://github.com/berrli/Environmental-Insights>

[5] Interactive Dashboard for Paper 1 Data: https://berrli.github.io/England-Ambient-air-Pollution-Model-Dashboard/england_AIUK_2024_air_pollution_demo.html

[6] Interactive Dashboard for Paper 2 Data: https://berrli.github.io/Global-Ambient-air-Pollution-Model-Dashboard/global_AIUK_2024_air_pollution_demo.html

[7] Example Private Open Dataset and Code for Peer Review for one research project: <https://datadryad.org/stash/share/fUrhJytdndI6xB-h1GjSxqelb-st4udQk4IHrxXYCDU>

[8] British Science Festival Twitter Post: <https://x.com/liberrisford/status/1700490538628022395>

[9] Pint of Science Game Repo: <https://github.com/berrli/InteractiveAirPollutionDecisionTree>

Postscript, 10th February 2026

Open Research Award Impact: Environmental Insights

The Open Research Award was awarded to me during the latter stages of my PhD, at a time when several of the planned open research outputs were still in development. When the award was received, a number of the associated research papers remained under peer review. Since then, all publications have successfully passed peer review, and the broader programme of work, including datasets, software infrastructure, and an interactive platform, has been completed and released. This update summarises the progress made following the award in 2024 and outlines the subsequent work undertaken in alignment with open research principles.

The research focused on developing scalable machine-learning approaches to estimate hourly ambient air pollution concentrations across space and time. A central objective, consistent with the ethos of the Open Research Award, was to ensure that the work extended beyond traditional academic outputs and instead established accessible, reusable research infrastructure.

Open Datasets

Three major datasets were developed and published:

- **ML-HAPPG (Global)** – Hourly global air pollution estimates for 2022 at $0.25^\circ \times 0.25^\circ$ spatial resolution.
CEDA Dataset Record: <https://dx.doi.org/10.5285/7f91b1326a324caa9e436b8fdef4a0d8>
- **ML-HAPPE (England)** – Hourly high-resolution (1 km \times 1 km) air pollution estimates across England for 2018.
CEDA Dataset
Record: <https://catalogue.ceda.ac.uk/uuid/fc735f9878ed43e293b85f85e40df24d>
- **SynthHAPPE (England)** – Synthetic “typical day” hourly pollution profiles derived from multi-year observations (2014–2018), representing characteristic patterns rather than a single historical year.
CEDA Dataset
Record: <https://catalogue.ceda.ac.uk/uuid/fe877f3035c042478fd67de21f5f445a/>

These datasets were designed to be openly accessible, fully documented, and suitable for reuse across research, policy, and applied environmental contexts.

Environmental Insights: From Research to Reusable Tools

To support reproducibility and facilitate reuse, I developed the Environmental Insights Python package:

GitHub Repository:

<https://github.com/liamjberrisford/Environmental-Insights>

Documentation:

https://liamjberrisford.github.io/Environmental-Insights/home_page.html

The package enables users to retrieve historical air pollution data, run trained machine-learning models, and analyse spatial and temporal pollution patterns in a structured and reproducible manner. It was intentionally designed to be accessible to a broad audience, including academic researchers, public sector organisations, and non-governmental bodies, thereby lowering the technical barrier to engaging with high-resolution environmental data.

Environmental Insights Interactive

To further extend accessibility beyond programmatic users, the modelling framework was translated into a fully featured web application:

<https://liamjberrisford.github.io/Environmental-Insights-Interactive/>

Impact of the Open Research Awards

I greatly valued the opportunity to participate in the Open Research Awards. The process provided space to reflect critically on my own research practice and played a formative role in shaping the subsequent direction of the project. In particular, it helped crystallise a longer-term vision for translating methodological research into durable, openly accessible infrastructure, ultimately leading to the datasets, software package, and interactive platform described above. Engaging with other award recipients across disciplines was especially enriching; understanding how colleagues in different fields conceptualise openness, reproducibility, and impact offered perspectives that I have since brought back into my work within computer science and environmental modelling.

Find out more about the [University of Exeter Open Research Awards](#)