

GREENER principles for environmentally sustainable computational science

Received: 6 November 2022

Accepted: 9 May 2023

Published online: 26 June 2023

 Check for updates

Loïc Lannelongue^{1,2,3,4}✉, Hans-Erik G. Aronson⁵, Alex Bateman⁶, Ewan Birney⁶, Talia Caplan⁷, Martin Jukes⁸, Johanna McEntyre⁶, Andrew D. Morris⁵, Gerry Reilly⁵ & Michael Inouye^{1,2,3,4,9,10,11}

The carbon footprint of scientific computing is substantial, but environmentally sustainable computational science (ESCS) is a nascent field with many opportunities to thrive. To realize the immense green opportunities and continued, yet sustainable, growth of computer science, we must take a coordinated approach to our current challenges, including greater awareness and transparency, improved estimation and wider reporting of environmental impacts. Here, we present a snapshot of where ESCS stands today and introduce the GREENER set of principles, as well as guidance for best practices moving forward.

Scientific research and development have transformed and immeasurably improved the human condition, whether by building instruments to unveil the mysteries of the universe, developing treatments to fight cancer or improving our understanding of the human genome. Yet, science can, and frequently does, impact the environment, and the magnitude of these impacts is not always well understood. Given the connection between climate change and human health, it is becoming increasingly apparent to biomedical researchers in particular, as well as their funders, that the environmental effects of research should be taken into account^{1–5}.

Recent studies have begun to elucidate the environmental impacts of scientific research, with an initial focus on scientific conferences and experimental laboratories⁶. The 2019 Fall Meeting of the American Geophysical Union was estimated to emit 80,000 metric tonnes of CO₂ equivalent (tCO₂e), equivalent to the average weekly emissions of the city of Edinburgh, UK⁷ (CO₂e, or CO₂-equivalent, summarizes the global warming impacts of a range of greenhouse gases (GHGs) and is the standard metric for carbon footprints, although its accuracy is sometimes debated⁸). The annual meeting of the Society for Neuroscience was estimated to emit 22,000 tCO₂e, approximately the annual carbon footprint of 1,000 medium-sized laboratories⁹.

The life-cycle impact (including construction and usage) of university buildings has been estimated at $-0.125 \text{ tCO}_2\text{e m}^{-2} \text{ yr}^{-1}$ (ref. 10), and the yearly carbon footprint of a typical life-science laboratory at $-20 \text{ tCO}_2\text{e}$ (ref. 9). The Laboratory Efficiency Assessment Framework (LEAF) is a widely adopted standard to monitor and reduce the carbon footprint of laboratory-based research¹¹. Other recent frameworks can help to raise awareness: GES Ipoint5¹² provides an open-source tool to estimate the carbon footprint of research laboratories and covers buildings, procurement, commuting and travel, and the Environmental Responsibility 5-R Framework provides guidelines for ecologically conscious research¹³.

With the increasing scale of high-performance and cloud computing, the computational sciences are susceptible to having silent and unintended environmental impacts. The sector of information and communication technologies (ICT) was responsible for between 1.8% and 2.8% of global GHG emissions in 2020¹⁴—more than aviation (1.9%¹⁵)—and, if unchecked, the ICT carbon footprint could grow exponentially in coming years¹⁴. Although the environmental impact of experimental ‘wet’ laboratories is more immediately obvious, with their large pieces of equipment and high plastic and reagent usage, the impact of algorithms is less clear and often underestimated. The risks

¹Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ³Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK. ⁴Health Data Research UK Cambridge, Wellcome Genome Campus and University of Cambridge, Cambridge, UK. ⁵Health Data Research (HDR) UK, London, UK. ⁶European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, UK. ⁷Wellcome Trust, London, UK. ⁸RAL Space, Science and Technology Facilities Council, Harwell Campus, Didcot, UK. ⁹Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. ¹⁰British Heart Foundation Centre of Research Excellence, University of Cambridge, Cambridge, UK. ¹¹The Alan Turing Institute, London, UK.

✉e-mail: ll582@medschl.cam.ac.uk

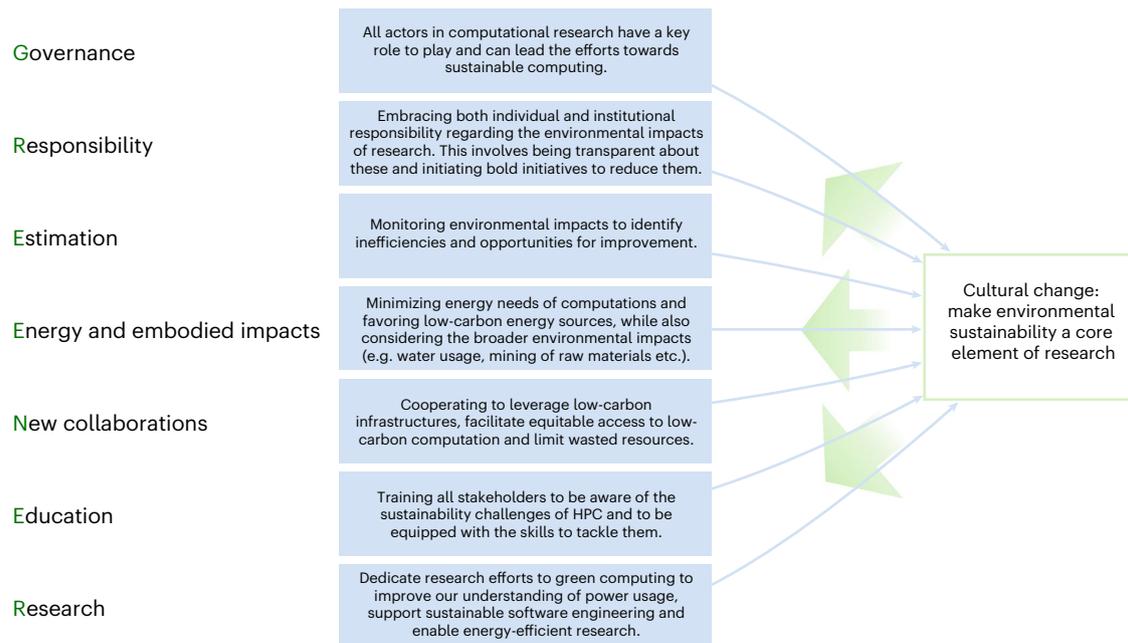


Fig. 1 | GREENER principles for ESCS. The GREENER principles enable cultural change (blue arrows), which in turn facilitates their implementation (green arrows) and triggers a virtuous circle.

of seeking performance at any cost and the importance of considering energy usage and sustainability when developing new hardware for high-performance computing (HPC) was raised as early as 2007¹⁶. Since then, continuous improvements have been made by developing new hardware, building lower-energy data centers and implementing more efficient HPC systems^{17,18}. However, it is only in the past five years that these concerns have reached HPC users, in particular researchers. Notably, the field of artificial intelligence (AI) has first taken note of its environmental impacts, in particular those of the very large language models developed^{19–23}. It is unclear, however, to what extent this has led the field towards more sustainable research practices. A small number of studies have also been performed in other fields, including bioinformatics²⁴, astronomy and astrophysics^{25–28}, particle physics²⁹, neuroscience³⁰ and computational social sciences³¹. Health data science is starting to address the subject, but a recent systematic review found only 25 publications in the field over the past 12 years³². In addition to the environmental effects of electricity usage, manufacturing and disposal of hardware, there are also concerns around data centers’ water usage and land footprint³³. Notably, computational science, in particular AI, has the potential to help fight climate change, for example, by improving the efficiency of wind farms, by facilitating low-carbon urban mobility and by better understanding and anticipating severe weather events³⁴.

In this Perspective we highlight the nascent field of environmentally sustainable computational science (ESCS)—what we have learned from the research so far, and what scientists can do to mitigate their environmental impacts. In doing so, we present GREENER (Governance, Responsibility, Estimation, Energy and embodied impacts, New collaborations, Education and Research; Fig. 1), a set of principles for how the computational science community could lead the way in sustainable research practices, maximizing computational science’s benefit to both humanity and the environment.

Environmental impacts of the computational sciences

The past three years have seen increased concerns regarding the carbon footprint of computations, and only recently have tools^{21,35–37} and guidelines³⁸ been widely available to computational scientists to

allow them to estimate their carbon footprint and be more environmentally sustainable.

Most calculators that estimate the carbon footprint of computations are targeted at machine learning tasks and so are primarily suited to Python pipelines, graphics processing units (GPUs) and/or cloud computing^{36,37,39,40}. Python libraries have the benefit of integrating well into machine learning pipelines or online calculators for cloud GPUs^{21,41}. Recently, a flexible online tool, the Green Algorithms calculator³⁵, enabled the estimation of the carbon footprint for nearly any computational task, empowering sustainability metrics across fields, hardware, computing platforms and locations.

Some publications, such as ref. 38, have listed simple actions that computational scientists can take regarding their environmental impact, including estimating the carbon footprint of running algorithms, both a posteriori to acknowledge the impact of a project and before starting as part of a cost–benefit analysis. A 2020 report from The Royal Society formalizes this with the notion of ‘energy proportionality’, meaning the environmental impacts of an innovation must be outweighed by its environmental or societal benefits³⁴. It is also important to minimize electronic waste by keeping devices for longer and using second-hand hardware when possible. A 2021 report by the World Health Organization⁴² warns of the dramatic effect of e-waste on population health, particularly children. The unregulated informal recycling industry, which handles more than 80% of the 53 million tonnes of e-waste, causes a high level of water, soil and air pollution, often in low- and middle-income countries⁴³. The up to 56 million informal waste workers are also exposed to hazardous chemicals such as heavy metals and persistent organic pollutants⁴². Scientists can also choose energy-efficient hardware and computing facilities, while favoring those powered by green energy. Writing efficient code can substantially reduce the carbon footprint as well, and this can be done alongside making hardware requirements and carbon footprints clear when releasing new software. The Green Software Foundation (<https://greensoftware.foundation>) promotes carbon-aware coding to reduce the operational carbon footprint of the softwares used in all aspects of society. There is, however, a rebound effect to making algorithms and hardware more efficient: instead of reducing computing usage, increased efficiency encourages more analyses to

be performed, which leads to a reevaluation of the cost–benefit but often results in increased carbon footprints. The rebound effect is a key example of why research practice should adapt to technological advances so that they lead to carbon footprint reductions.

GREENER computational science

ESCS is an emerging field, but one that is of rapidly increasing importance given the climate crisis. In the following, our proposed set of principles (Fig. 1) outlines the main axes where progress is needed, where opportunities lie and where we believe efforts should be concentrated.

Governance and responsibility

Everyone involved in computational science has a role to play in making the field more sustainable, and many do already, from grassroots movements to large institutions. Individual and institutional responsibility is a necessary step to ensure transparency and reduction of GHG emission. Here we highlight key stakeholders alongside existing initiatives and future opportunities for involvement.

Grassroots initiatives led by graduate students, early career researchers and laboratory technicians have shown great success in tackling the carbon footprint of laboratory work, including Green Labs Netherlands⁴⁴, the Nottingham Technical Sustainability Working Group or the Digital Humanities Climate Coalition⁴⁵. International coalitions such as the Sustainable Research (SuRe) Symposium, initially set up for wet laboratories, have started to address the impact of computing as well. IT teams in HPC centers are naturally key, both in terms of training and ensuring that the appropriate information is logged so that scientists can follow the carbon footprints of their work. Principal investigators can encourage their teams to think about this issue and provide access to suitable training when needed.

Simultaneously, top–down approaches are needed, with funding bodies and journals occupying key positions in both incentivizing carbon-footprint reduction and in promoting transparency. Funding bodies can directly influence the researchers they fund and those applying for funding via their funding policies. They can require estimates of carbon footprints to be included in funding applications as part of ‘environmental impacts statements’. Many funding bodies include sustainability in their guidelines already; see, for example, the UK’s NIHR carbon reduction guidelines¹, the brief mention of the environment in UKRI’s terms and conditions⁴⁶, and the Wellcome Trust’s carbon-offsetting travel policy⁴⁷.

Although these are important first steps, bolder action is needed to meet the urgency of climate change. For example, UKRI’s digital research infrastructure scoping project⁴⁸, which seeks to provide a roadmap to net zero for its digital infrastructure, sends a clear message that sustainable research includes minimizing the GHG emissions from computation. The project not only raises awareness but will hopefully result in reductions in GHG emissions.

Large research institutes are key to managing and expanding centralized data infrastructures and trusted research environments (TREs). For example, EMBL’s European Bioinformatics Institute manages more than 40 data resources⁴⁹, including AlphaFold DB⁵⁰, which contains over 200,000,000 predicted protein structures that can be searched, browsed and retrieved according to the FAIR principles (findable, accessible, interoperable, reusable)⁵¹. As a consequence, researchers do not need to run the carbon-intensive AlphaFold algorithm for themselves and instead can just query the database. AlphaFold DB was queried programmatically over 700 million times and the web page was accessed 2.4 million times between August 2021 and October 2022. Institutions also have a role in making procurement decisions carefully, taking into account both the manufacturing and operational footprint of hardware purchases. This is critical, as the lifetime footprint of a computational facility is largely determined by the date it is purchased. Facilities could also better balance investment decisions, with a focus on attracting staff based on sustainable and efficient working environments, rather than high-powered hardware⁵².

However, increases in the efficiencies of digital technology alone are unlikely to prove sufficient in ensuring sustainable resource use⁵³. Alongside these investments, funding bodies should support a shift towards more positive, inclusive and green research cultures, recognizing that more data or bigger models do not always translate into greater insights and that a ‘fit for purpose’ approach can ultimately be more efficient. Organizations such as Health Data Research UK and the UK Health Data Research Alliance have a key convening role in ensuring that awareness is raised around the climate impact of both infrastructure investment and computational methods.

Journals may incentivize authors to acknowledge and indeed estimate the carbon footprint of the work presented. Some authors already do this voluntarily (for example, refs. 54–59), mostly in bioinformatics and machine learning so far, but there is potential to expand it to other areas of computational science. In some instances, showing that a new tool is greener can be an argument in support of a new method⁶⁰.

International societies in charge of organizing annual conferences may help scientists reduce the carbon footprint of presenting their work by offering hybrid options. The COVID-19 pandemic boosted virtual and hybrid meetings, which have a lower carbon footprint while increasing access and diversity^{7,61}. Burtscher and colleagues found that running the annual meeting of the European Astronomical Society online emitted >3,000-fold less CO₂e than the in-person meeting (0.582 tCO₂e compared to 1,855 tCO₂e)²⁵. Institutions are starting to tackle this; for example, the University of Cambridge has released new travel guidelines encouraging virtual meetings whenever feasible and restricting flights to essential travel, while also acknowledging that different career stages have different needs⁶².

Industry partners will also need to be part of the discussion. Acknowledging and reducing computing environmental impact comes with added challenges in industry, such as shareholder interests and/or public relations. While the EU has backed some initiatives helping ICT-reliant companies to address their carbon footprint, such as ICTfootprint.eu, other major stakeholders have expressed skepticism regarding the environmental issues of machine learning models^{63,64}. Although challenging, tech industry engagement and inclusion is nevertheless essential for tackling GHG emissions.

Estimate and report the energy consumption of algorithms

Estimating and monitoring the carbon footprint of computations is an essential step towards sustainable research as it identifies inefficiencies and opportunities for improvement. User-level metrics are crucial to understanding environmental impacts and promoting personal responsibility. In some HPC situations, particularly in academia, the financial cost of running computations is negligible and scientists may have the impression of unlimited and inconsequential computing capacity. Quantifying the carbon footprint of individual projects helps raise awareness of the true costs of research.

Although progress has been made in estimating energy usage and carbon footprints over the past few years, there are still barriers that prevent the routine estimation of environmental impacts. From task-agnostic, general-purpose calculators³⁵ and task-specific packages^{36,37,65} to server-side softwares^{66,67}, each estimation tool is a trade-off between ease of use and accuracy. A recent primer⁶⁸ discusses these different options in more detail and provides recommendations as to which approach fits a particular need.

Regardless of the calculator used, for these tools to work effectively and for scientists to have an accurate representation of their energy consumption, it is important to understand the power management for different components. For example, the power usage of processing cores such as central processing units (CPUs) and GPUs is not a readily available metric; instead, thermal design power (meaning, how much heat the chip can be expected to dissipate in a normal setting) is used. Although an acceptable approximation, it has also been shown to substantially underestimate power usage in some

situations⁶⁹. The efficiency of data centers is measured by the power usage effectiveness (PUE), which quantifies how much energy is needed for non-computing tasks, mainly cooling (efficient data centers have PUEs close to 1). This metric is widely used, with large cloud providers reporting low PUEs (for example, 1.11 for Google⁷⁰ compared to a global average of 1.57⁷¹), but discrepancies in how it is calculated can limit PUE interpretation and thus its impact^{72–74}. A standard from the International Organization for Standardization is trying to address this⁷⁵. Unfortunately, the PUE of a particular data center, whether cloud or institutional, is rarely publicly documented. Thus, an important step is the data science and infrastructure community making both hardware and data centers' energy consumption metrics available to their users and the public. Ultimately, tackling unnecessary carbon footprints will require transparency³⁴.

Tackling energy and embodied impacts through new collaborations

Minimizing carbon intensity (meaning the carbon footprint of producing electricity) is one of the most immediately impactful ways to reduce GHG emissions. Carbon intensities depend largely on geographical location, with up to three orders of magnitude between the top and bottom performing high-income countries in terms of low carbon energies (from 0.10 gCO₂e kWh⁻¹ in Iceland to 770 gCO₂e kWh⁻¹ in Australia⁷⁶). Changing the carbon intensity of a local state or national government is nearly always impractical as it would necessitate protracted campaigns to change energy policies. An alternative is to relocate computations to low-carbon settings and countries, but, depending on the type of facility or the sensitivity of the data, this may not always be possible. New inter-institutional cooperation may open up opportunities to enable access to low-carbon data centers in real time.

It is, however, essential to recognize and account for inequalities between countries in terms of access to green energy sources. International cooperation is key to providing scientists from low- and middle-income countries (LMICs), who frequently only have high-carbon-intensity options available to them, access to low-carbon computing infrastructures for their work. In the longer term, international partnerships between organizations and nations can help build low-carbon computing capacity in LMICs.

Furthermore, the footprint of user devices should not be forgotten. In one estimate, the energy footprint of streaming a video to a laptop is mainly on the laptop (72%), with 23% used in transmission and a mere 5% at the data center⁷⁷. Zero clients (user devices with no compute or storage capacity) can be used in some research use cases and drastically reduce the client-side footprint⁷⁸.

It can be tempting to reduce the environmental impacts of computing to electricity needs, as these are the easiest ones to estimate. However, water usage, ecological impacts and embodied carbon footprints from manufacturing should also be addressed. For example, for personal hardware, such as laptops, 70–80% of the life-cycle impact of these devices comes from manufacturing only⁷⁹, as it involves mining raw materials and assembling the different components, which require water and energy. Moreover, manufacturing often takes place in countries that have a higher carbon intensity for power generation and a slower transition to zero-carbon power⁸⁰. Currently, hardware renewal policies, either for work computers or servers in data centers, are often closely dependent on warranties and financial costs, with environmental costs rarely considered. For hardware used in data centers, regular updates may be both financially and environmentally friendly, as efficiency gains may offset manufacturing impacts. Estimating these environmental impacts will allow HPC teams to know for sure. Reconditioned and remanufactured laptops and servers are available, but growth of this sector is currently limited by negative consumer perception⁸¹. Major suppliers of hardware are making substantial commitments, such as 100% renewable energy supply by 2030⁸² or net zero by 2050⁸³.

Another key consideration is data storage. Scientific datasets are now measured in petabytes (PB). In genomics, the popular UK Biobank cohort⁸⁴ is expected to reach 15 PB by 2025⁸⁵, and the first image of a black hole required the collection of 5 PB of data⁸⁶. The carbon footprint of storing data depends on numerous factors, but based on some manufacturers' estimations, the order of magnitude of the life-cycle footprint of storing 1 TB of data for a year is -10 kg CO₂e (refs. 87,88). This issue is exacerbated by the duplication of such datasets in order for each institution, and sometimes each research group, to have a copy. Centralized and collaborative computing resources (such as TREs) holding both data and computing hardware may help alleviate redundant resources. TRE efforts in the UK span both health (for example, NHS Digital⁸⁹) and administrative data (for example, the SAIL databank on the UK Secure Research Platform⁹⁰ and the Office for National Statistics Secure Research Service⁹¹). Large (hyperscale) data centers are expected to be more energy-efficient⁹², but they may also encourage unnecessary increases in the scale of computing (rebound effect).

The importance of dedicated education and research efforts for ESCS

Education is essential to raise awareness with different stakeholders. In lieu of incorporating some aspects into more formal undergraduate programs, integrating sustainability into computational training courses is a tangible first step toward reducing carbon footprints. An example is the 'Green Computing' Workshop on Education at the 2022 conference on Intelligent Systems for Molecular Biology.

Investing in research that will catalyze innovation in the field of ESCS is a crucial role for funders and institutions to play. Although global data centers' workloads have increased more than sixfold between 2010 and 2018, their total electricity usage has been approximately stable due to the use of power-efficient hardware⁹³, but environmentally sustainable investments will be needed to perpetuate this trend. Initiatives like Wellcome's Research Sustainability project⁹⁴, which look to highlight key gaps where investment could deliver the next generation of ESCS tools and technology, are key to ensuring that growth in energy demand beyond current efficiency trends can be managed in a sustainable way. Similarly, the UKRI Data and Analytics Research Environments UK program (DARE UK) needs to ensure that sustainability is a key evaluation criterion for funding and infrastructure investments for the next generation of TREs.

Recent studies found that the most widely used programming languages in research, such as R and Python⁹⁵, tend to be the least energy-efficient ones^{96,97}, and, although it is unlikely that forcing the community to switch to more efficient languages would benefit the environment in the short term (due to inefficient coding for example), this highlights the importance of having trained research software engineers within research groups to ensure that the algorithms used are efficiently implemented. There is also scope to use current tools more efficiently by better understanding and monitoring how coding choices impact carbon footprints. Algorithms also come with high memory requirements, sometimes using more energy than processors⁹⁸. Unfortunately, memory power usage remains poorly optimized, as speed of access is almost always favored over energy efficiency⁹⁹. Providing users and software engineers with the flexibility to opt for energy efficiency would present an opportunity for a reduction in GHG emissions^{100,101}.

Cultural change

In parallel to the technological reductions in energy usage and carbon footprints, research practices will also need to change to avoid rebound effects³⁸. Similar to the aviation industry, there is a tendency to count on technology to solve sustainability concerns without having to change usage¹⁰² (that is, waiting on computing to become zero-carbon rather than acting on how we use it). Cultural change in the computing community to reconsider how we think about computing costs will

be necessary. Research strategies at all levels will need to consider environmental impacts and corresponding approaches to carbon footprint minimization. The upcoming extension of the LEAF standard for computational laboratories will provide researchers with tangible tools to do so. Day to day, there is a need to solve trade-offs between the speed of computation, accuracy and GHG emissions, keeping in mind the goal of GHG reduction. These changes in scientific practices are challenging, but, importantly, there are synergies between open computational science and green computing¹⁰³. For example, making code, data and models FAIR so that other scientists avoid unnecessary computations can increase the reach and impact of a project. FAIR practices can result in highly efficient code implementations, reduce the need to retrain models, and reduce unnecessary data generation/storage, thus reducing the overall carbon footprint. As a result, green computing and FAIR practices may both stimulate innovation and reduce financial costs.

Moreover, computational science has downstream effects on carbon footprints in other areas. In the biomedical sciences, developments in machine learning and computer vision impact the speed and scale of medical imaging processing. Discoveries in health data science make their way to clinicians and patients through, for example, connected devices. In each of these cases and many others, environmental impacts propagate through the whole digital health sector³². Yet, here too synergies exist. In many cases, such as telemedicine, there may be a net benefit in terms of both carbon and patient care, provided that all impacts have been carefully accounted for. These questions are beginning to be tackled in medicine, such as assessments of the environmental impact of telehealth¹⁰⁴ or studies into ways to sustainably handle large volumes of medical imaging data¹⁰⁵. For the latter, NHS Digital (the UK's national provider of information, data and IT systems for health and social care) has released guidelines to this effect¹⁰⁶. Outside the biomedical field, there are immense but, so far, unrealized opportunities for similar efforts.

Conclusion

The computational sciences have an opportunity to lead the way in sustainability, which may be achieved through the GREENER principles for ESCS (Fig. 1): Governance, Responsibility, Estimation, Energy and embodied impacts, New collaborations, Education and Research. This will require more transparency on environmental impacts. Although some tools already exist to estimate carbon footprints, more specialized ones will be needed alongside a clearer understanding of the carbon footprint of hardware and facilities, as well as more systematic monitoring and acknowledgment of carbon footprints. Measurement is a first step, followed by a reduction in GHG emissions. This can be achieved with better training and sensible policies for renewing hardware and storing data. Cooperation, open science and equitable access to low-carbon computing facilities will also be crucial¹⁰⁷. Computing practices will need to adapt to include carbon footprints in cost–benefit analyses, as well as consider the environmental impacts of downstream applications. The development of sustainable solutions will need particularly careful consideration, as they frequently have the least benefit for populations, often in LMICs, who suffer the most from climate change^{22,108}. All stakeholders have a role to play, from funding bodies, journals and institutions to HPC teams and early career researchers. There is now a window of time and an immense opportunity to transform computational science into an exemplar of broad societal impact and sustainability.

References

- NIHR Carbon Reduction Guidelines (National Institute for Health and Care Research, 2019); <https://www.nihr.ac.uk/documents/nihr-carbon-reduction-guidelines/21685>
- NHS Becomes the World's First National Health System to Commit to Become 'Carbon Net Zero', Backed by Clear Deliverables and Milestones (NHS England, 2020); <https://www.england.nhs.uk/2020/10/nhs-becomes-the-worlds-national-health-system-to-commit-to-become-carbon-net-zero-backed-by-clear-deliverables-and-milestones/>
- Climate and COVID-19: converging crises. *Lancet* **397**, 71 (2021).
- Marazziti, D. et al. Climate change, environment pollution, COVID-19 pandemic and mental health. *Sci. Total Environ.* **773**, 145182 (2021).
- Wellcome Commissions Report on Science's Environmental Impact (Wellcome, 2022); <https://wellcome.org/news/wellcome-commissions-report-sciences-environmental-impact>
- Towards Climate Sustainability of the Academic System in Europe and Beyond (ALLEA, 2022); <https://doi.org/10.26356/climate-sust-acad>
- Klöwer, M., Hopkins, D., Allen, M. & Higham, J. An analysis of ways to decarbonize conference travel after COVID-19. *Nature* **583**, 356–359 (2020).
- Allen, M. R. et al. A solution to the misrepresentations of CO₂-equivalent emissions of short-lived climate pollutants under ambitious mitigation. *npj Clim. Atmos. Sci.* **1**, 16 (2018).
- Nathans, J. & Sterling, P. How scientists can reduce their carbon footprint. *eLife* **5**, e15928 (2016).
- Helmers, E., Chang, C. C. & Dauwels, J. Carbon footprinting of universities worldwide part II: first quantification of complete embodied impacts of two campuses in Germany and Singapore. *Sustainability* **14**, 3865 (2022).
- Marshall-Cook, J. & Farley, M. *Sustainable Science and the Laboratory Efficiency Assessment Framework (LEAF)* (UCL, 2023).
- Mariette, J. et al. An open-source tool to assess the carbon footprint of research. *Environ. Res. Infrastruct. Sustain.* **2**, 035008 (2022).
- Murray, D. S. et al. The environmental responsibility framework: a toolbox for recognizing and promoting ecologically conscious research. *Earth's Future* **11**, e2022EF002964 (2023).
- Freitag, C. et al. The real climate and transformative impact of ICT: a critique of estimates, trends and regulations. *Patterns* **2**, 100340 (2021).
- Ritchie, H. Climate change and flying: what share of global CO₂ emissions come from aviation? *Our World in Data* (22 October 2022); <https://ourworldindata.org/co2-emissions-from-aviation>
- Feng, W. & Cameron, K. The Green500 list: encouraging sustainable supercomputing. *Computer* **40**, 50–55 (2007).
- Garg, S. K., Yeo, C. S., Anandasivam, A. & Buyya, R. Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers. *J. Parallel Distrib. Comput.* **71**, 732–749 (2011).
- Katal, A., Dahiya, S. & Choudhury, T. Energy efficiency in cloud computing data centers: a survey on software technologies. *Clust. Comput.* <https://doi.org/10.1007/s10586-022-03713-0> (2022).
- Strubell, E., Ganesh, A. & McCallum, A. Energy and policy considerations for deep learning in NLP. In *Proc. 57th Annual Meeting of the Association for Computational Linguistics* 3645–3650 (Association for Computational Linguistics, 2019); <https://doi.org/10.18653/v1/P19-1355>
- Schwartz, R., Dodge, J., Smith, N. A. & Etzioni, O. Green AI. Preprint at <https://arxiv.org/abs/1907.10597> (2019).
- Lacoste, A., Luccioni, A., Schmidt, V. & Dandres, T. Quantifying the carbon emissions of machine learning. Preprint at <https://arxiv.org/abs/1910.09700> (2019).
- Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the dangers of stochastic parrots: can language models be too big? In *Proc. 2021 ACM Conference on Fairness, Accountability and Transparency* 610–623 (Association for Computing Machinery, 2021); <https://doi.org/10.1145/3442188.3445922>
- Memmel, E., Menzen, C., Schuurmans, J., Wesel, F. & Batselier, K. Towards Green AI with tensor networks—sustainability and innovation enabled by efficient algorithms. Preprint at <https://doi.org/10.48550/arXiv.2205.12961> (2022).

24. Grealey, J. et al. The carbon footprint of bioinformatics. *Mol. Biol. Evol.* **39**, msac034 (2022).
25. Burtscher, L. et al. The carbon footprint of large astronomy meetings. *Nat. Astron.* **4**, 823–825 (2020).
26. Jahnke, K. et al. An astronomical institute's perspective on meeting the challenges of the climate crisis. *Nat. Astron.* **4**, 812–815 (2020).
27. 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).
28. Portegies Zwart, S. The ecological impact of high-performance computing in astrophysics. *Nat. Astron.* **4**, 819–822 (2020).
29. Bloom, K. et al. Climate impacts of particle physics. Preprint at <https://arxiv.org/abs/2203.12389> (2022).
30. Aron, A. R. et al. How can neuroscientists respond to the climate emergency? *Neuron* **106**, 17–20 (2020).
31. Leslie, D. *Don't 'Research Fast and Break Things': on the Ethics of Computational Social Science* (Zenodo, 2022); <https://doi.org/10.5281/zenodo.6635569>
32. Samuel, G. & Lucassen, A. M. The environmental sustainability of data-driven health research: a scoping review. *Digit. Health* **8**, 205520762211112 (2022).
33. Al Kez, D., Foley, A. M., Laverty, D., Del Rio, D. F. & Sovacool, B. Exploring the sustainability challenges facing digitalization and internet data centers. *J. Clean. Prod.* **371**, 133633 (2022).
34. *Digital Technology and the Planet—Harnessing Computing to Achieve Net Zero* (The Royal Society, 2020); <https://royalsociety.org/topics-policy/projects/digital-technology-and-the-planet/>
35. Lannelongue, L., Grealey, J. & Inouye, M. Green algorithms: quantifying the carbon footprint of computation. *Adv. Sci.* **8**, 2100707 (2021).
36. Henderson, P. et al. Towards the systematic reporting of the energy and carbon footprints of machine learning. *J. Mach. Learn. Res.* **21**, 10039–10081 (2020).
37. 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).
38. Lannelongue, L., Grealey, J., Bateman, A. & Inouye, M. Ten simple rules to make your computing more environmentally sustainable. *PLoS Comput. Biol.* **17**, e1009324 (2021).
39. Valeye, F. Tracarbon. *GitHub* <https://github.com/fvaley/tracarbon> (2022).
40. Trébaol, T. *CUMULATOR—a Tool to Quantify and Report the Carbon Footprint of Machine Learning Computations and Communication in Academia and Healthcare* (École Polytechnique Fédérale de Lausanne, 2020).
41. Cloud Carbon Footprint—An open source tool to measure and analyze cloud carbon emissions. <https://www.cloudcarbonfootprint.org/> (2023).
42. *Children and Digital Dumpsites: E-Waste Exposure and Child Health* (World Health Organization, 2021); <https://apps.who.int/iris/handle/10665/341718>
43. Sepúlveda, A. et al. A review of the environmental fate and effects of hazardous substances released from electrical and electronic equipments during recycling: examples from China and India. *Environ. Impact Assess. Rev.* **30**, 28–41 (2010).
44. Franssen, T. & Johnson, H. *The Implementation of LEAF at Public Research Organisations in the Biomedical Sciences: a Report on Organisational Dynamics* (Zenodo, 2021); <https://doi.org/10.5281/ZENODO.5771609>
45. DHCC Information, Measurement and Practice Action Group. *A Researcher Guide to Writing a Climate Justice Oriented Data Management Plan* (Zenodo, 2022); <https://doi.org/10.5281/ZENODO.6451499>
46. UKRI. *UKRI Grant Terms and Conditions* (UKRI, 2022); <https://www.ukri.org/wp-content/uploads/2022/04/UKRI-050422-FullEconomicCostingGrantTermsConditionsGuidance-Apr2022.pdf>
47. *Carbon Offset Policy for Travel—Grant Funding* (Wellcome, 2021); <https://wellcome.org/grant-funding/guidance/carbon-offset-policy-travel>
48. Juckes, M., Pascoe, C., Woodward, L., Vanderbauwhede, W. & Weiland, M. *Interim Report: Complexity, Challenges and Opportunities for Carbon Neutral Digital Research* (Zenodo, 2022); <https://zenodo.org/record/7016952>
49. Thakur, M. et al. EMBL's European Bioinformatics Institute (EMBL-EBI) in 2022. *Nucleic Acids Res.* **51**, D9–D17 (2022).
50. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
51. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
52. Bichsel, J. *Research Computing: The Enabling Role of Information Technology* (Educause, 2012); <https://library.educause.edu/resources/2012/11/research-computing-the-enabling-role-of-information-technology>
53. Creutzig, F. et al. Digitalization and the Anthropocene. *Annu. Rev. Environ. Resour.* **47**, 479–509 (2022).
54. Yang, L. & Chen, J. A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. *Microbiome* **10**, 130 (2022).
55. 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).
56. Lannelongue, L. & Inouye, M. *Inference Mechanisms and Prediction of Protein-Protein Interactions*. Preprint at <http://biorxiv.org/lookup/doi/10.1101/2022.02.07.479382> (2022).
57. Dubois, F. *The Vehicle Routing Problem for Flash Floods Relief Operations* (Univ. Paul Sabatier, 2022).
58. Thiele, L., Cranmer, M., Coulton, W., Ho, S. & Spergel, D. N. Predicting the thermal Sunyaev-Zel'dovich field using modular and equivariant set-based neural networks. Preprint at <https://arxiv.org/abs/2203.00026> (2022).
59. Armstrong, G. et al. Efficient computation of Faith's phylogenetic diversity with applications in characterizing microbiomes. *Genome Res.* **31**, 2131–2137 (2021).
60. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
61. Estien, C. O., Myron, E. B., Oldfield, C. A. & Alwin, A. & Ecological Society of America Student Section Virtual scientific conferences: benefits and how to support underrepresented students. *Bull. Ecol. Soc. Am.* **102**, e01859 (2021).
62. University of Cambridge. *Guidelines for Sustainable Business Travel* (Univ. Cambridge, 2022); https://www.environment.admin.cam.ac.uk/files/guidelines_for_sustainable_business_travel_approved.pdf
63. Patterson, D. et al. Carbon emissions and large neural network training. Preprint at <https://arxiv.org/abs/2104.10350> (2021).
64. Patterson, D. et al. The carbon footprint of machine learning training will plateau, then shrink. *Computer* **55**, 18–28 (2022).
65. *Neuroimaging Pipeline Carbon Tracker Toolboxes* (OHBM SEA-SIG, 2023); <https://ohbm-environment.org/carbon-tracker-toolboxes/>
66. Lannelongue, L. *Green Algorithms for High Performance Computing* (GitHub, 2022); <https://github.com/Lannelongue/GreenAlgorithms4HPC>
67. *Carbon Footprint Reporting—Customer Carbon Footprint Tool* (Amazon Web Services, 2023); <https://aws.amazon.com/aws-cost-management/aws-customer-carbon-footprint-tool/>
68. Lannelongue, L. & Inouye, M. Carbon footprint estimation for computational research. *Nat. Rev. Methods Prim.* **3**, 9 (2023).

69. Cutress, I. *Why Intel Processors Draw More Power Than Expected: TDP and Turbo Explained* (Anandtech, 2018); <https://www.anandtech.com/show/13544/why-intel-processors-draw-more-power-than-expected-tdp-turbo>
70. Efficiency. *Google Data Centers* <https://www.google.com/about/datacenters/efficiency/>
71. *Uptime Institute Releases 2021 Global Data Center Survey* (Facility Executive, 2021); <https://facilityexecutive.com/2021/09/uptime-institute-releases-2021-global-data-center-survey/>
72. Zoie, R. C., Mihaela, R. D. & Alexandru, S. An analysis of the power usage effectiveness metric in data centers. In *Proc. 2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)* 1–6 (IEEE, 2017); <https://doi.org/10.1109/ISEEE.2017.8170650>
73. Yuventi, J. & Mehdiadeh, R. A critical analysis of power usage effectiveness and its use in communicating data center energy consumption. *Energy Build.* **64**, 90–94 (2013).
74. Avelar, V., Azevedo, D. & French, A. (eds) *PUE: A Comprehensive Examination of the Metric* White Paper No. 49 (Green Grid, 2012).
75. *Power Usage Effectiveness (PUE)* (ISO/IEC); <https://www.iso.org/obp/ui/#iso:std:iso-iec:30134:-2:ed-1:v1:en>
76. *2022 Country Specific Electricity Grid Greenhouse Gas Emission Factors* (Carbon Footprint, 2023); https://www.carbonfootprint.com/docs/2023_02_emissions_factors_sources_for_2022_electricity_v10.pdf
77. Kamiya, G. *The Carbon Footprint of Streaming Video: Fact-Checking the Headlines—Analysis* (IEA, 2020); <https://www.iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-headlines>
78. Rot, A., Chrobak, P. & Sobinska, M. Optimisation of the use of IT infrastructure resources in an institution of higher education: a case study. In *Proc. 2019 9th International Conference on Advanced Computer Information Technologies (ACIT)* 171–174 (IEEE, 2019); <https://doi.org/10.1109/ACITT.2019.8780018>
79. 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).
80. Kamal, K. Y. The silicon age: trends in semiconductor devices industry. *JESTR* **15**, 110–115 (2022).
81. Gåvertsson, I., Milios, L. & Dalhammar, C. Quality labelling for re-used ICT equipment to support consumer choice in the circular economy. *J. Consum. Policy* **43**, 353–377 (2020).
82. *Intel Corporate Responsibility Report 2021–2022* (Intel, 2022); <https://csrreportbuilder.intel.com/pdfbuilder/pdfs/CSR-2021-22-Full-Report.pdf>
83. *TSMC Task Force on Climate-related Financial Disclosures* (TSMC, 2020); https://esg.tsmc.com/download/file/TSMC_TCFD_Report_E.pdf
84. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
85. *UK Biobank Creates Cloud-Based Health Data Analysis Platform to Unleash the Imaginations of the World’s Best Scientific Minds* (UK Biobank, 2020); <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-creates-cloud-based-health-data-analysis-platform-to-unleash-the-imaginations-of-the-world-s-best-scientific-minds>
86. Jackson, K. A picture is worth a petabyte of data. *Science Node* (5 June 2019).
87. Nguyen, B. H. et al. Architecting datacenters for sustainability: greener data storage using synthetic DNA. In *Proc. Electronics Goes Green 2020* (ed. Schneider-Ramelow, F.) 105 (Fraunhofer, 2020).
88. *Seagate Product Sustainability* (Seagate, 2023); <https://www.seagate.com/gb/en/global-citizenship/product-sustainability/>
89. Madden, S. & Pollard, C. *Principles and Best Practices for Trusted Research Environments* (NHS England, 2021); <https://transform.england.nhs.uk/blogs/principles-and-practice-for-trusted-research-environments/>
90. Jones, K. H., Ford, D. V., Thompson, S. & Lyons, R. A profile of the SAIL Databank on the UK secure research platform. *Int. J. Popul. Data Sci.* **4**, 1134 (2020).
91. *About the Secure Research Service* (Office for National Statistics); <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/aboutthesecureresearchservice>
92. Shehabi, A. et al. *United States Data Center Energy Usage Report* Report no. LBNL-1005775, 1372902 (Office of Scientific and Technical Information, 2016); <http://www.osti.gov/servlets/purl/1372902/>
93. Masanet, E., Shehabi, A., Lei, N., Smith, S. & Koomey, J. Recalibrating global data center energy-use estimates. *Science* **367**, 984–986 (2020).
94. Caplan, T. *Help Us Advance Environmentally Sustainable Research* (Wellcome, 2022); <https://medium.com/wellcome-data/help-us-advance-environmentally-sustainable-research-3c11fe2a8298>
95. Choueiry, G. *Programming Languages Popularity in 12,086 Research Papers* (Quantifying Health, 2023); <https://quantifyinghealth.com/programming-languages-popularity-in-research/>
96. Pereira, R. et al. Ranking programming languages by energy efficiency. *Sci. Comput. Program.* **205**, 102609 (2021).
97. Lin, Y. & Danielsson, J. *Choosing a Numerical Programming Language for Economic Research: Julia, MATLAB, Python or R* (Centre for Economic Policy Research, 2022); <https://cepr.org/voxeu/columns/choosing-numerical-programming-language-economic-research-julia-matlab-python-or-r>
98. Appuswamy, R., Olma, M. & Ailamaki, A. Scaling the memory power wall with DRAM-aware data management. In *Proc. 11th International Workshop on Data Management on New Hardware* 1–9 (ACM, 2015); <https://doi.org/10.1145/2771937.2771947>
99. Guo, B., Yu, J., Yang, D., Leng, H. & Liao, B. Energy-efficient database systems: a systematic survey. *ACM Comput. Surv.* **55**, 111 (2022).
100. Karyakin, A. & Salem, K. An analysis of memory power consumption in database systems. In *Proc. 13th International Workshop on Data Management on New Hardware—DAMON ’17* 1–9 (ACM Press, 2017); <https://doi.org/10.1145/3076113.3076117>
101. Karyakin, A. & Salem, K. DimmStore: memory power optimization for database systems. *Proc. VLDB Endow.* **12**, 1499–1512 (2019).
102. Caset, F., Boussauw, K. & Storme, T. Meet & fly: sustainable transport academics and the elephant in the room. *J. Transp. Geogr.* **70**, 64–67 (2018).
103. Govaart, G. H., Hofmann, S. M. & Medawar, E. The sustainability argument for open science. *Collabra Psychol.* **8**, 35903 (2022).
104. Cockrell, H. C. et al. Environmental impact of telehealth use for pediatric surgery. *J. Pediatr. Surg.* **57**, 865–869 (2022).
105. Alshqaqeeq, F., McGuire, C., Overcash, M., Ali, K. & Twomey, J. Choosing radiology imaging modalities to meet patient needs with lower environmental impact. *Resour. Conserv. Recycl.* **155**, 104657 (2020).
106. *Sustainability Annual Report 2020–2021* (NHS, 2021); <https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/sustainability/sustainability-reports/sustainability-annual-report-2020-21>
107. *UNESCO Recommendation on Open Science* (UNESCO, 2021); <https://en.unesco.org/science-sustainable-future/open-science/recommendation>
108. Samuel, G. & Richie, C. Reimagining research ethics to include environmental sustainability: a principled approach, including a case study of data-driven health research. *J. Med. Ethics* <https://doi.org/10.1136/jme-2022-108489> (2022).

Acknowledgements

L.L. was supported by the University of Cambridge MRC DTP (MR/S502443/1) and the BHF program grant (RG/18/13/33946). M.I. was supported by the Munz Chair of Cardiovascular Prediction and Prevention and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014; NIHR203312). M.I. was also supported by the UK Economic and Social Research 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; NIHR203312). 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 the British Heart Foundation and Wellcome.

Author contributions

L.L. conceived and coordinated the manuscript. M.I. organized and edited the manuscript. All authors contributed to the writing and revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Loïc Lannelongue.

Peer review information *Nature Computational Science* thanks Bernabe Dorronsoro and Kirk Cameron for their contribution to the peer review of this work. Primary Handling Editors: Kaitlin McCardle and Ananya Rastogi, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2023