https://doi.org/10.1038/s43586-023-00202-5

Carbon footprint estimation for computational research

Loïc Lannelongue & Michael Inouye

Data analysis relies heavily on computation, and algorithms have grown more demanding in terms of hardware and energy. Monitoring their environmental impacts is and will continue to be an essential part of sustainable research. Here, we provide guidance on how to do so accurately and with limited overheads.

The urgency for environmentally sustainable research practices has never been greater. The gap between the carbon footprint of scientists (approximately 4–25 tonnes of CO_2-equivalent (CO_2e) per year per scientist[1,2]) and the Intergovernmental Panel on Climate Change carbon budget per person to keep global warming below 1.5 °C (<2 tonnes of CO_2e per year per person[3]) is substantial.

Estimation of the carbon footprint of a computational project is best done by the researchers themselves. The main source of greenhouse gas (GHG) emissions is the power draw of computers during the analyses. Carbon footprints are generally measured in grams of CO_2e (gCO_2e), which summarizes the global warming impact of a range of GHGs. The carbon footprint of a task depends on the total energy used by the computer and on the amount of GHGs emitted to produce this energy, called carbon intensity. Carbon intensities mostly depend on the sources of electricity production, and various resources make this data available, such as ElectricityMap. Estimating the energy consumption of a computer or a particular computation is a trade-off between accuracy and practicality, and it has been shown that accurate estimates can be obtained by focusing on the power draw of processing cores (CPUs and/or GPUs) and the quantity of memory available. The other relevant parameters are the runtime and the efficiency of the computing facility (measured by its power usage effectiveness, or PUE). These components have been combined to estimate overall energy consumption[4].

There are different strategies for estimating carbon footprints in practice. Depending on the strategy, some metrics mentioned above can be collected automatically. If computations are performed on a local computer (like a laptop or desktop computer), then the tools need to be user-side, meaning that data collection happens on the user's device. In this case, one can use an online calculator, such as GreenAlgorithms[4], or embed a software, package or library in their code to track computing usage to then produce carbon footprint estimates, like CarbonTracker[5] and CodeCarbon. When using a high performance computing (HPC) facility, where algorithms are run in a dedicated data centre or on the cloud, details about the computations are often logged for accounting purposes. In this case, there is also the possibility of using a server-side tool – like GreenAlgorithms4HPC[6] – that estimates carbon footprints a posteriori based on server logs.

Each strategy offers pros and cons (Table 1). The online calculator is the most flexible method, as it works for any hardware, programming language and field of research. It also enables estimations of carbon footprints a priori for planning and a posteriori to investigate previous computations. However, it can be cumbersome for large numbers of jobs, and it requires the user to collect details about the runs (runtime, memory, core usage, etc.). A package embedded in the code can be more accurate, as it tracks metrics in real time and can scale better to a large number of jobs, but only if such a tool exists for the particular combination of programming language, hardware and research field. An embedded package cannot be used to investigate past computations if the tool did not track the parameters at the time. On HPC servers, a server-side tool is expected to combine the strengths of the two other options, as it would not interfere with the code, is independent of the application and does not require manual operations. Server-side tools can also estimate aggregated carbon footprints over a long period of time (months to years). However, the appropriate information needs to be logged and made available by the workload manager (the piece of software allocating resources to users) and errors in the

Table 1 | Pros and cons of each strategy for carbon footprint estimation

Characteristic of each strategy	Online calculator	Embedded package	Server-side tool
Compatible with any hardware	Yes	No, a dedicated tool needs to be built for each	Yes, provided hardware usage can be monitored
Compatible with any programming language	Yes		Yes
Compatible with any task or research field	Yes		Yes
Computing metrics are collected automatically	No	Yes	Yes
Does not interfere with existing code	Yes	No	Yes
Estimates can be obtained beforehand	Yes	No	No
Estimates can be obtained in real time	No	Yes	Yes
Estimates can be obtained retrospectively	Yes	Only if the tracker was active at the time	Yes
Scalable with large numbers of jobs	No	Yes	Yes
Scalable over long periods of time	No	Only if the tracker is used every time	Yes

logs are harder to identify, so underestimations or overestimations can easily stay uncorrected. Finally, carbon footprint estimation is still at its infancy, so if no tool exists yet for a particular application or a particular platform, building one is also an option, using existing open-source code as a starting point, and making the tool available to the community afterwards.

Beyond the direct carbon footprint of running an analysis, other aspects of environmental impacts should be included when assessing research projects. Hardware manufacturing and disposal have significant environmental impacts and should be considered when discussing procurement or renewing policies. For example, around 70% of the life cycle footprint of a laptop is due to manufacturing alone[7]. In some fields, storage needs are high, and the impact of long-term data storage should be included. Estimates vary depending on hardware, but the order of magnitude of the carbon footprint of storing 1 terabyte of data is around 10 kgCO_2e per year. Most projects involve travelling, either for fieldwork or conferences, which can also cause a substantial amount of GHG emissions. If the research project involves work in a laboratory, then the LEAF framework can be used to estimate and reduce carbon footprints.

Systematically estimating the carbon footprint of computational research can limit the waste of resources, encourage the development of energy-efficient software and raise environmental awareness. Environmental impacts can also be included in cost–benefit analyses similarly to how financial costs are considered, for example, as part of funding applications to show the proposal's environmental sustainability. Upon completion of a project, its total carbon footprint can be reported and acknowledged in publications to raise awareness (for example, see ref. [8]). As part of the bigger picture, the aforementioned methods and tools empower research groups and institutions to monitor the computing footprint of all their research projects and use this information to inform future sustainability decisions.

Loïc Lannelongue [1,2,3,4] ✉ **& Michael Inouye**[1,2,3,4,5,6,7]

[1]Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [2]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [3]Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. [4]Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK. [5]Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. [6]British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK. [7]The Alan Turing Institute, London, UK.
✉e-mail: ll582@medschl.cam.ac.uk

Published online: 16 February 2023

References

1. ALLEA. Towards Climate Sustainability of the Academic System in Europe and beyond. *ALLEA* https://doi.org/10.26356/climate-sust-acad (2022).
2. Stevens, A. R. H., Bellstedt, S., Elahi, P. J. & Murphy, M. T. The imperative to reduce carbon emissions in astronomy. *Nat. Astron.* **4**, 843–851 (2020).
3. Arias, P. A. et al. Technical Summary. in *Climate Change 2021: The Physical Science Basis* (eds Masson-Delmotte, V. et al.) (Cambridge Univ. Press, 2021).
4. Lannelongue, L., Grealey, J. & Inouye, M. Green algorithms: quantifying the carbon footprint of computation. *Adv. Sci.* **8**, 2100707 (2021).
5. Anthony, L. F. W., Kanding, B. & Selvan, R. Carbontracker: tracking and predicting the carbon footprint of training deep learning models. Preprint at https://arxiv.org/abs/2007.03051 (2020).
6. Lannelongue, L. GreenAlgorithms4HPC. *GitHub* https://github.com/GreenAlgorithms/GreenAlgorithms4HPC (2022).
7. Clément, L.-P. P.-V. P., Jacquemotte, Q. E. S. & Hilty, L. M. Sources of variation in life cycle assessments of smartphones and tablet computers. *Environ. Impact Assess. Rev.* **84**, 106416 (2020).
8. Qin, Y. et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* **54**, 134–142 (2022).

Acknowledgements

L.L. was supported by the University of Cambridge MRC DTP (MR/S502443/1). M.I. was supported by the Munz Chair of Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014)*. M.I. was also supported by the UK Economic and Social Research 878 Council (ES/T013192/1). This work was supported by core funding from the British Heart Foundation (RG/13/13/30194; RG/18/13/33946) and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014)*. *The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), and British Heart Foundation and Wellcome.

Competing interests

The authors declare no competing interests.