UNIVERSITY OF CAMBRIDGE
Department of Public Health and Primary Care

British Heart Foundation

NIHR | Cambridge Biomedical Research Centre

HDRUK Health Data Research UK

Baker HEART & DIABETES INSTITUTE

Software Sustainability Institute

Jesus College Cambridge

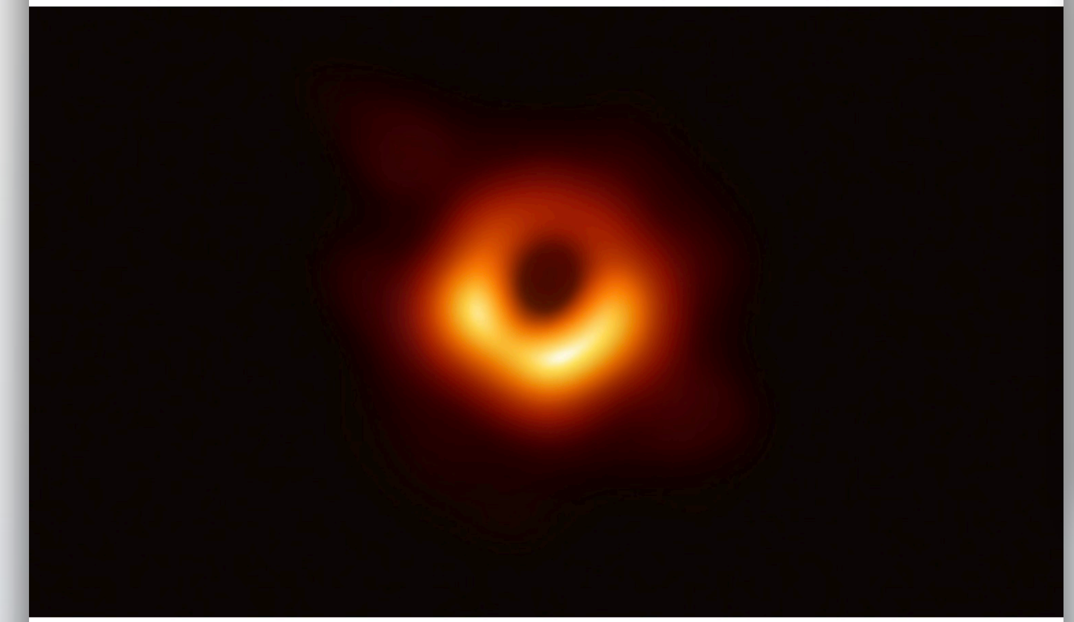
TACKLING THE HIDDEN COSTS OF COMPUTING: GREENER PRINCIPLES FOR SUSTAINABLE RESEARCH

Loïc Lannelongue, PhD
Research Associate, University of Cambridge
Software Sustainability Institute Fellow
Post-Doctoral Associate, Jesus College, Cambridge

LL582@medschl.cam.ac.uk | in | @Loic_Lnlg | www.lannelongue.eu

HIGH-PERFORMANCE COMPUTING HAS ENABLED AMAZING DISCOVERIES

NATIONAL GEOGRAPHIC LOGIN Renew SUBSCRIBE MENU



The Event Horizon Telescope—a planet-scale array of ground-based radio telescopes—has obtained the first image of a supermassive black hole and its shadow. The image reveals the central black hole of Messier 87, a massive galaxy in the Virgo cluster.
PHOTOGRAPH BY EVENT HORIZON TELESCOPE COLLABORATION

SCIENCE | STARSTRUCK

First-ever picture of a black hole unveiled

Event Horizon Telescope

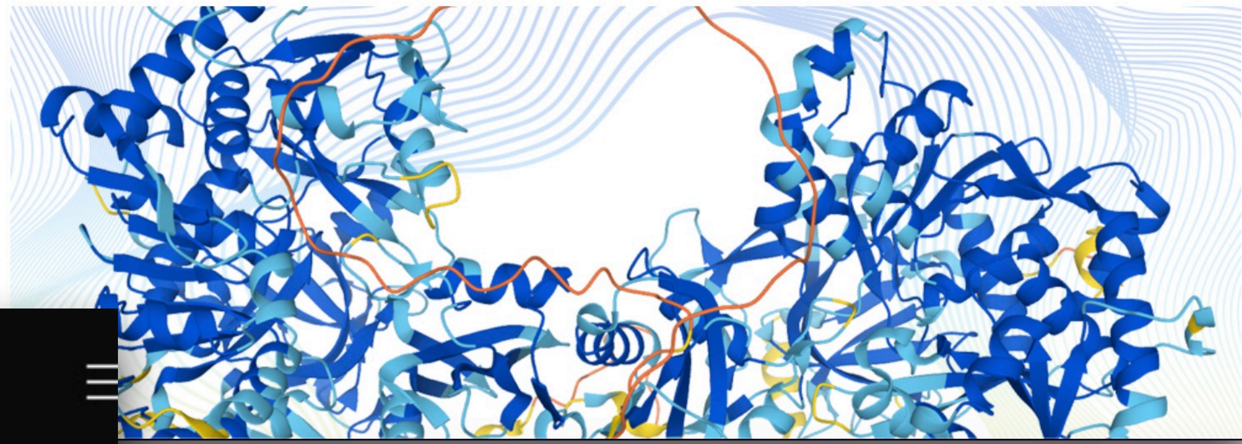
Astronomers reveal first image of the black hole at the heart of our galaxy

May 12, 2022

DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins

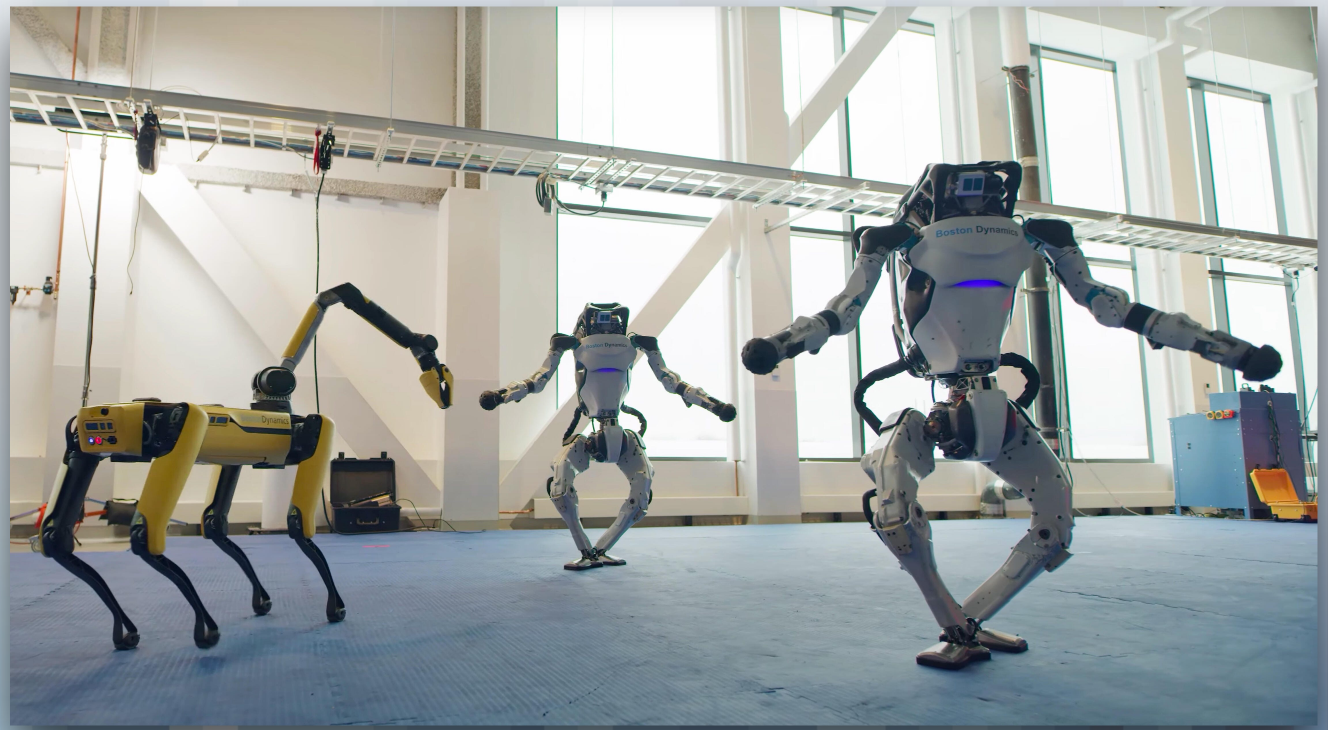
This article is also available in Deutsch, Français, Italiano and Español.

Partners use AlphaFold, the AI system recognised last year as a solution to the protein structure prediction problem, to release more than 350,000 protein structure predictions including the entire human proteome to the scientific community.



Climate Change AI

Climate Change AI is a global initiative to catalyze impactful work at the intersection of climate change and machine learning.



and much more

...INCLUDING IN FIGHTING CLIMATE CHANGE

Climate Change AI

Climate Change AI is a global initiative to catalyze impactful work at the intersection of climate change and machine learning.

Met Office

NEWS

Up to £1.2billion for weather and climate supercomputer

- More sophisticated rainfall predictions, helping the Environment Agency rapidly deploy mobile flood defences
- Better forecasting at airports so they can plan for potential disruption; and
- More detailed information for the energy sector to help them mitigate against potential energy blackouts and surges

Research | Blog | Impact | Safety & Ethics | About | Careers

Research

Machine learning can boost the value of wind energy

February 26, 2019

Carbon-free technologies like renewable energy
but many of them have not reached their full pot
over the past decade, wind farms have become a
free electricity as the cost of turbines has plumm
However, the variable nature of wind itself makes
source—less useful than one that can reliably de

...BUT AT A COST: **DATA CENTRES**

Global GHG emissions of data centres

~126 Mt of CO_2e/year

Equivalent to American commercial aviation

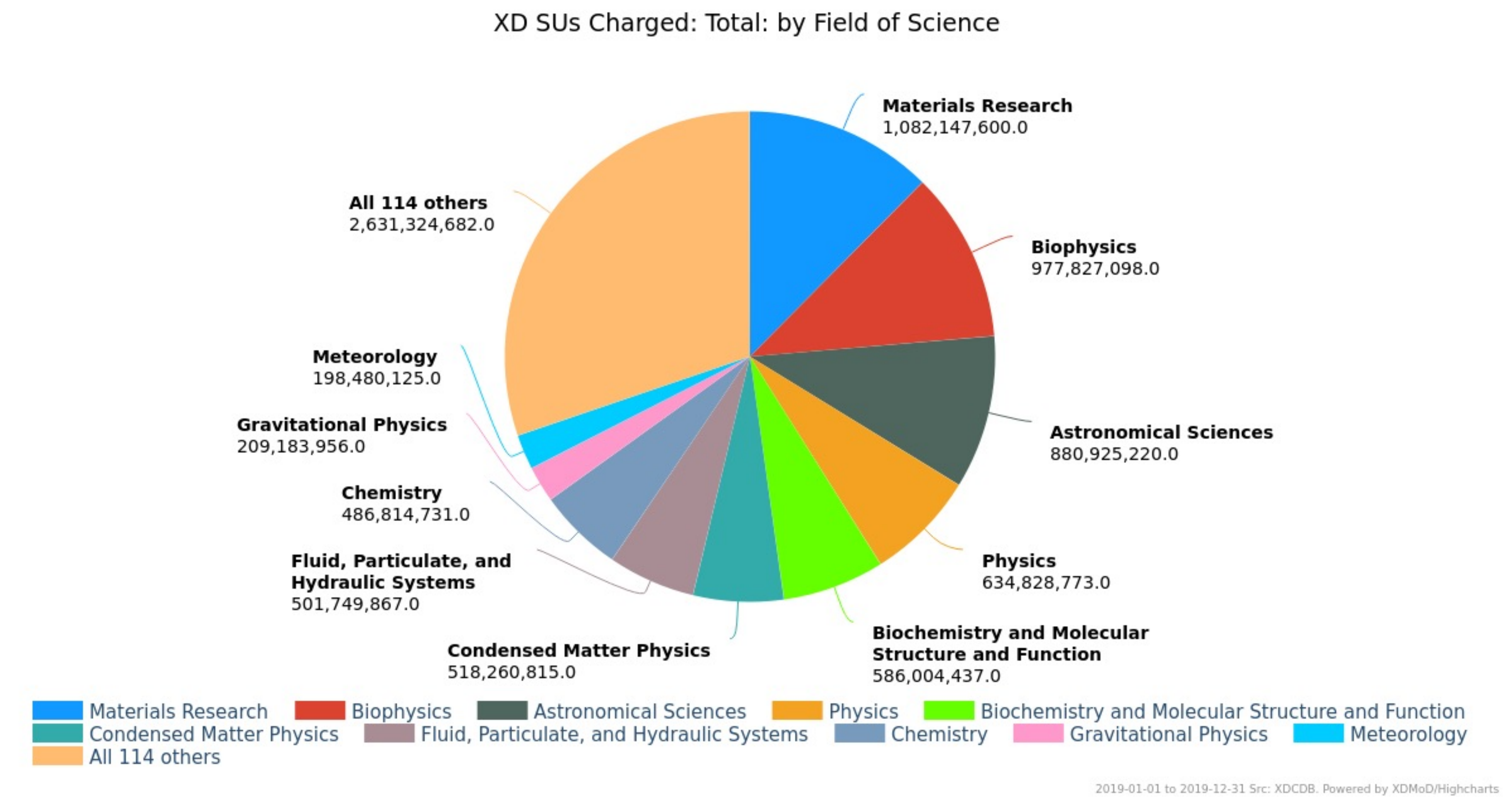
Malmodin et al. 2024
"ICT sector electricity consumption and greenhouse gas emissions – 2020 outcome"

https://www.epa.gov/system/files/documents/2024-05/420f24022.pdf

DATA CENTRES AND SCIENCE

XSEDE: network of institutional data centres in the US

In 2020:
9 billion compute hours
24 million h/day



https://portal.xsede.org/#/gallery/total_su_last_year

WHAT ABOUT THE SCIENCE WE DO? AI

ESTIMATING THE CARBON FOOTPRINT OF BLOOM, A 176B PARAMETER LANGUAGE MODEL

Alexandra Sasha Luccioni
Hugging Face
sasha.luccioni@hf.co

Sylvain Viguier
Graphcore
sylvainv@graphcore.ai

Anne-Laure Ligozat
LISN & ENSIIE
anne-laure.ligozat@lisn.upsaclay.fr

Model name	Number of parameters	Datacenter PUE	Carbon intensity of grid used	Power consumption	CO_2eq emissions	CO_2eq emissions × PUE
GPT-3	175B	1.1	429 gCO_2eq/kWh	1,287 MWh	*502 tonnes*	552 tonnes
Gopher	280B	1.08	330 gCO_2eq/kWh	*1,066 MWh*	*352 tonnes*	380 tonnes
OPT	175B	*1.09* [2]	*231 gCO_2eq/kWh*	*324 MWh*	70 tonnes	*76.3 tonnes* [3]
BLOOM	176B	1.2	57 gCO_2eq/kWh	433 MWh	25 tonnes	30 tonnes

UNFORTUNATELY…

Computing is **not** free

But from a **user** point of view, **it may look like it**

FOCUSING ON COMPUTING

<u>Day-to-day computing</u>

Emails

Writing on Words

Web surfing

Zoom

<u>Intense computations:</u>

long runtimes (several hours)

and/or large memory requirements (10s GB)

MORE FIELDS ARE NOW TACKLING THE ISSUE

ACL Anthology FAQ Corrections Submissions Search...

Energy and Policy Considerations for Deep Learning in NLP

Emma Strubell, Ananya Ganesh, Andrew McCallum

Abstract

Recent progress i... networks trained ... However, these a... necessitate simila... financially, due to... footprint required... researchers by qu... neural network m... improve equity in...

arXiv.org > cs > arXiv:1910.09700 Search... Help | Advance

Computer Science > Computers and Society

[Submitted on 21 Oct 2019 (v1), last revised 4 Nov 2019 (this version, v2)]

Quantifying the Carbon Emissions of Machine Learning

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, Thomas Dandres

From an environmental standpoint, there are a few crucial aspects of training a neural network that have a major impact on the quantity of carbon that it emits. These factors include: the location of the server used for training and the energy grid that it uses, the length of the training procedure, and even the make and model of hardware on which the training takes place. In order to approximate these emissions, we present our Machine Learning Emissions Calculator, a tool for our community to better understand the environmental impact of training ML models. We accompany this tool with an explanation of the factors cited above, as well as concrete actions that individual practitioners and organizations can take to mitigate their carbon

arXiv.org > cs > arXiv:1907.10597 Search... Help | Advance

Computer Science > Computers and Society

[Submitted on 22 Jul 2019 (v1), last revised 13 Aug 2019 (this version, v3)]

Green AI

Roy Schwartz, Jesse Dodge, Noah A. Smith, Oren Etzioni

The computations required ... [2]. These computations have a surprisingly large carbon ... ial cost of the computations can make it difficult for ac... This position paper advoc... n, we propose reporting the financial cost or "price tag ... to make AI both greener and more inclusive---enabling ... for AI.

ARTICLE OPEN ACCESS

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

Authors: Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, Shmargaret Shmitchell

Authors Info & Affiliations

Publication: FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accou... 2021 • Pages 610–623 • https://doi.org/10.1145/3442188.3445922

Thinking Geographically about AI Sustainability

Meilin Shi[1], Kitty Currier[2], Zilong Liu[1], Krzysztof Janowicz[1,2], Nina Wiedemann[3], Judith Verstegen[4], Grant McKenzie[5], Anita Graser[6], Rui Zhu[7], and Gengchen Mai[8]

arXiv > cs > arXiv:2206.06370

Computer Science > Computers and Society

[Submitted on 12 Jun 2022]

Don't "research fast and break things": On the ethics of Computational Social Science

David Leslie

arXiv > physics > arXiv:2203.12389

Physics > Physics and Society

[Submitted on 23 Mar 2022 (v1), last revised 23 Aug 2022 (this version, v2)]

Climate impacts of particle physics

Kenneth Bloom, Veronique Boisvert, Daniel Britzger, Micah Buuck, Astrid Eichhorn, Michael Headley, Kristin Lohwasser, Petra Merkel

nature astronomy

Explore content ⌄ Journal information ⌄ Publish with us ⌄

nature > nature astronomy > comment > article

Comment | Published: 10 September 2020

The ecological impact of high-performance computing in astrophysics

Simon Portegies Zwart

Nature Astronomy 4, 819–822 (2020) | Cite this article

1835 Accesses | 10 Citations | 500 Altmetric | Metrics

Computer use in astronomy continues to increase, and so also its impact on the environment. To minimize the effects, astronomers should avoid interpreted scripting languages such as Python, and favour the optimal use of energy-efficient workstations.

OHBM Sustainability & Environmental Action

Special Interest Group

Keep IT Green

KIG: a tool for Carbon footprint monitoring in physics research

Francesco Minarini, PhD student in Physics @University of Bologna, Italy and INFN

research area: Green Computing for High Energy Physics

International Symposium on Grids & Clouds (ISGC) 2023, Academia Sinica, Taipei, Taiwan.

03/24/2023

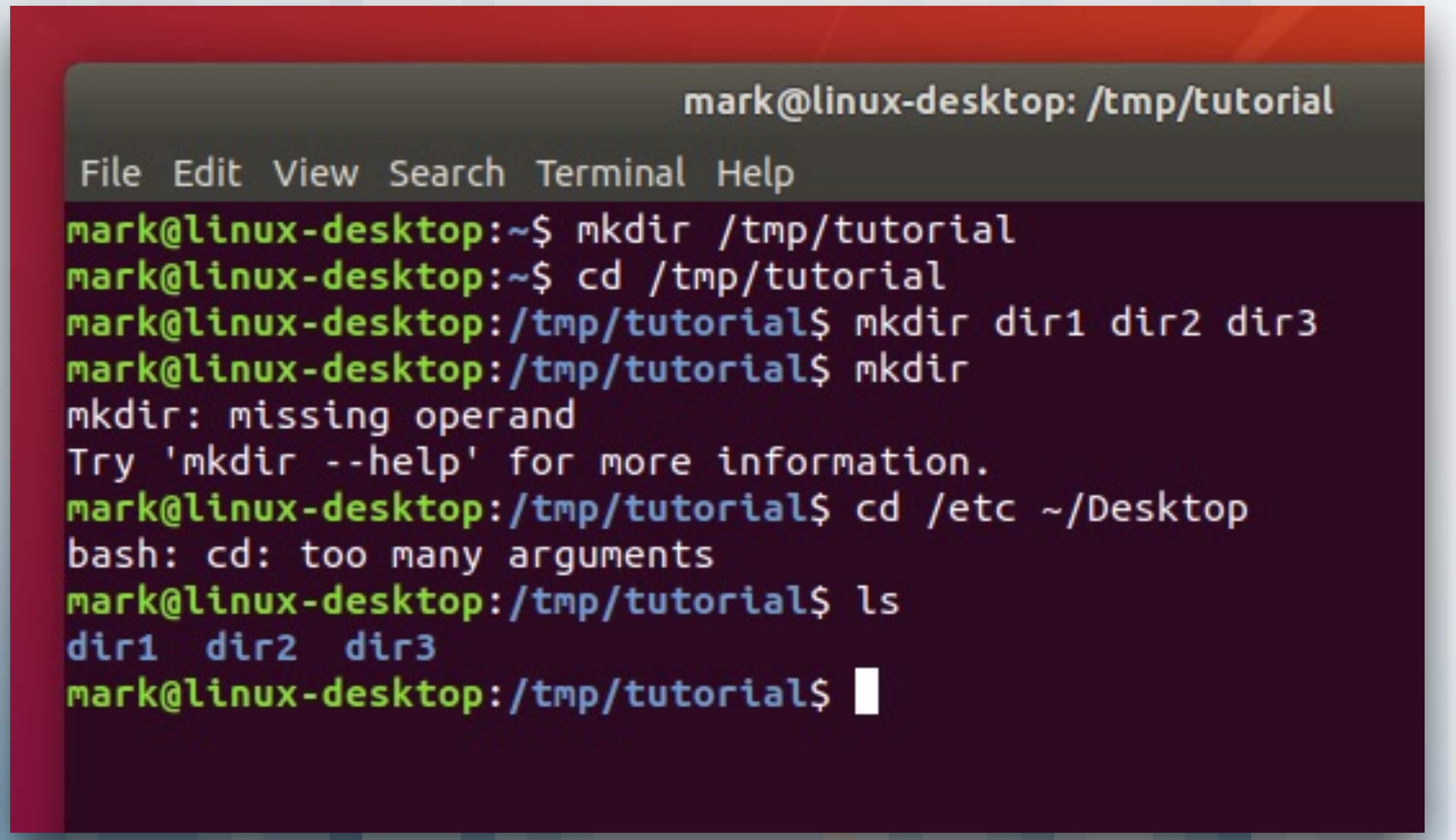
NEUROVIEW | VOLUME 106, ISSUE 1, P17-20, APRIL 08, 2020

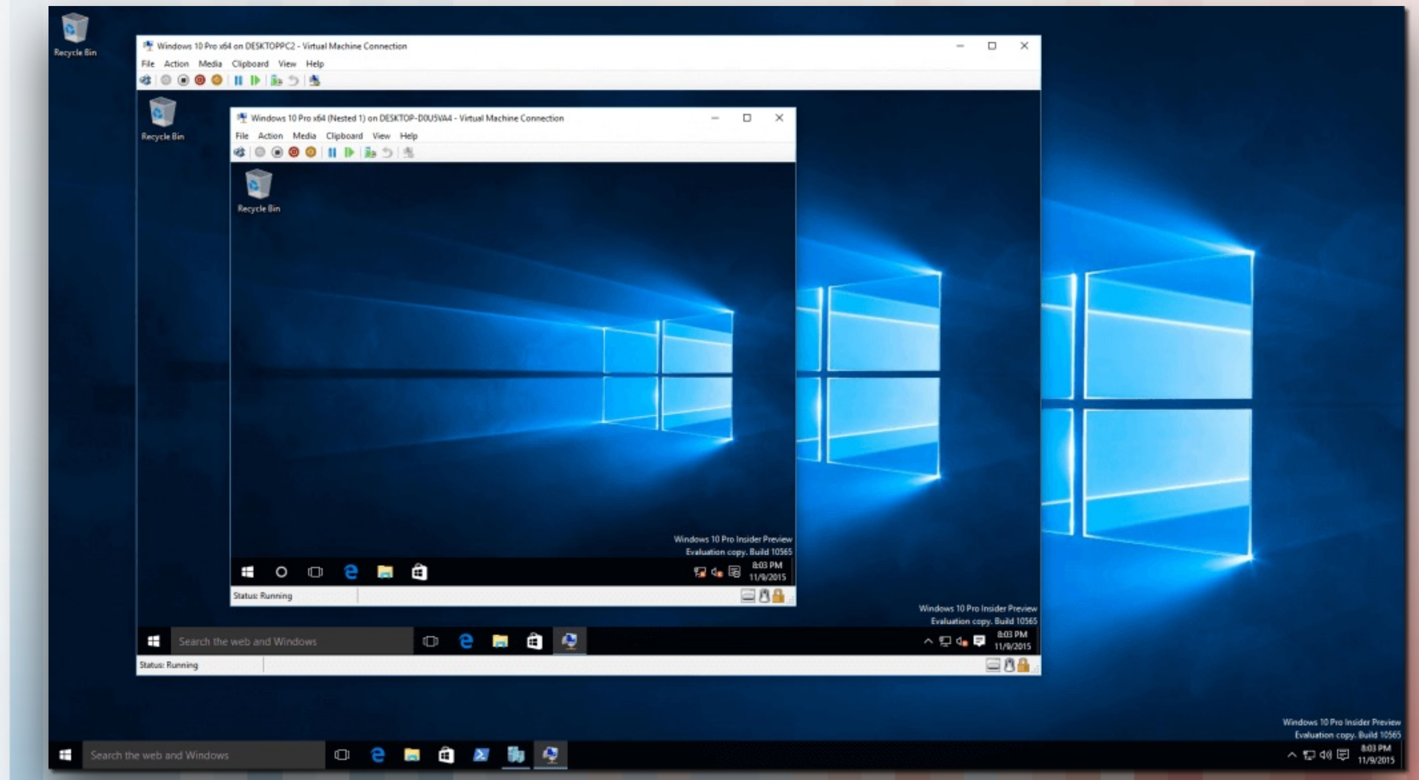
How Can Neuroscientists Respond to the Climate Emergency?

Adam R. Aron • Richard B. Ivry • Kate J. Jeffery • ... Robert Schmidt • Christopher Summerfield • Anne E. Urai • Show all authors

Open Access • DOI: https://doi.org/10.1016/j.neuron.2020.02.019 • Check for updates

IT'S ALL THE SAME (ISH)





BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

70-98% of the cradle-to-grave impact is from production (consumer devices)

15-60% for servers in data centres

Manufacturing/EoL impacts largely underestimated / underreported on

Keep, Repair, Reuse

BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

>82% of the 54m of tonnes of e-waste are handled by 12-56m informal waste workers worldwide

18m children work in industries involving waste processing

E-waste are predicted to raise by 40% by 2030

BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

HAL open science

A comprehensive review of the end-of-life modeling in LCAs of digital equipment

Marion Ficher, Tom Bauer, Anne-Laure Ligozat

"No clear consensus exists on modeling EoL in an LCA."

"Most studies employ a substitution approach with recycling and avoided impacts assessment."

"The substitution approach leads to several limitations: invisibilization of environmental impacts of EoL treatments and underestimations of potential environmental burdens."

BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

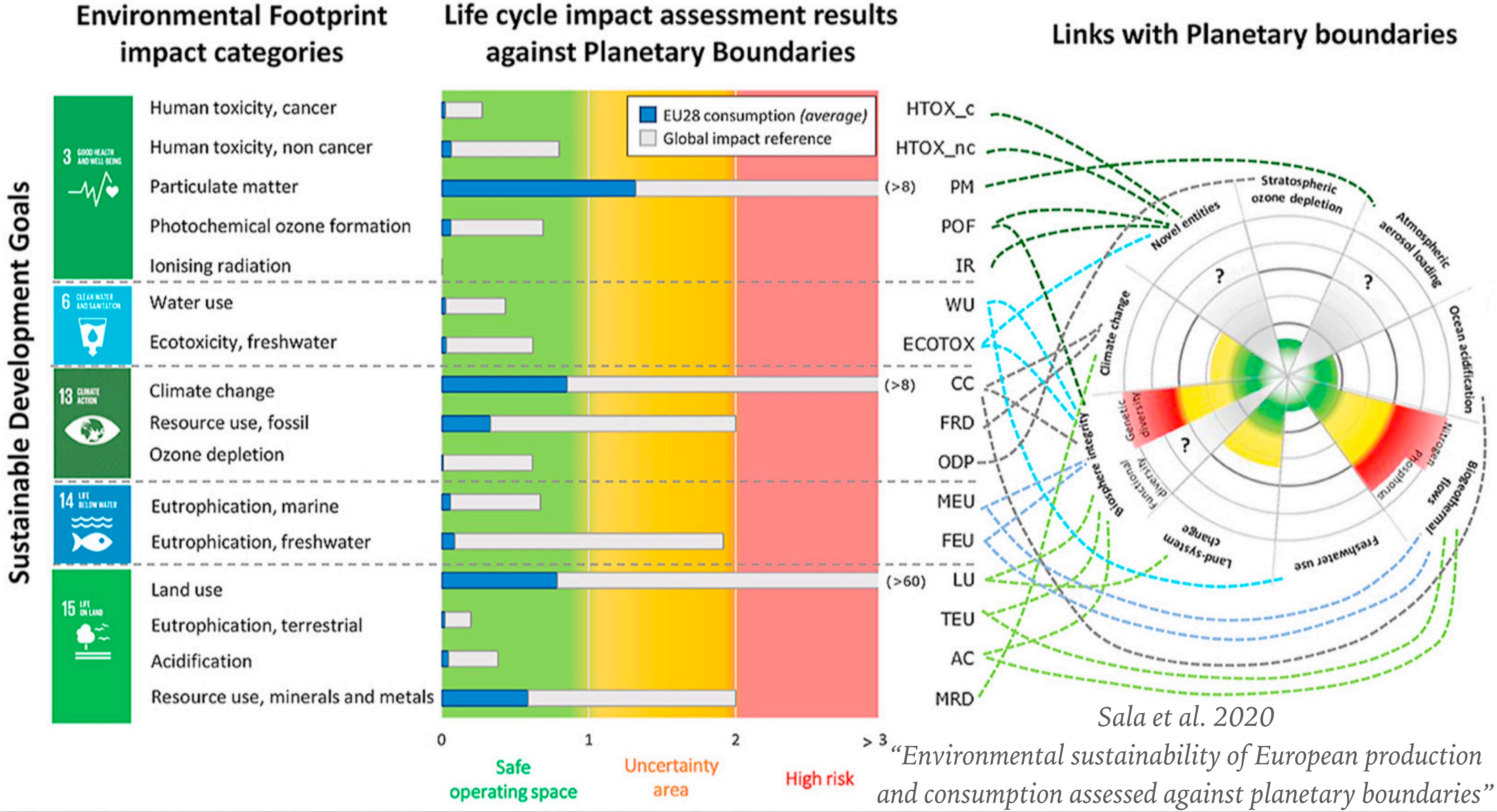
"A medium-sized data centre (15 MW) uses as much water as **three average-sized hospitals**"

Perspective | Open access | Published: 15 February 2021

Data centre water consumption

David Mytton

npj Clean Water **4**, Article number: 11 (2021) | Cite this article



Sala et al. 2020

“Environmental sustainability of European production and consumption assessed against planetary boundaries”

BREAKING DOWN THE ENVIRONMENTAL IMPACTS OF COMPUTING

Embodied
vs
use-stage

Carbon footprint
vs
broader environmental impacts

Computing
vs
storage

Energy bill of a data centre:

- *50% servers*
- *10% storage*
- *40% overheads (cooling)*

Storage
~10 $kgCO_2e$/TB/year

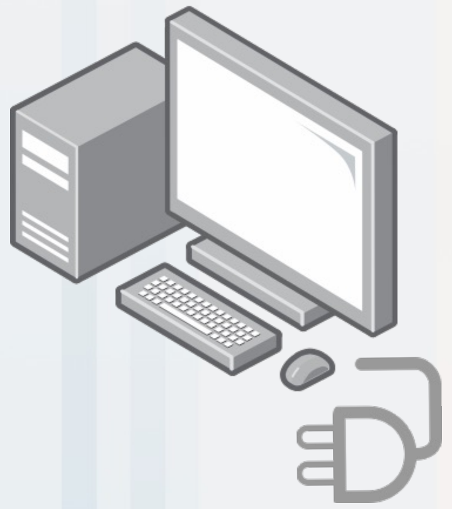
Don't store useless data

https://www.seagate.com/gb/en/global-citizenship/product-sustainability/

THE CARBON FOOTPRINT OF COMPUTATION

Carbon footprint = energy used x carbon intensity

gCO_2e $\quad$ kWh $\quad$ gCO_2e/kWh



THE CARBON FOOTPRINT OF COMPUTATION: **ENERGY NEEDED**

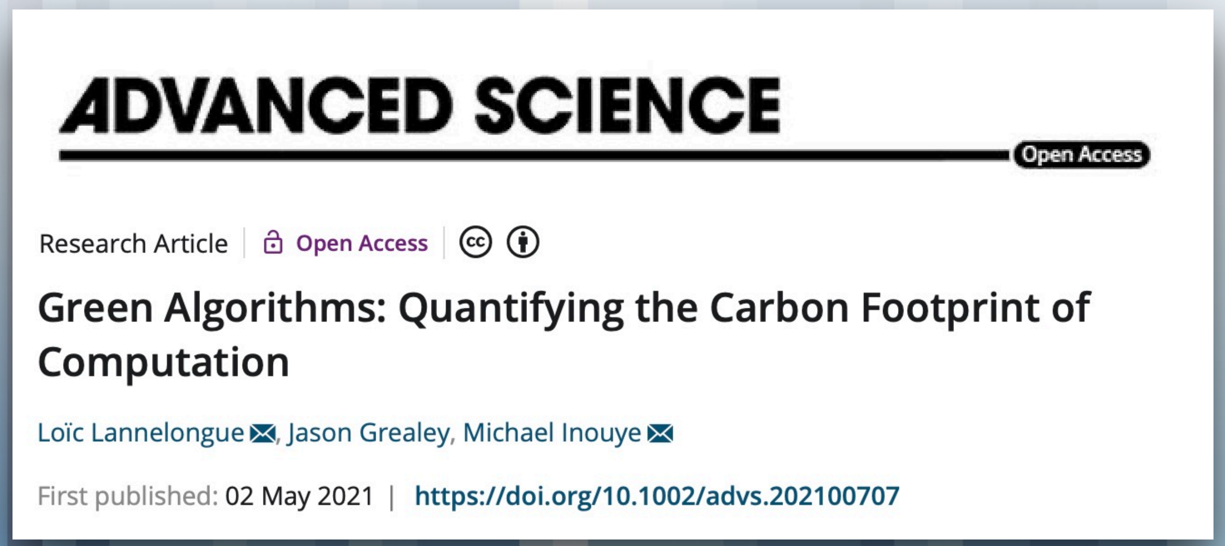
$$E = t \times (P_c + P_m) \times PUE$$

Running time (h)

Power draw of processing cores (W)

Power draw from memory (W)

Efficiency of the data centre


ADVANCED SCIENCE Open Access

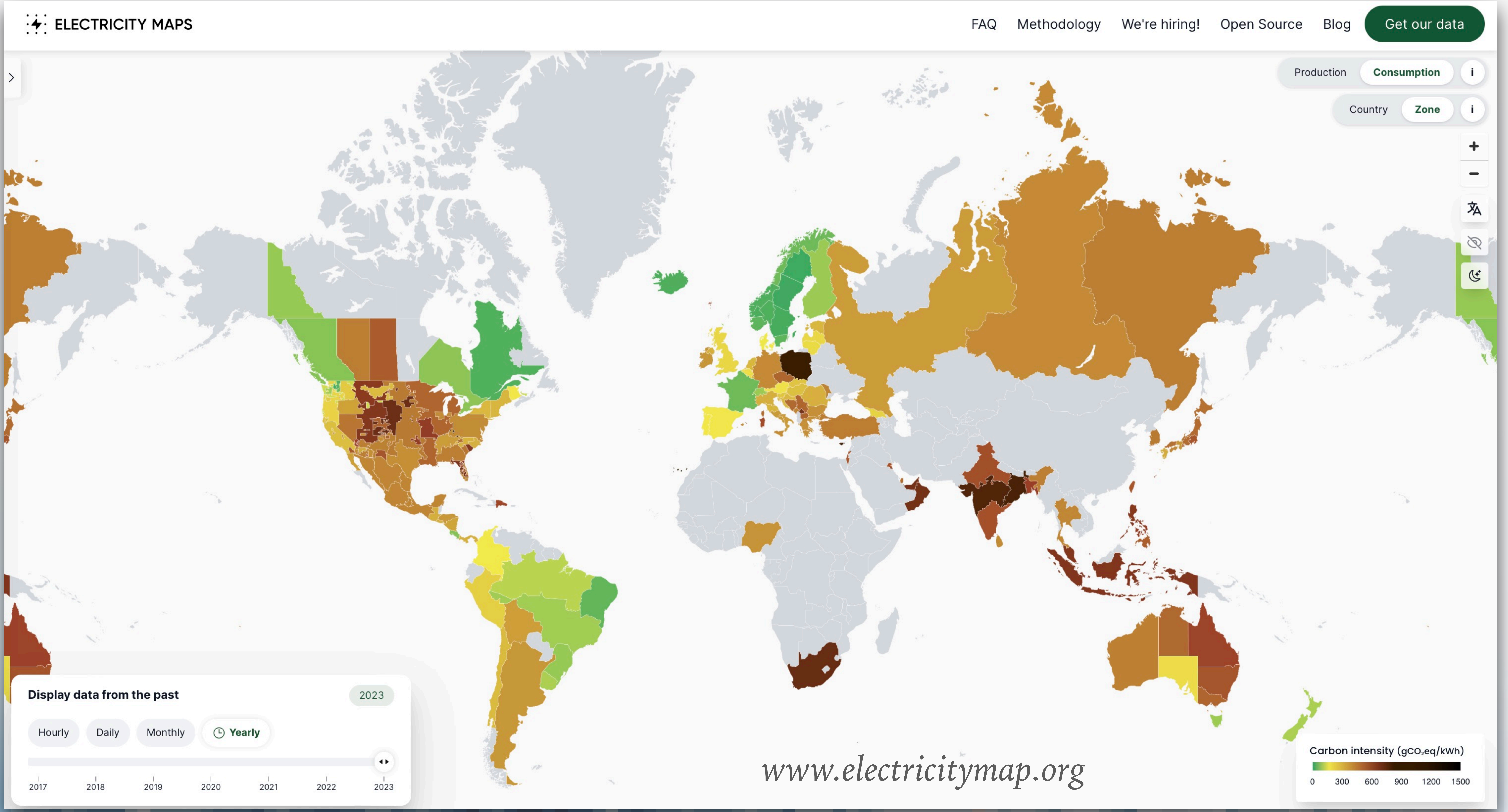
Research Article | Open Access

Green Algorithms: Quantifying the Carbon Footprint of Computation

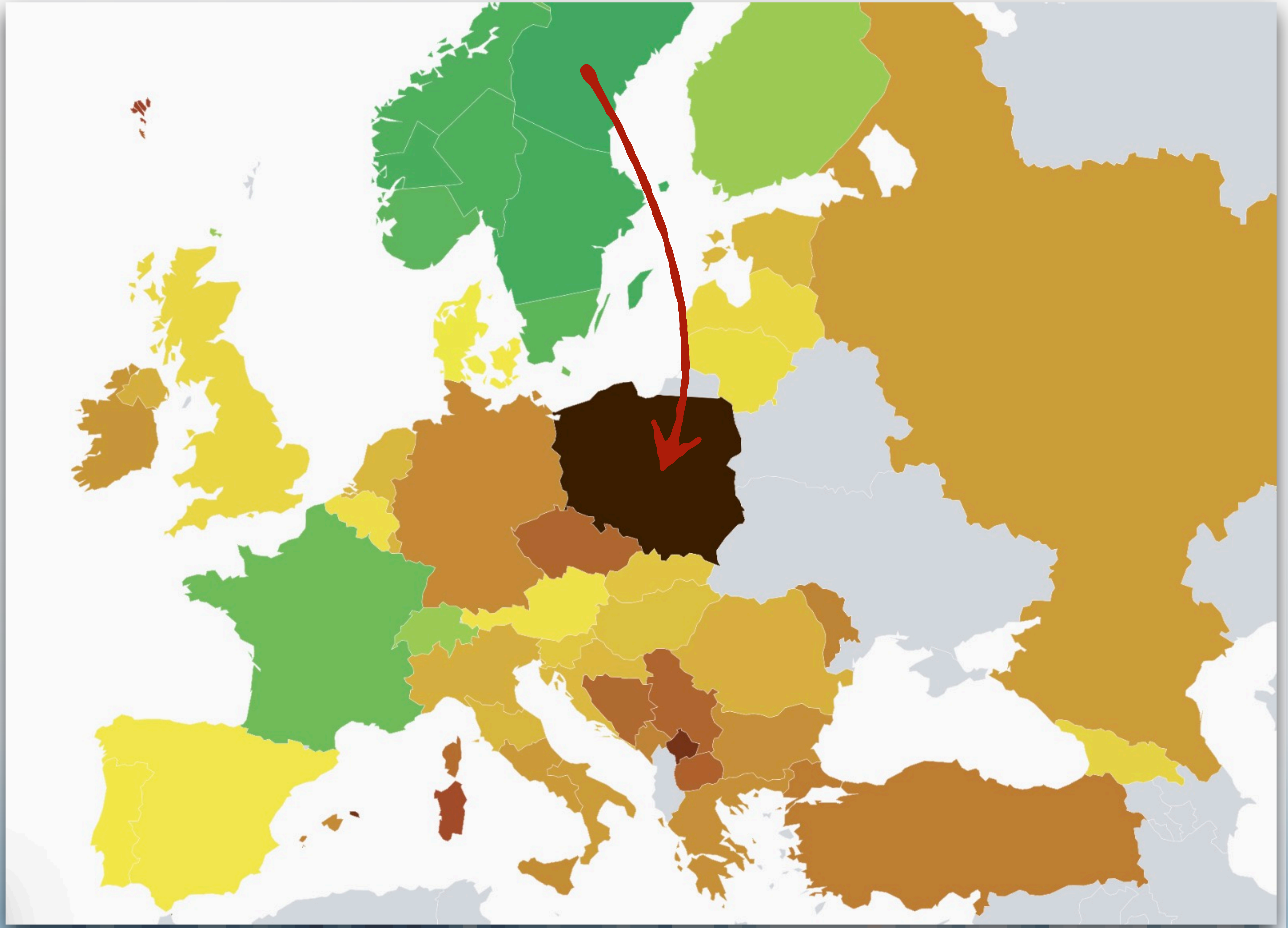
Loïc Lannelongue, Jason Grealey, Michael Inouye

First published: 02 May 2021 | https://doi.org/10.1002/advs.202100707

THE CARBON FOOTPRINT OF COMPUTATION: **CARBON INTENSITY**

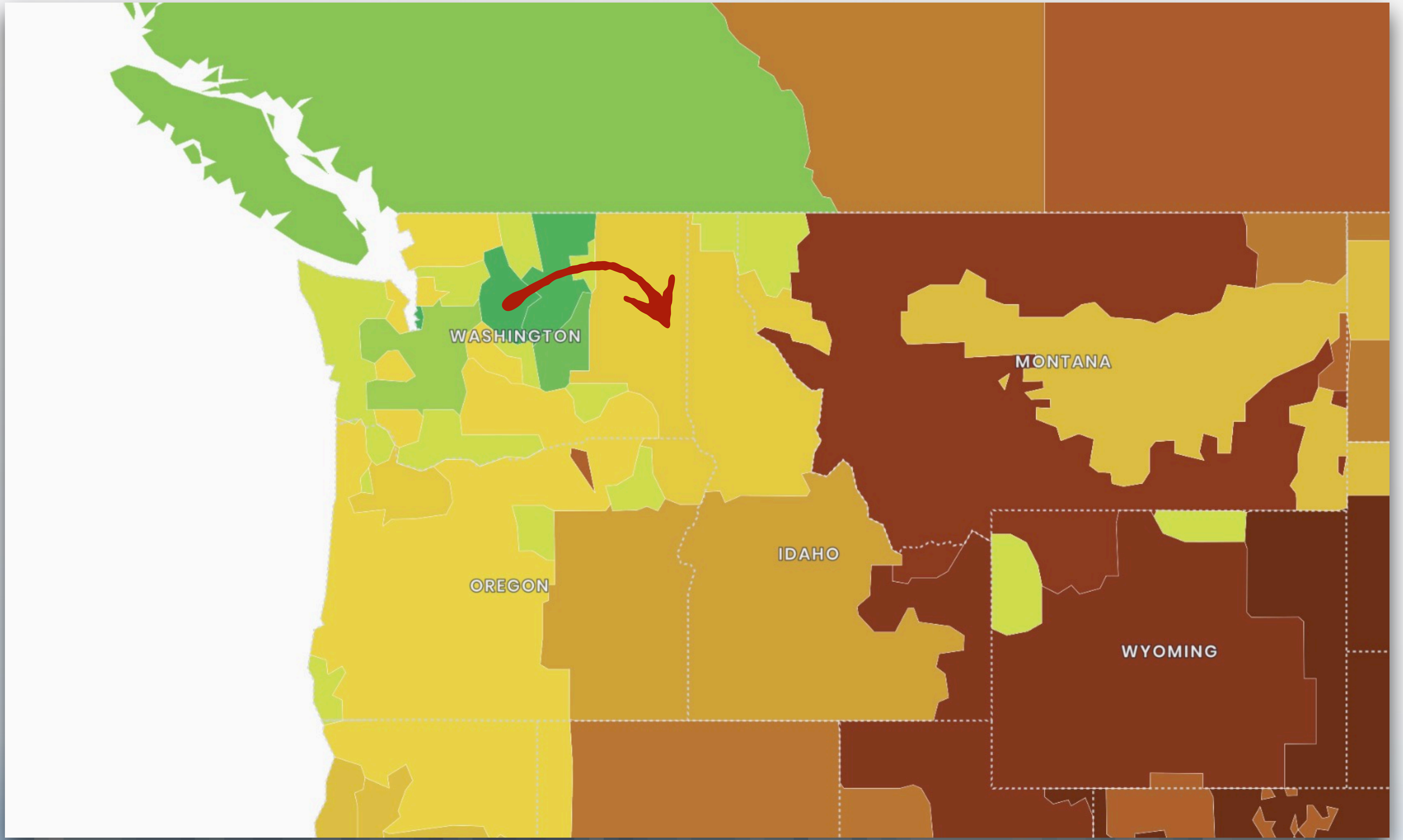


THE CARBON FOOTPRINT OF COMPUTATION: **CARBON INTENSITY**



x46

THE CARBON FOOTPRINT OF COMPUTATION: **CARBON INTENSITY**



x9

BUT…

It doesn't mean we should stop doing science

Synergies exist

GREENER PRINCIPLES FOR SUSTAINABLE COMPUTATIONAL SCIENCE

nature computational science

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature computational science > perspectives > article

Perspective | Published: 26 June 2023

GREENER principles for environmentally sustainable computational science

Loïc Lannelongue, Hans-Erik G. Aronson, Alex Bateman, Ewan Birney, Talia Caplan, Martin Juckes, Johanna McEntyre, Andrew D. Morris, Gerry Reilly & Michael Inouye

Nature Computational Science **3**, 514–521 (2023) | Cite this article

Abstract

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.

Collaboration with UK stakeholders

UNIVERSITY OF CAMBRIDGE
Department of Public Health and Primary Care

UK Research and Innovation

HDRUK
Health Data Research UK

EMBL-EBI

DARE UK

wellcome

GREENER PRINCIPLES FOR SUSTAINABLE COMPUTATIONAL SCIENCE

Governance

All actors in computational research have a key role to play and can lead the efforts towards sustainable computing.

Responsibility

Embracing both individual and institutional responsibility regarding the environmental impacts of research. This involves being transparent about these and initiating bold initiatives to reduce them.

Estimation

Monitoring environmental impacts to identify inefficiencies and opportunities for improvement.

Energy and embodied impacts

Minimising energy needs of computations and favouring low-carbon energy sources, while also considering the broader environmental impacts (e.g. water usage, mining of raw materials etc.).

New collaborations

Cooperating to leverage low-carbon infrastructures, facilitate equitable access to low-carbon computation and limit wasted resources.

Education

Training all stakeholders to be aware of the sustainability challenges of HPC and to be equipped with the skills to tackle them.

Research

Dedicate research efforts to green computing to improve our understanding of power usage, support sustainable software engineering and enable energy-efficient research.

Cultural change: make environmental sustainability a core element of research

GOVERNANCE AND RESPONSIBILITY

ALL **STAKEHOLDERS** HAVE A ROLE TO PLAY

Grassroots movements

Graduate students

Lab technicians

Postdocs

Undergrads

Bottom up ↑

↓ Top down

HPC teams

PIs

Training team

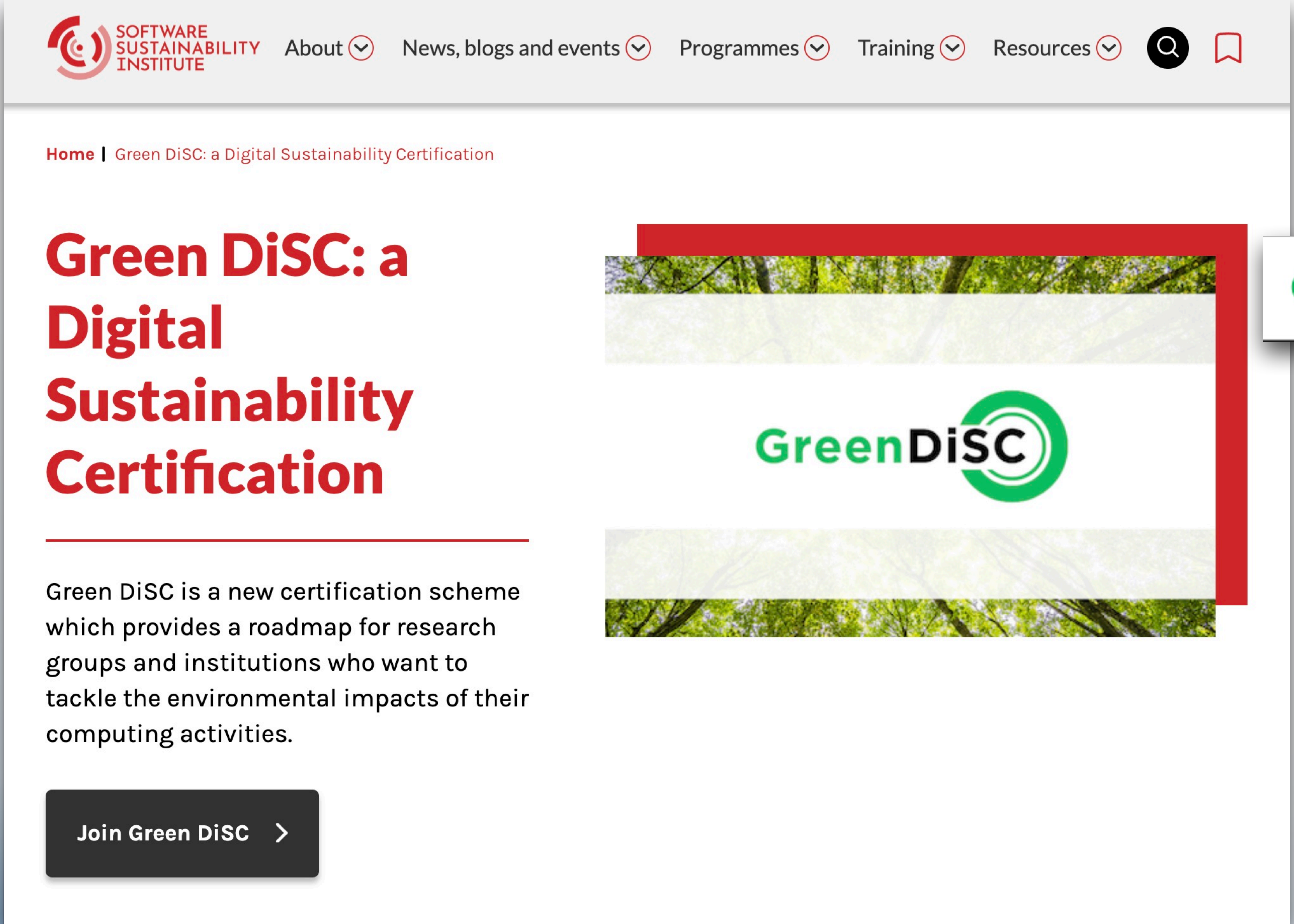
Journals

Funding bodies

Governments

Industry

GREEN DISC: A DIGITAL SUSTAINABILITY CERTIFICATION



info@greendisc.org

Subscribe for updates about the scheme:

ALL **STAKEHOLDERS** HAVE A ROLE TO PLAY: **GOVERNMENTS AND FUNDERS**

France 2030 : Lancement de l'appel à projets « Accélération des usages de l'IA générative dans l'économie »

5 Avril 2024 | Communiqué de presse

The Green Algorithms tools now used for all AI funding calls ran by the French government.

Le dossier **déposé comportera une auto-évaluation de l'impact environnemental du projet,** conformément aux critères énoncés dans le paragraphe « critères de sélection » ci-dessous**. Le dossier précisera obligatoirement la méthodologie utilisée**

pour parvenir aux résultats présentés dans le cadre de cette auto-évaluation. Le dossier précisera le pays de localisation des serveurs. Le calcul des diverses quantités rentrant dans l'évaluation de l'impact environnemental du système d'IA développé **s'appuiera obligatoirement sur l'outil Green Algorithms**[5] (GT, Lannelongue et al ; Jay et al.[6]), lorsque cet outil permet ce calcul.

Les efforts des porteurs de projets en matière d'écoconception, de maîtrise des consommations énergétiques et de ressources ainsi que de recyclabilité seront valorisés.

FROM THEORY TO PRACTICE

Estimating and reporting the carbon footprint of algorithms

ESTIMATING CARBON FOOTPRINTS **IN PRACTICE**

nature reviews methods primers

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature reviews methods primers › comment › article

Comment | Published: 16 February 2023

Carbon footprint estimation for computational research

Loïc Lannelongue & Michael Inouye

Nature Reviews Methods Primers **3**, Article number: 9 (2023) | Cite this article

187 Accesses | **23** Altmetric | Metrics

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.

Michael Inouye

EXISTING TOOLS

carbontracker

pypi v1.1.6 | python 3.8 | 3.9 | 3.10 | build passing | License MIT

About

carbontracker is a tool for tracking and predicting the energy consumption and carbon footprint of training deep learning models as described in Anthony et al. (2020).

experiment-impact-tracker

The experiment-impact-tracker is meant to be a simple drop-in method to track energy usage, carbon emissions, and compute utilization of your system. Currently, on Linux systems with Intel chips (that support the RAPL or powergadget interfaces) and NVIDIA GPUs, we record: power draw from CPU and GPU, hardware information, python package versions, estimated carbon emissions information, etc. In California we even support realtime carbon emission information by querying caiso.com!

Once all this information is logged, you can generate an online appendix which shows off this information like seen here:

https://breakend.github.io/RL-Energy-Leaderboard/reinforcement_learning_energy_leaderboard/pongnoframeskip-v4_experiments/ppo2_stable_baselines,_default_settings/0.html

CODE CARBON

What it is

A lightweight and easy-to-use Python pip package

Emissions tracked based on your power consumption & location-dependent carbon intensity

Cloud Carbon Footprint

Cloud Carbon Emissions Measurement and Analysis Tool

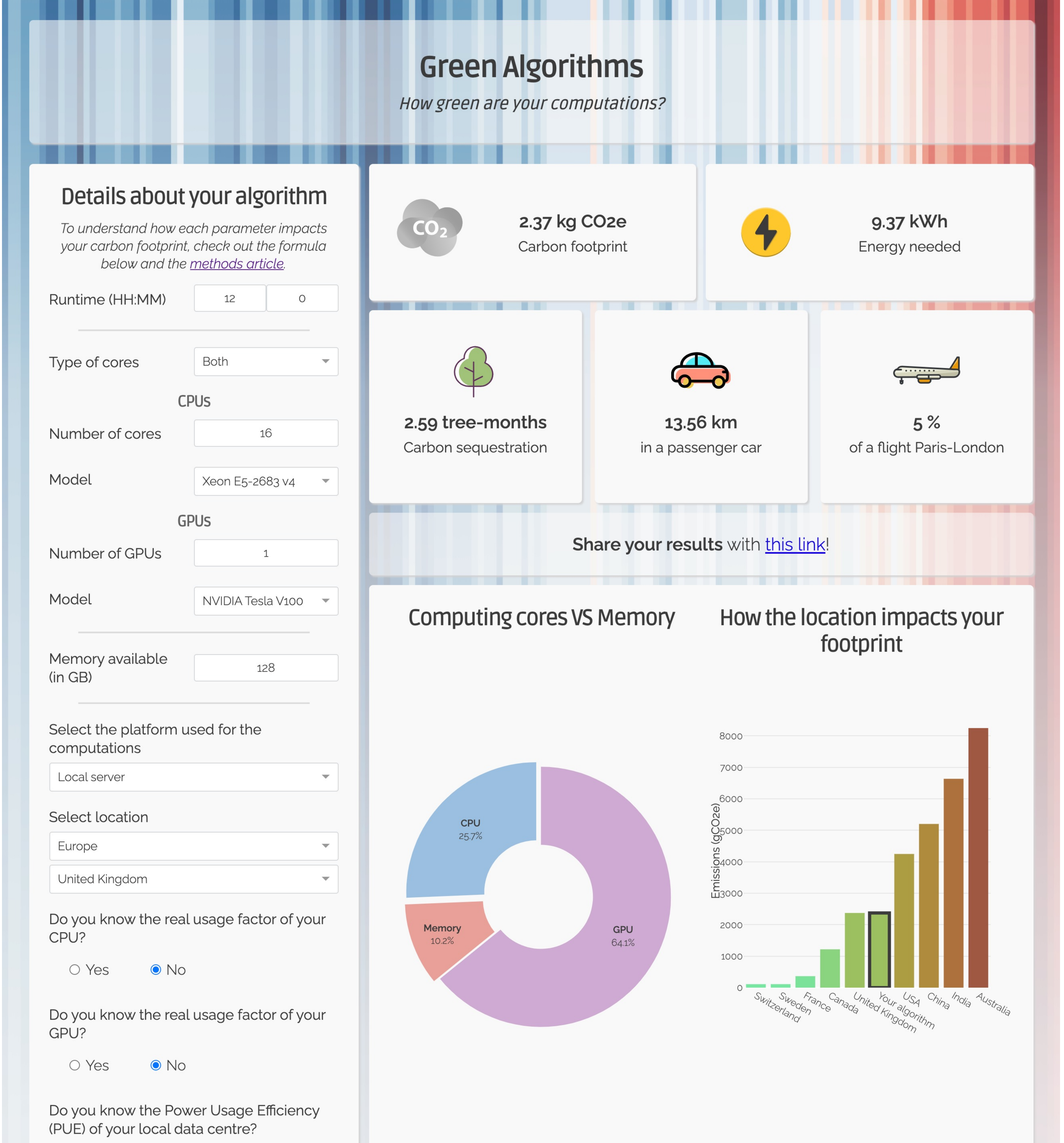
Understand how your cloud usage impacts our environment and what you can do about it

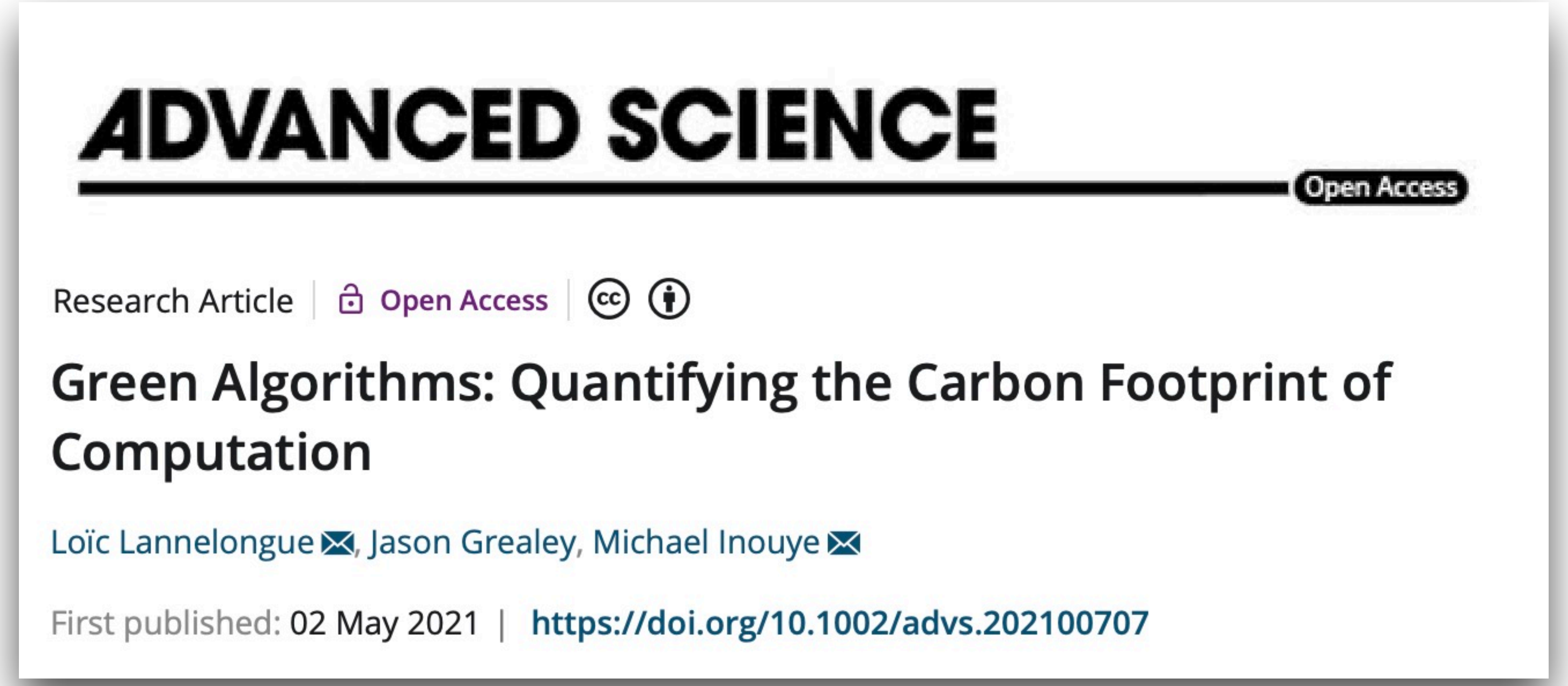
TRACARBON

CUMULATOR — a tool to quantify and report the carbon footprint of machine learning computations and communication in academia and healthcare

Trébaol, Tristan

2020





ADVANCED SCIENCE Open Access

Research Article | Open Access

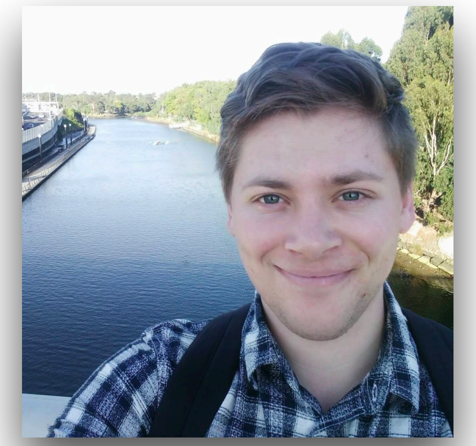
Green Algorithms: Quantifying the Carbon Footprint of Computation

Loïc Lannelongue, Jason Grealey, Michael Inouye

First published: 02 May 2021 | https://doi.org/10.1002/advs.202100707

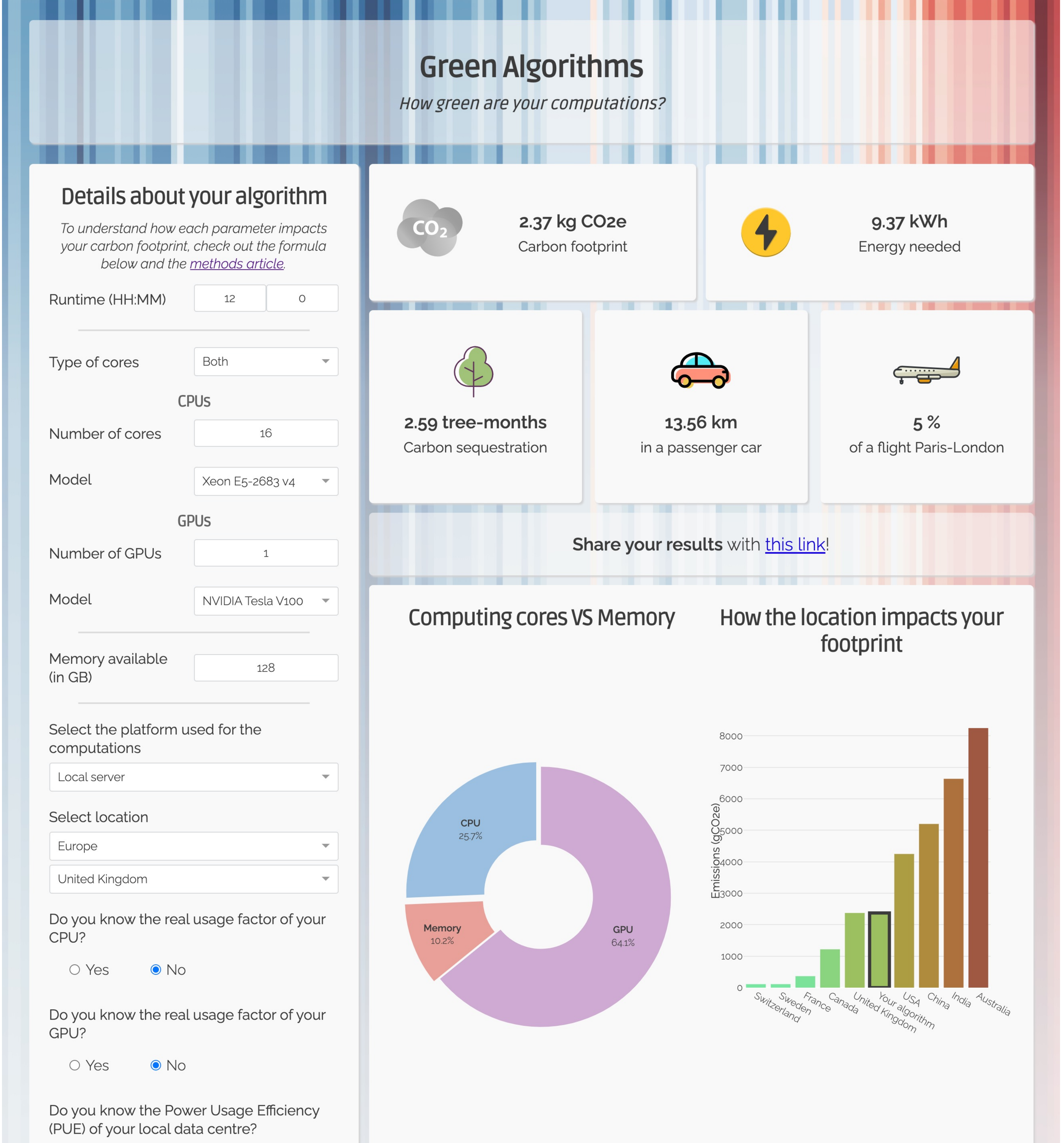
THE GREEN ALGORITHMS CALCULATOR

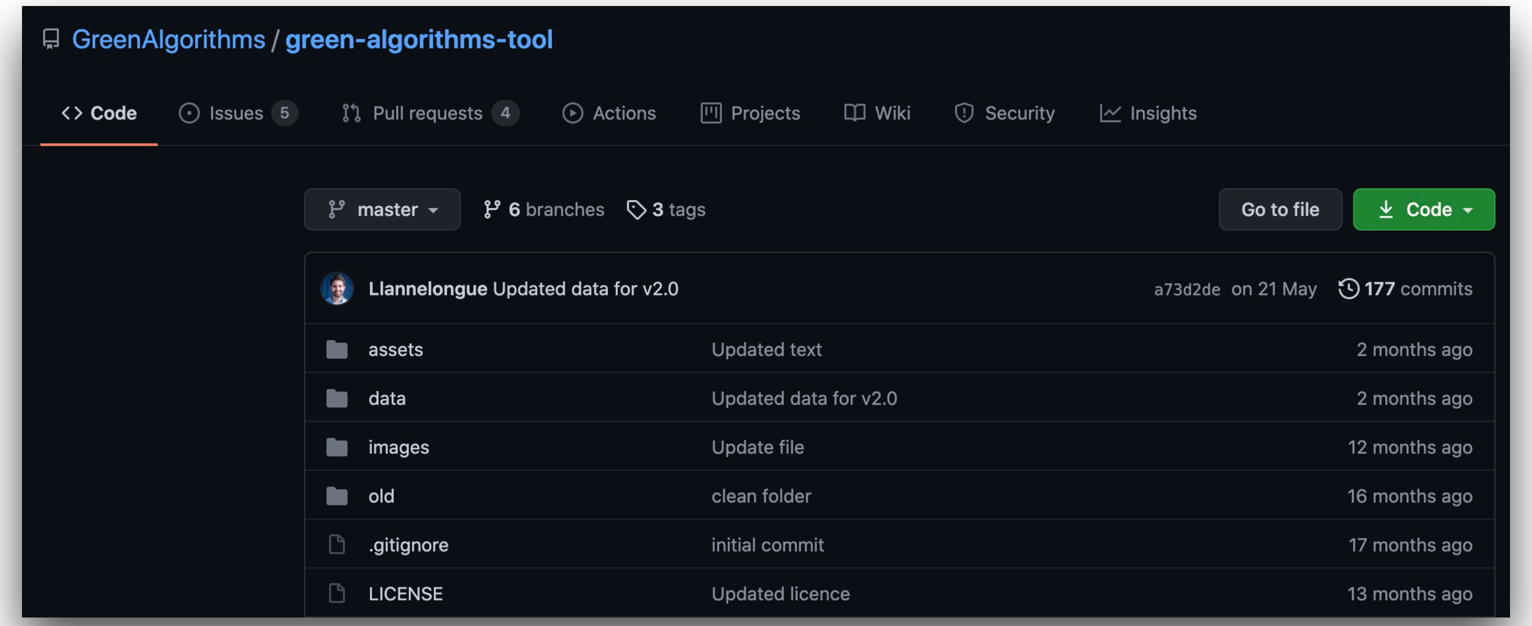
calculator.green-algorithms.org



Jason Grealey

Michael Inouye

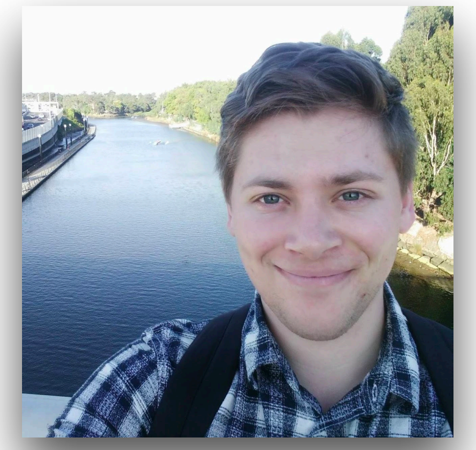




https://github.com/GreenAlgorithms/green-algorithms-tool

THE GREEN ALGORITHMS CALCULATOR

calculator.green-algorithms.org



Jason Grealey

Michael Inouye

GREEN ALGORITHMS 4 HPC

```
GreenAlgorithms4HPC]$ myCarbonFootprint.sh --STARTDAY 2020-01-01 --ENDDAY 2020-06-01
```

GREEN ALGORITHMS 4 HPC

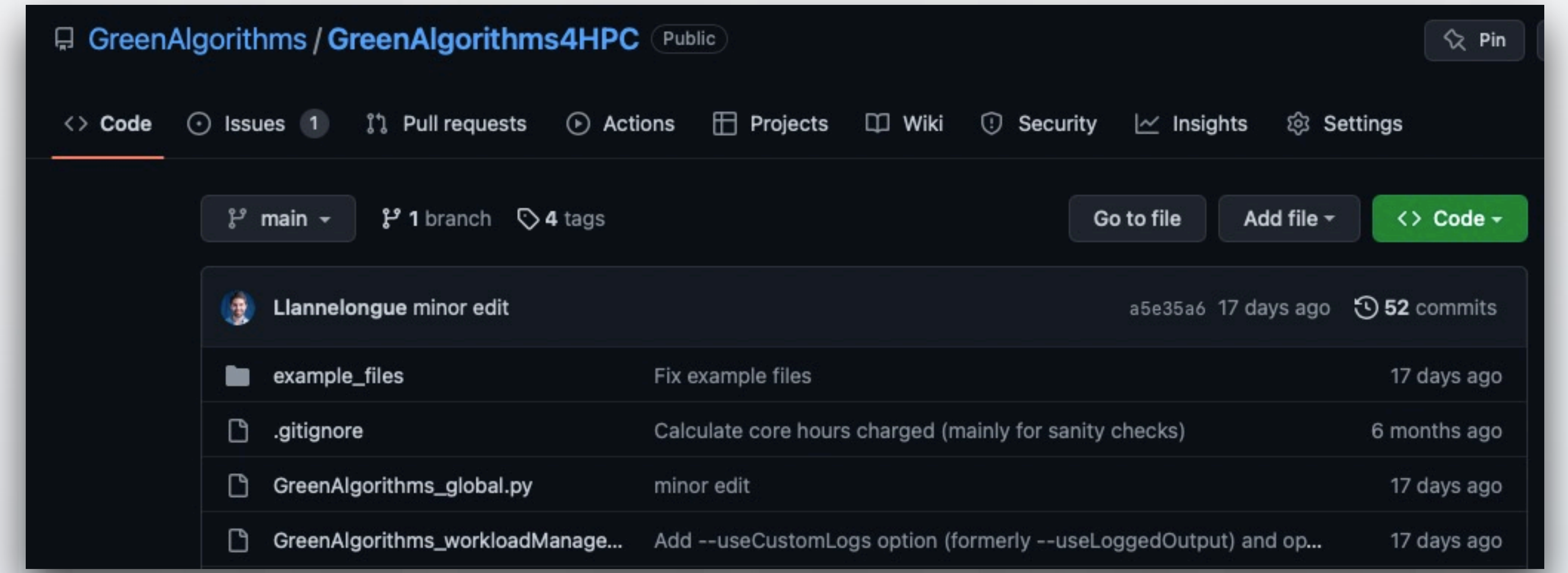
```
GreenAlgorithms4HPC]$ myCarbonFootprint.sh --STARTDAY 2020-01-01 --ENDDAY 2020-06-01
```

```
    ###################################
    #                                 #
    #  Your carbon footprint on CSD3  #
    #    (2021-01-01 / 2021-12-31)    #
    #                                 #
    ###################################

            ----------------
           |   222 kgCO2e   |
            ----------------

...This is equivalent to:
     - 20 tree-years
     - driving 1,268 km
     - 4.44 flights between Paris and London

...26.0% of your jobs failed, which represents a waste of 51 kgCO2e (55.26 tree-months).
...On average, you request at least 1.0 times the memory you need. By only requesting the memory you needed, you could have saved 0 gCO2e (0.00 tree-months).

Energy used: 960.17 kWh
     - CPUs: 88.91 kWh (9%)
     - GPUs: 713.81 kWh (74%)
     - Memory: 32.22 kWh (3%)
     - Data centre overheads: 125.24 kWh (13%)
Carbon intensity used for the calculations: 231.12 gCO2e/kWh

Summary of your usage:
     - First/last job recorded on that period: 2021-01-01/2021-12-08
     - Number of jobs: 1,490 (1,102 completed)
     - Core hours used/charged: 1,302.1 (CPU), 2,852.0 (GPU), 4,154.1 (total).
     - Total usage time (i.e. when cores were performing computations):
        - CPU: 430 days 03:58:39
        - GPU: 118 days 20:01:30
     - Total wallclock time: 132 days 10:49:44
     - Total memory requested: 40,981 GB

Limitations to keep in mind:
     - The workload manager doesn't alway log the exact CPU usage time, and when this information is missing, we assume that all cores are used at 100%.
     - For now, we assume that GPU jobs only use 1 GPU and the GPU is used at 100% (as the information needed for more accurate measurement is not available)
     (both of these may lead to slightly overestimated carbon footprints, although the order of magnitude is likely to be correct)
     - Conversely, the wasted energy due to memory overallocation may be largely underestimated, as the information needed is not always logged.

Any bugs, questions, suggestions? Email LL582@medschl.cam.ac.uk
--------------------------------------------------------------------------------
Calculated using the Green Algorithms framework: www.green-algorithms.org
Please cite https://onlinelibrary.wiley.com/doi/10.1002/advs.202100707
```

GREEN ALGORITHMS 4 HPC

```
GreenAlgorithms4HPC]$ myCarbonFootprint.sh --STARTDAY 2020-01-01 --ENDDAY 2020-06-01
```

https://github.com/GreenAlgorithms/GreenAlgorithms4HPC



Let us know if you try it!

```
####################################
#                                  #
#  Your carbon footprint on CSD3   #
#    (2021-01-01 / 2021-12-31)     #
#                                  #
####################################

        ----------------
        |   222 kgCO2e   |
        ----------------

...This is equivalent to:
    - 20 tree-years
    - driving 1,268 km
    - 4.44 flights between Paris and London

...26.0% of your jobs failed, which represents a waste of 51 kgCO2e (55.26 tree-months).
...On average, you request at least 1.0 times the memory you need. By only requesting the memory you r

Energy used: 960.17 kWh
    - CPUs: 88.91 kWh (9%)
    - GPUs: 713.81 kWh (74%)
    - Memory: 32.22 kWh (3%)
    - Data centre overheads: 125.24 kWh (13%)
Carbon intensity used for the calculations: 231.12 gCO2e/kWh

Summary of your usage:
    - First/last job recorded on that period: 2021-01-01/2021-12-08
    - Number of jobs: 1,490 (1,102 completed)
    - Core hours used/charged: 1,302.1 (CPU), 2,852.0 (GPU), 4,154.1 (total).
    - Total usage time (i.e. when cores were performing computations):
        - CPU: 430 days 03:58:39
        - GPU: 118 days 20:01:30
    - Total wallclock time: 132 days 10:49:44
    - Total memory requested: 40,981 GB
```

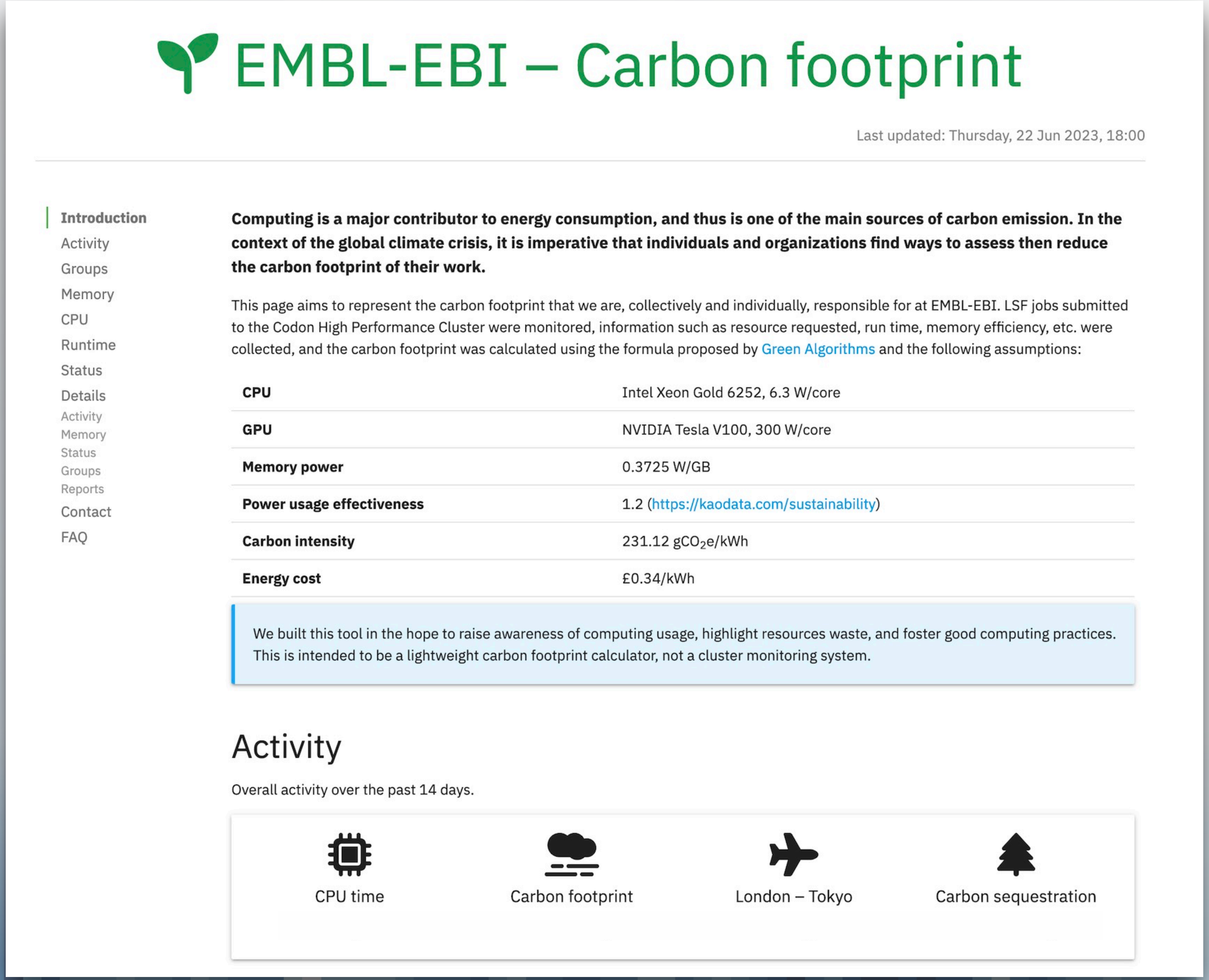
COMING NEXT: GREEN ALGORITHMS **DASHBOARD**

Institutional dashboard to **monitor computing carbon footprint** across research groups, units and departments

COMING NEXT: GREEN ALGORITHMS DASHBOARD

EMBL-EBI

Concept pioneered by EMBL-EBI
(and others!)



EMBL-EBI – Carbon footprint

Last updated: Thursday, 22 Jun 2023, 18:00

- Introduction
- Activity
- Groups
- Memory
- CPU
- Runtime
- Status
- Details
  - Activity
  - Memory
  - Status
  - Groups
  - Reports
- Contact
- FAQ

Computing is a major contributor to energy consumption, and thus is one of the main sources of carbon emission. In the context of the global climate crisis, it is imperative that individuals and organizations find ways to assess then reduce the carbon footprint of their work.

This page aims to represent the carbon footprint that we are, collectively and individually, responsible for at EMBL-EBI. LSF jobs submitted to the Codon High Performance Cluster were monitored, information such as resource requested, run time, memory efficiency, etc. were collected, and the carbon footprint was calculated using the formula proposed by Green Algorithms and the following assumptions:

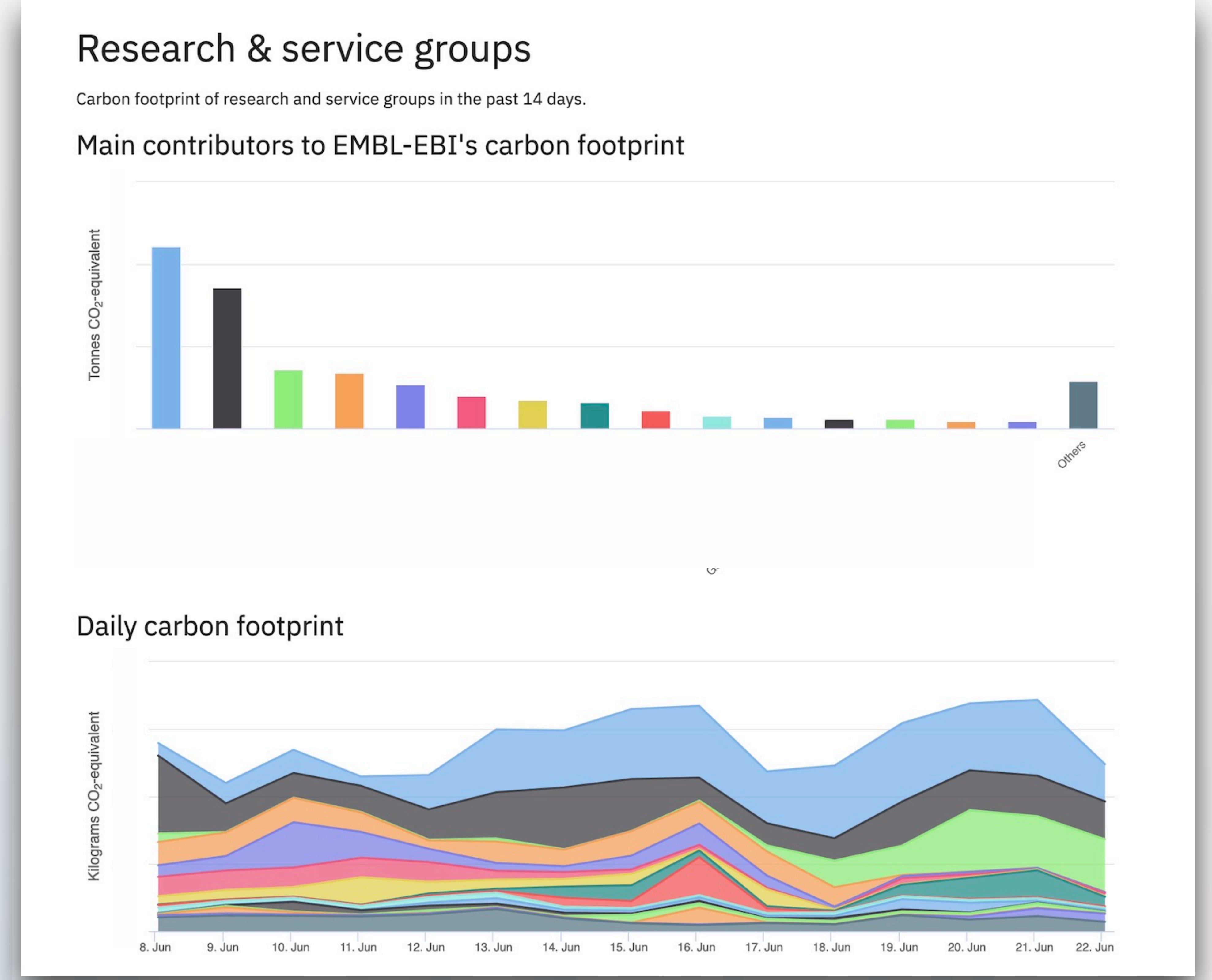
CPU	Intel Xeon Gold 6252, 6.3 W/core
GPU	NVIDIA Tesla V100, 300 W/core
Memory power	0.3725 W/GB
Power usage effectiveness	1.2 (https://kaodata.com/sustainability)
Carbon intensity	231.12 gCO_2e/kWh
Energy cost	£0.34/kWh

We built this tool in the hope to raise awareness of computing usage, highlight resources waste, and foster good computing practices. This is intended to be a lightweight carbon footprint calculator, not a cluster monitoring system.

Activity

Overall activity over the past 14 days.

CPU time | Carbon footprint | London – Tokyo | Carbon sequestration



Matthias Blum

Alex Bateman

COMING NEXT: GREEN ALGORITHMS **DASHBOARD**

Next step: an **open source, easy to deploy, reliable and transparent** SLURM-based dashboard implementing GA4HPC in computing facilities

Interested in piloting it in your organisation? Let's chat!

Already have such a dashboard? Let's chat!

Matthias Blum

Alex Bateman

Michael Inouye

TRANSPARENCY, FROM ALL OF US

Hardware manufacturers

Data centres

Institutions

Cloud providers

Scientists

TRANSPARENCY, FROM ALL OF US

Carbon impact and offsetting

We used GreenAlgorithms v.1.0 (ref. [121]) to estimate that the main computational work in this study had a carbon impact of at least 2,660 kg of CO_2 emissions (CO_2e), corresponding to 233 tree-years. As a commitment to the reduction of carbon emissions associated with computation in research, we consequently funded planting of 30 trees through a local Australian charity, which across their lifetime will sequester a combined estimated 8,040 kg of CO_2e, or three times the amount of CO_2e generated by this study.

Youwen Qin et al., Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort, Nature Genetics, 2022

Carbon impact and offsetting

We used GreenAlgorithms v.1.0 (ref. [84]) to estimate that the main computational work in this study had a carbon impact of at least 1,004 kg of CO_2 emissions (CO_2e), corresponding to 94 tree-years. As a commitment to the reduction of carbon emissions associated with computation in research, we consequently funded the planting of 45 trees through a local Australian charity, which across their lifetime will sequester a combined estimated 12,000 kg of CO_2e, or 12 times the amount of CO_2e generated by this study.

Yu Xu et al., An atlas of genetic scores to predict multi-omic traits, Nature, 2023

Carbon footprint of this project

We did our best to minimise greenhouse gas emissions related to this project and, using the Green Algorithms calculator (v2.1) [35], we estimated that the carbon footprint of this project was 51 kgCO_2e, which corresponds to 4.7 tree-years.

Lannelongue & Inouye, Pitfalls of machine learning models for protein-protein interaction networks, Bioinformatics, 2023

TRANSPARENCY, FROM ALL OF US

tel-03726667, version 1 | Theses

en fr **The vehicle routing problem for flash floods relief operations**

Florent Dubois [1, 2] Details

1 IRIT-SEPIA - Système d'exploitation, systèmes répartis, de l'intergiciel à l'architecture
IRIT - Institut de recherche en informatique de Toulouse
2 UT3 - Université Toulouse III - Paul Sabatier

To conclude on the ethical aspect of my research, the carbon footprint of my thesis has been evaluated. It has been calculated using green-algorithms.org v2.0 [59]. The calculation considers the large-scale experimentation conducted on the platform Grid5000 located in France. Experimentation took a total of

576 hours of computations on 16 CPUs Xeon E5-2660 v3 and has drawn 207.47 kWh. It represents a carbon footprint of $8.08kg\ CO_{2e}$.

Research | Open Access | Published: 19 August 2022

A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions

Lu Yang & Jun Chen

Microbiome **10**, Article number: 130 (2022) | Cite this article

others (146.1s vs 1.2–57.8 s). For large sample sizes, ZicoSeq can complete the analysis at an average of 5 and 25 min for n = 1000 and 5000, respectively (Fig. S22). Based on the Green Algorithms (green-algorithms.org v2.1 [62]) and the geographic location of Minnesota, USA, ZicoSeq has a carbon footprint of 0.06 g CO2e, 0.59 g CO2e, and 3.16 g CO2e for n = 100, 1000, and 5000, respectively.

GENOME RESEARCH

HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CONTA

Efficient computation of Faith's phylogenetic diversity with applications in characterizing microbiomes

George Armstrong[1,2,3], Kalen Cantrell[2], Shi Huang[1,2], Daniel McDonald[1], Niina Haiminen[4], Anna… Imran McGrath[2,8], Kris… Guillaume Méric[12,13], … Mohit Jain[2,17,18], Mic… Laxmi Parida[4], Yoshiki…

0.5 GB of memory. Additionally, using Green Algorithms (Lannelongue et al. 2021), we estimated the carbon footprint of the scikit-bio reference implementation on the 20,000 sample table to be 12.84 g CO_2e, whereas we estimated the carbon footprint of SFPhD would be 0.04 g CO_2e in the United States, which is a 321-fold reduction in impact on global warming.

Equivariant and Modular DeepSets with Applications in Cluster Cosmology

Leander Thiele* Department of Physics, Princeton University, Princeton, NJ 08544

Miles Cranmer Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544

William Coulton, Shirley Ho, David N. Spergel Center for Computational Astrophysics

[6] Total compute cost is 13.4 (Tesla P100+9CPU) khr (1.09t CO_2e [26]) with a PyTorch [27] implementation.

TRANSPARENCY, FROM ALL OF US

NeurIPS 2021 Paper Checklist Guidelines

The NeurIPS Paper Checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. For each question in the checklist:

- (d) Did you include the amount of **compute** and the type of **resources** used (e.g., type of GPUs, internal cluster, or cloud provider)?
  - Ideally, you would provide the compute required for each of the individual experimental runs as well as the total compute.
  - Note that your full research project might have required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper). The total compute used may be harder to characterize, but if you can do that, that would be even better.
  - You are also encouraged to use a CO2 emissions tracker and provide that information. See, for example, the experiment impact tracker (Henderson et al.), the ML CO2 impact calculator (Lacoste et al.), and CodeCarbon.

TRANSPARENCY, FROM ALL OF US

NeurIPS 2021 Paper Checklist Guidelines

The NeurIPS Paper Checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. For each question in the checklist:

But… not anymore in 2023…

9. **Compute:** Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Ideally, you would provide the compute required for each of the individual experimental runs as well as the total compute. Note that your full research project might have required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper). Enter yes, no, n/a, or an explanation if appropriate. Answers are visible to reviewers.
   - Authors are encouraged to provide as much information as practical about the type and amount of compute used for experiments. The total compute used for all experiments may be harder to characterize, but if you can do that, that would be even better.

GPUs, internal cluster, or cloud

the individual experimental runs as well as the total

have required more compute than the experiments reported in the paper

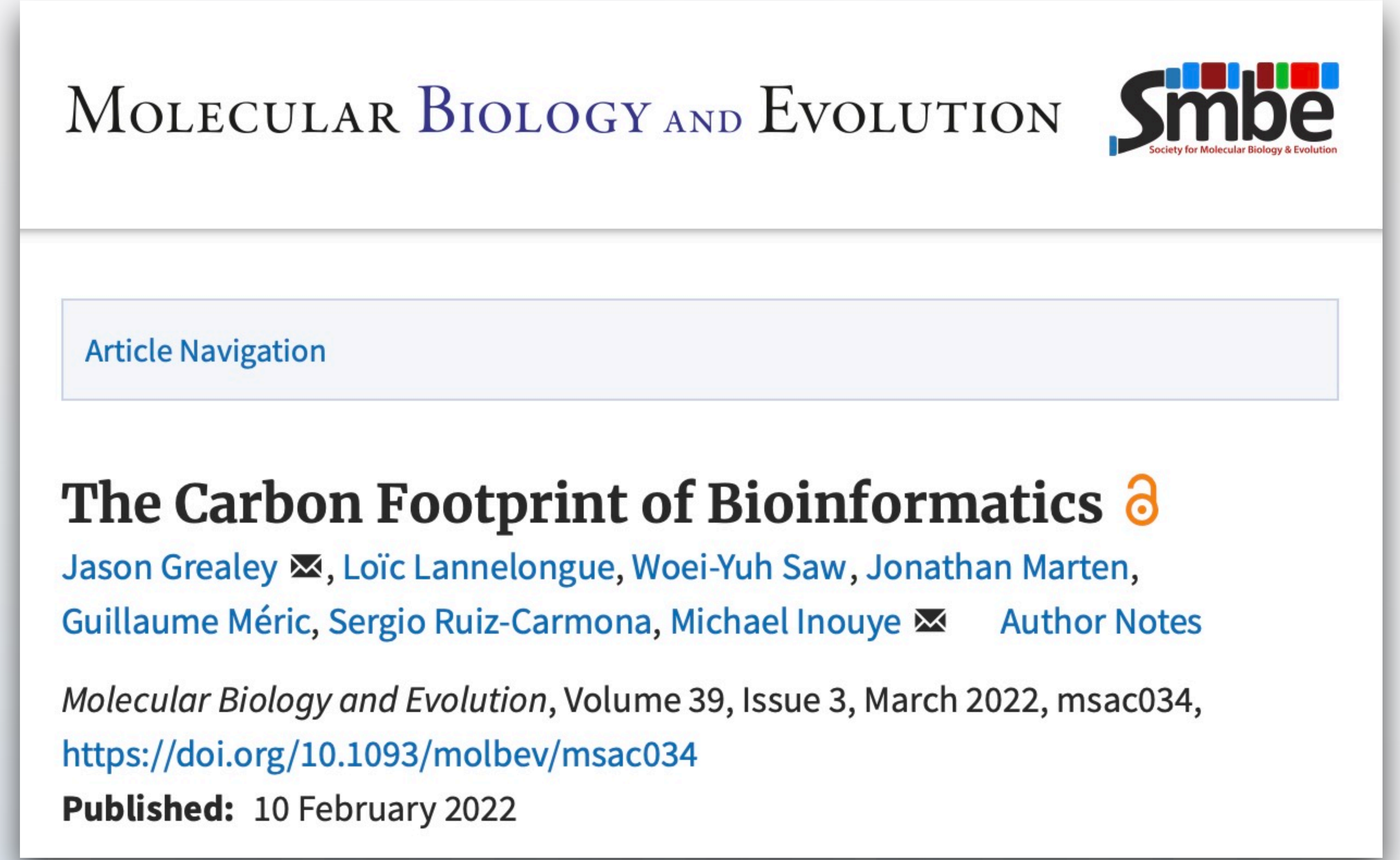
experiments that didn't make it into the paper). The total compute used may be harder to

you can do that, that would be even better.

also encouraged to use a CO2 emissions tracker and provide that information. See, for example, the experiment impact tracker (Henderson et al.), the ML CO2 impact calculator (Lacoste et al.), and CodeCarbon.

DIVING INTO SOME FIELDS

INVESTIGATING BIOINFORMATIC TOOLS


Molecular Biology and Evolution

Smbe
Society for Molecular Biology & Evolution

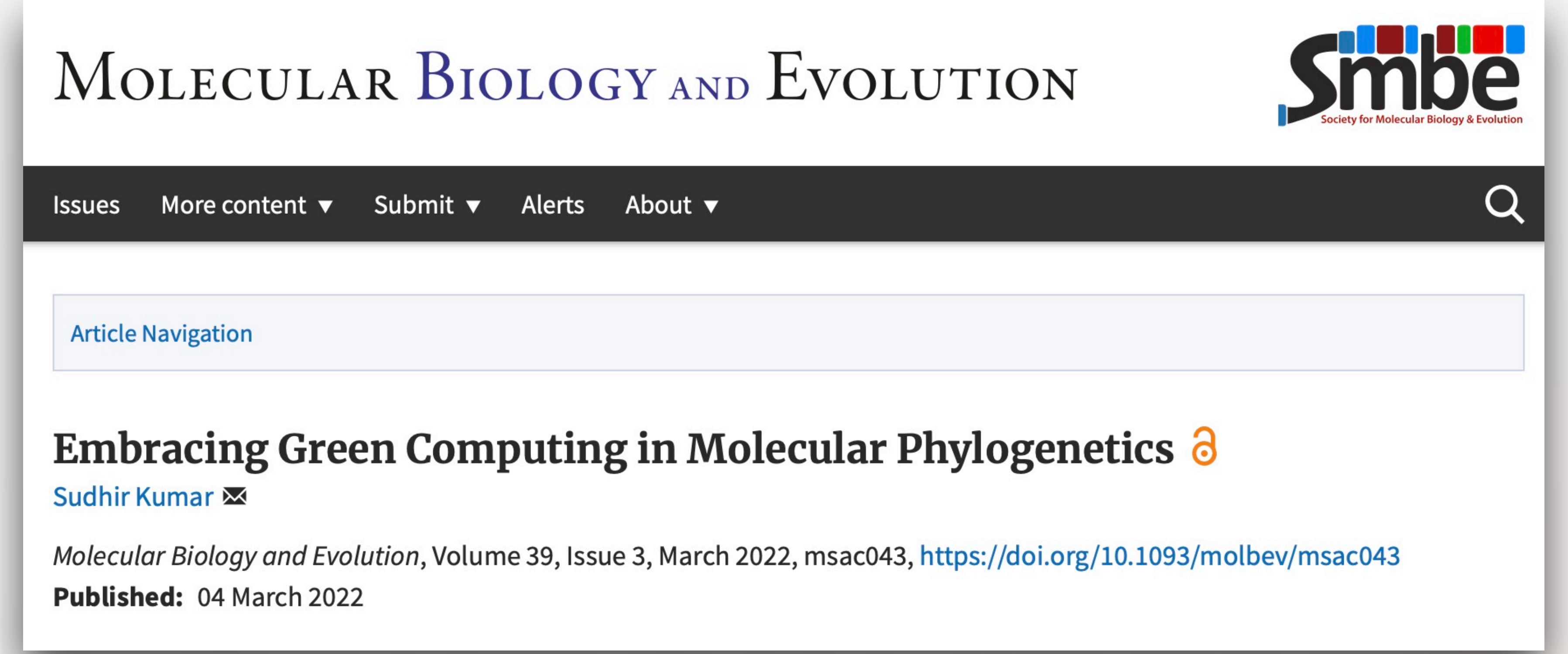
Article Navigation

The Carbon Footprint of Bioinformatics

Jason Grealey, Loïc Lannelongue, Woei-Yuh Saw, Jonathan Marten, Guillaume Méric, Sergio Ruiz-Carmona, Michael Inouye Author Notes

Molecular Biology and Evolution, Volume 39, Issue 3, March 2022, msac034, https://doi.org/10.1093/molbev/msac034

Published: 10 February 2022


Molecular Biology and Evolution

Smbe
Society for Molecular Biology & Evolution

Issues | More content ▾ | Submit ▾ | Alerts | About ▾

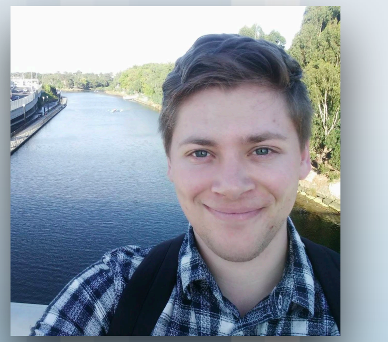
Article Navigation

Embracing Green Computing in Molecular Phylogenetics

Sudhir Kumar

Molecular Biology and Evolution, Volume 39, Issue 3, March 2022, msac043, https://doi.org/10.1093/molbev/msac043

Published: 04 March 2022


Jason Grealey


Woei Yuh Saw

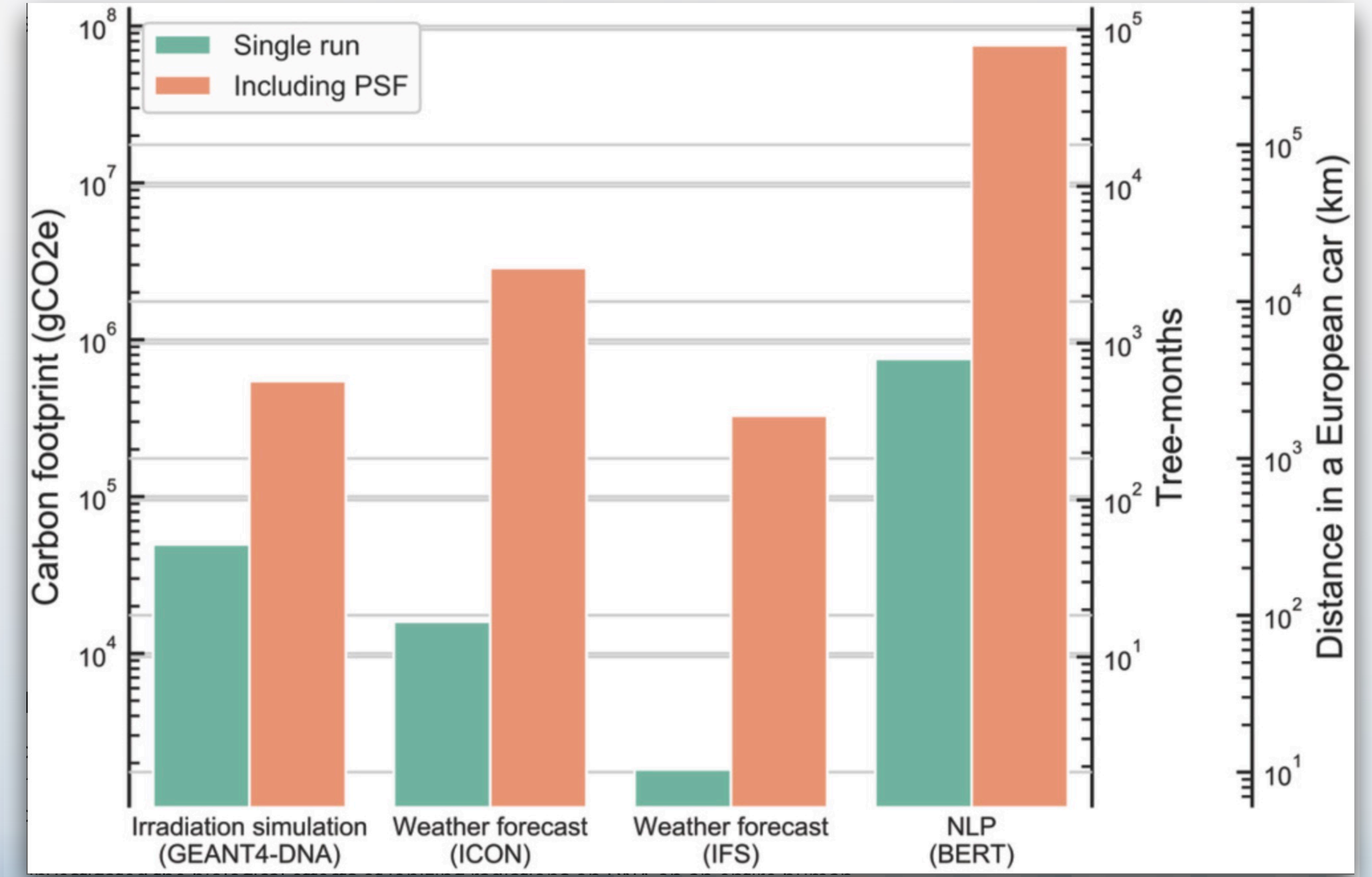

Jonathan Marten


Sergio Ruiz-Carmona


Guillaume Méric


Michael Inouye

EXAMPLES OF CARBON FOOTPRINTS FROM DIFFERENT FIELDS



Lannelongue, Grealey, Inouye, Advanced Science, 2021

FROM ACKNOWLEDGING TO REDUCING IMPACTS

Tackling Energy and embodied impact through New Collaborations

WHAT CAN WE DO NOW?


PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS

EDITORIAL

Ten simple rules to make your computing more environmentally sustainable

Loïc Lannelongue, Jason Grealey, Alex Bateman, Michael Inouye

Published: September 20, 2021 • https://doi.org/10.1371/journal.pcbi.1009324

Keep, Repair, Reuse

Promote efficient data centres

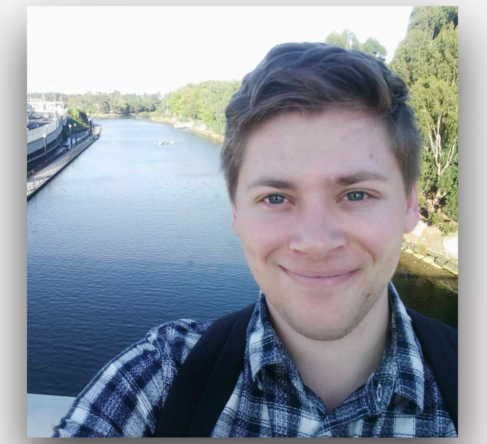
Estimate and report your own footprint for your projects

…and include it in your cost-benefit analysis

Sustainability should be accounted for in renewing policies

Carefully choose your computing facility

Optimise (or use optimised) code



Jason Grealey

Alex Bateman

Michael Inouye

COSTS AND BENEFITS

Estimate and report your own footprint for your projects

…and include it in your cost-benefit analysis

"Digital technologies developed and deployed in pursuit of net zero **must be energy-proportionate** – ie they must bring environmental or societal benefits that outweigh their own emissions."

COSTS AND BENEFITS

Estimate and report your own footprint for your projects

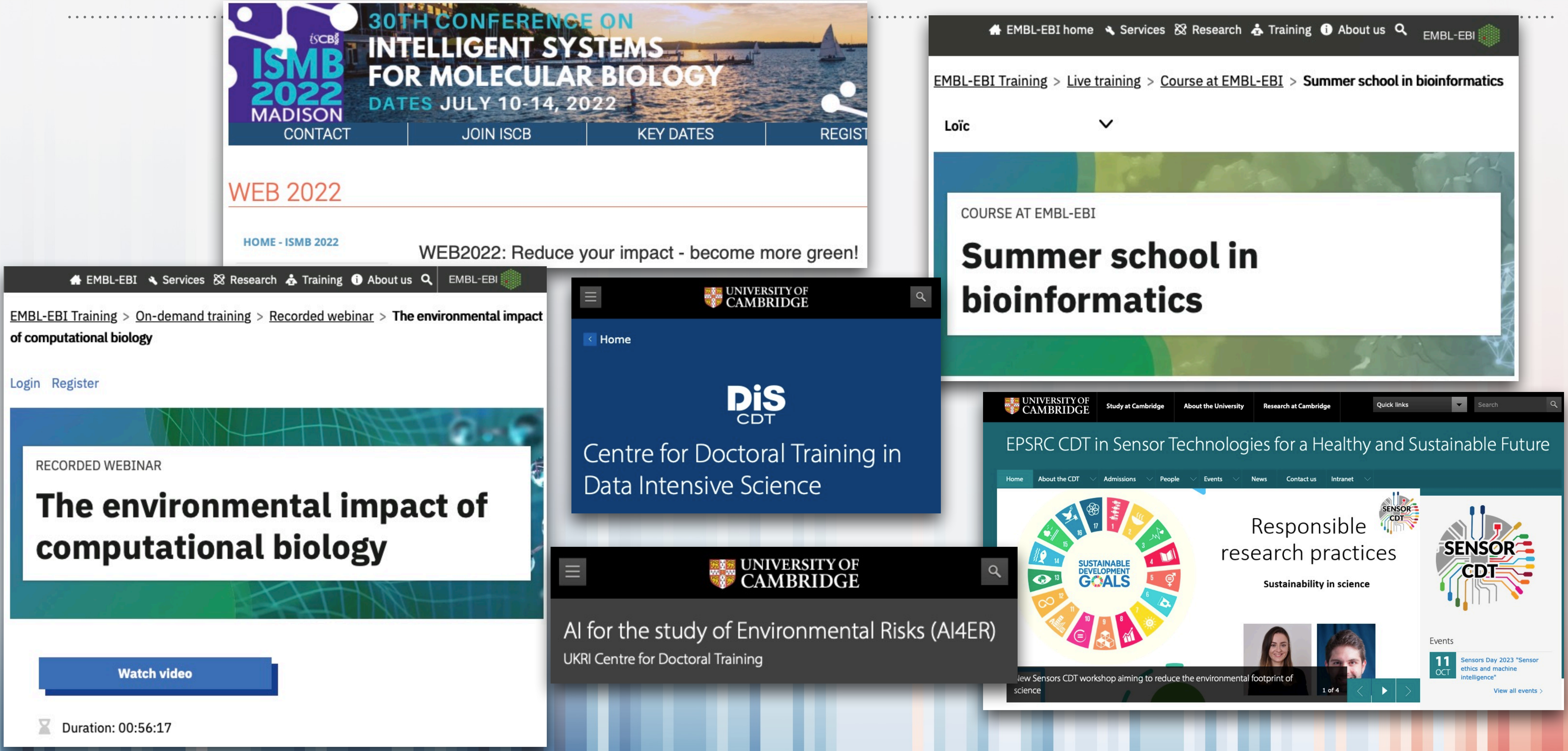
…and include it in your cost-benefit analysis

"Digital technologies developed and deployed in pursuit of net zero **must be energy-proportionate** – ie they must bring environmental or societal benefits that outweigh their own emissions."

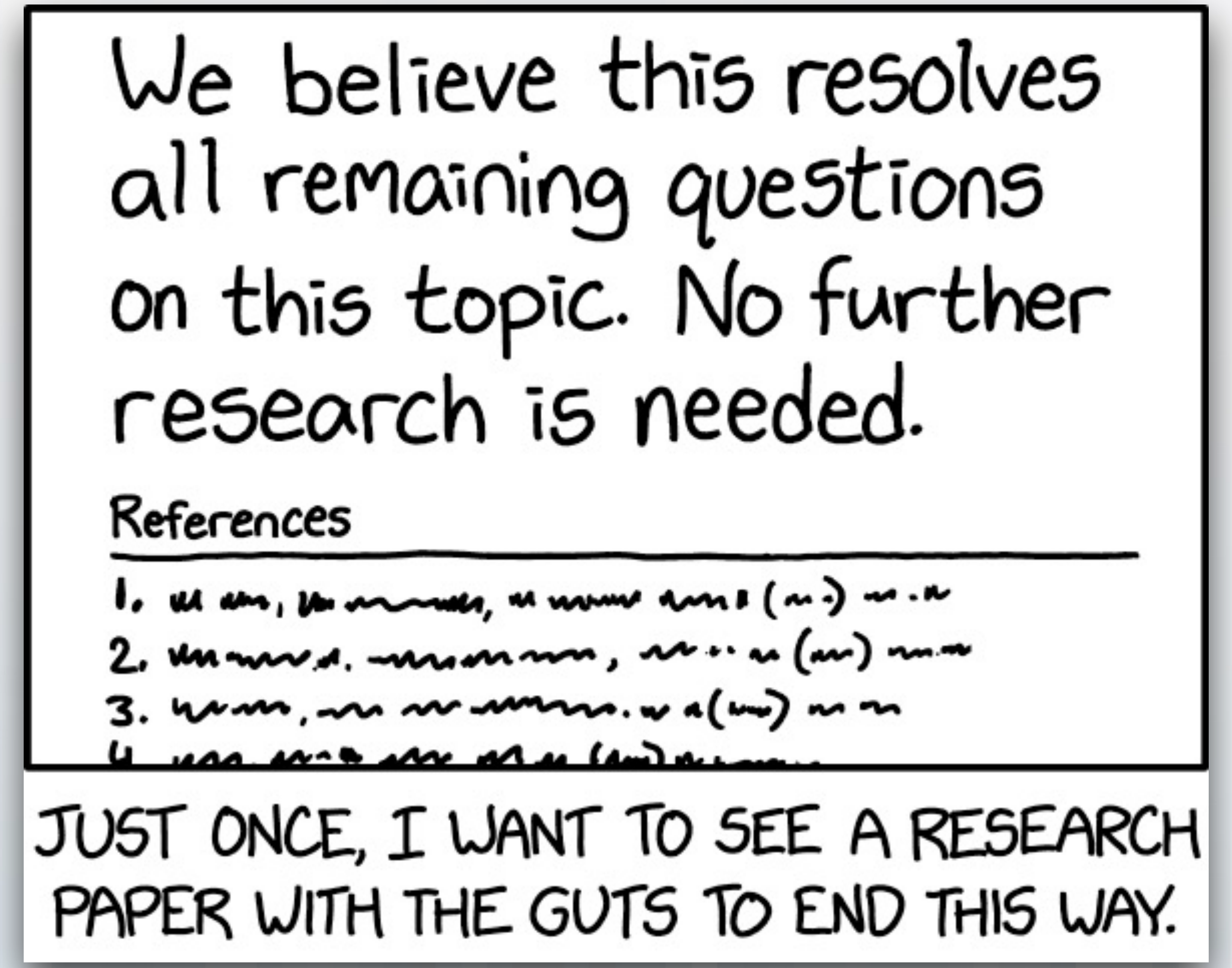
The solutions we develop **benefit the least populations who suffer the most from climate change**, it's true for health research, AI and probably most fields of science.

MOVING FORWARD: EDUCATION AND RESEARCH

MAKING SUSTAINABILITY PART OF SCIENTIFIC TRAINING



IDENTIFY FURTHER OPPORTUNITIES FOR MORE SUSTAINABLE COMPUTING



Sadly not yet

*So **dedicated** research efforts are needed*

xkcd.com

IDENTIFY FURTHER OPPORTUNITIES FOR MORE SUSTAINABLE COMPUTING

Table 3
Results for binary-trees, fannkuch-redux, and fasta.

binary-trees	Energy (J)	Time (ms)	Ratio (J/ms)	Mb
(c) C	39.80	1125	0.035	131
(c) C++	41.23	1129	0.037	132
(c) Rust $\Downarrow_2$	49.07	1263	0.039	180
(c) Fortran $\Uparrow_1$	69.82	2112	0.033	133
(c) Ada $\Downarrow_1$	95.02	2822	0.034	197
(c) Ocaml $\downarrow_1$ $\Uparrow_2$	100.74	3525	0.029	148
(v) Java $\uparrow_1$ $\Downarrow_{16}$	111.84	3306	0.034	1120
(v) Lisp $\downarrow_3$ $\Downarrow_3$	149.55	10570	0.014	373
(v) Racket $\downarrow_4$ $\Downarrow_6$	155.81	11261	0.014	467
(i) Hack $\uparrow_2$ $\Downarrow_9$	156.71	4497	0.035	502
(v) C# $\downarrow_1$ $\Downarrow_1$	189.74	10797	0.018	427
(v) F# $\downarrow_3$ $\Downarrow_1$	207.13	15637	0.013	432
(c) Pascal $\downarrow_3$ $\Uparrow_5$	214.64	16079	0.013	256
(c) Chapel $\uparrow_5$ $\Uparrow_4$	237.29	7265	0.033	335
(v) Erlang $\uparrow_5$ $\Uparrow_1$	266.14	7327	0.036	433
(c) Haskell $\uparrow_2$ $\Downarrow_2$	270.15	11582	0.023	494
(i) Dart $\downarrow_1$ $\Uparrow_1$	290.27	17197	0.017	475
(i) JavaScript $\downarrow_2$ $\Downarrow_4$	312.14	21349	0.015	916
(i) TypeScript $\downarrow_2$ $\Downarrow_2$	315.10	21686	0.015	915
(c) Go $\uparrow_3$ $\Uparrow_{13}$	636.71	16292	0.039	228
(i) Jruby $\uparrow_2$ $\Downarrow_3$	720.53	19276	0.037	1671
(i) Ruby $\Uparrow_5$	855.12	26634	0.032	482
(i) PHP $\Uparrow_3$	1,397.51	42316	0.033	786
(i) Python $\Uparrow_{15}$	1,793.46	45003	0.040	275
(i) Lua $\downarrow_1$	2,452.04	209217	0.012	1961
(i) Perl $\uparrow_1$	3,542.20	96097	0.037	2148
(c) Swift		n.e.		

Contents lists available at ScienceDirect
Science of Computer Programming
ELSEVIER
www.elsevier.com/locate/scico

Ranking programming languages by energy efficiency

Rui Pereira [a,b,*], Marco Couto [c,b], Francisco Ribeiro [c,b], Rui Rua [c,b], Jácome Cunha [c,b], João Paulo Fernandes [d], João Saraiva [c,b]

[a] C4 — Centro de Competências em Cloud Computing (C4-UBI), Universidade da Beira Interior, Rua Marquês d'Ávila e Bolama, 6201-001, Covilhã, Portugal
[b] HASLab/INESC Tec, Portugal
[c] Universidade do Minho, Portugal
[d] Departamento de Engenharia Informática, Faculdade de Engenharia da Universidade do Porto & CISUC, Portugal

(but many softwares we use e.g. in Python actually use efficient languages under the hood)

We need more trained Research Softwares Engineers

ALL THIS LEADING TO CULTURAL CHANGE

JEVON'S PARADOX

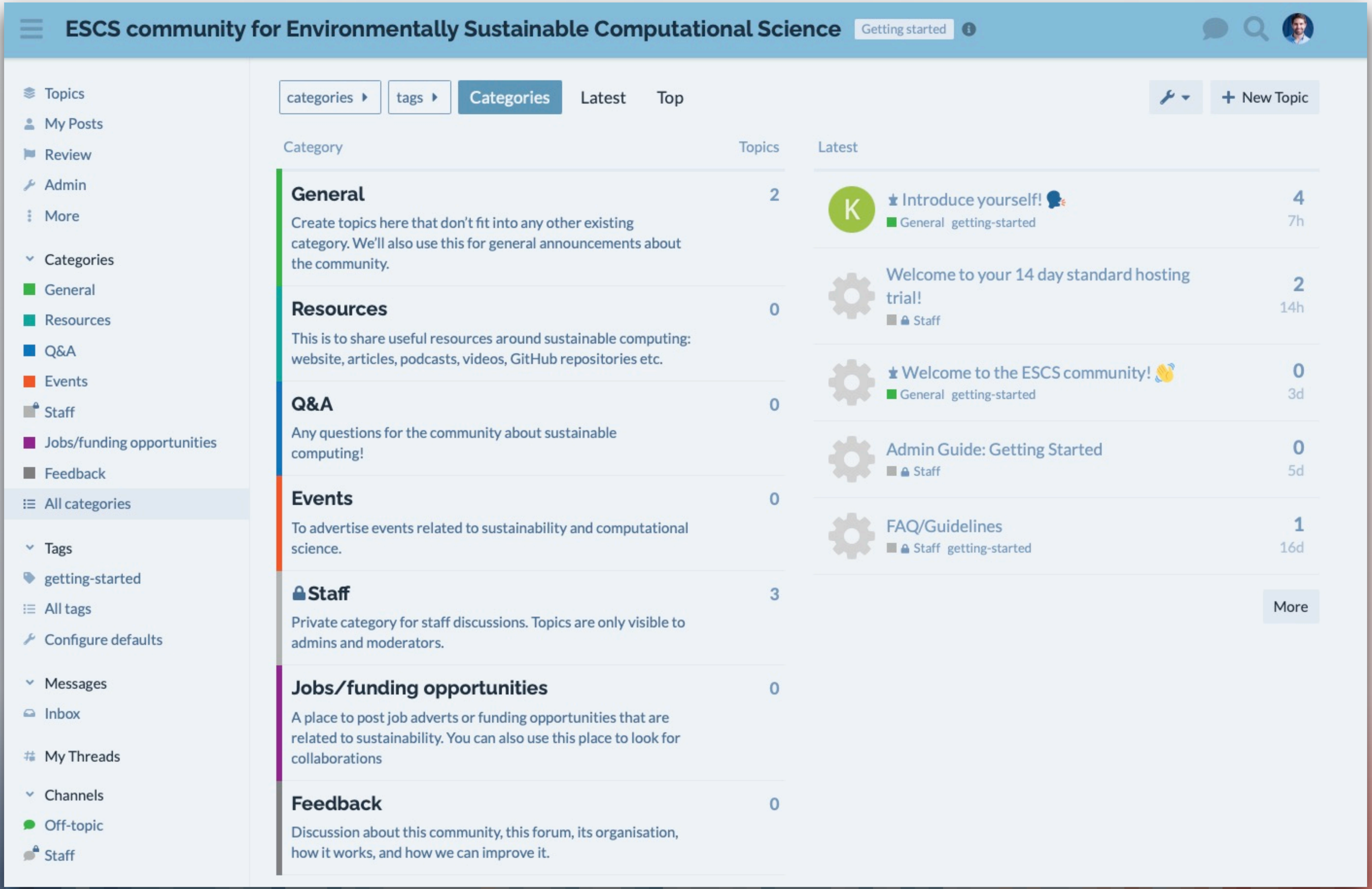
Rebound effect can ruin all our efforts

BUILDING THE ESCS COMMUNITY

An online forum
for discussions

Newsletter/mailing list

Regular online meet-ups



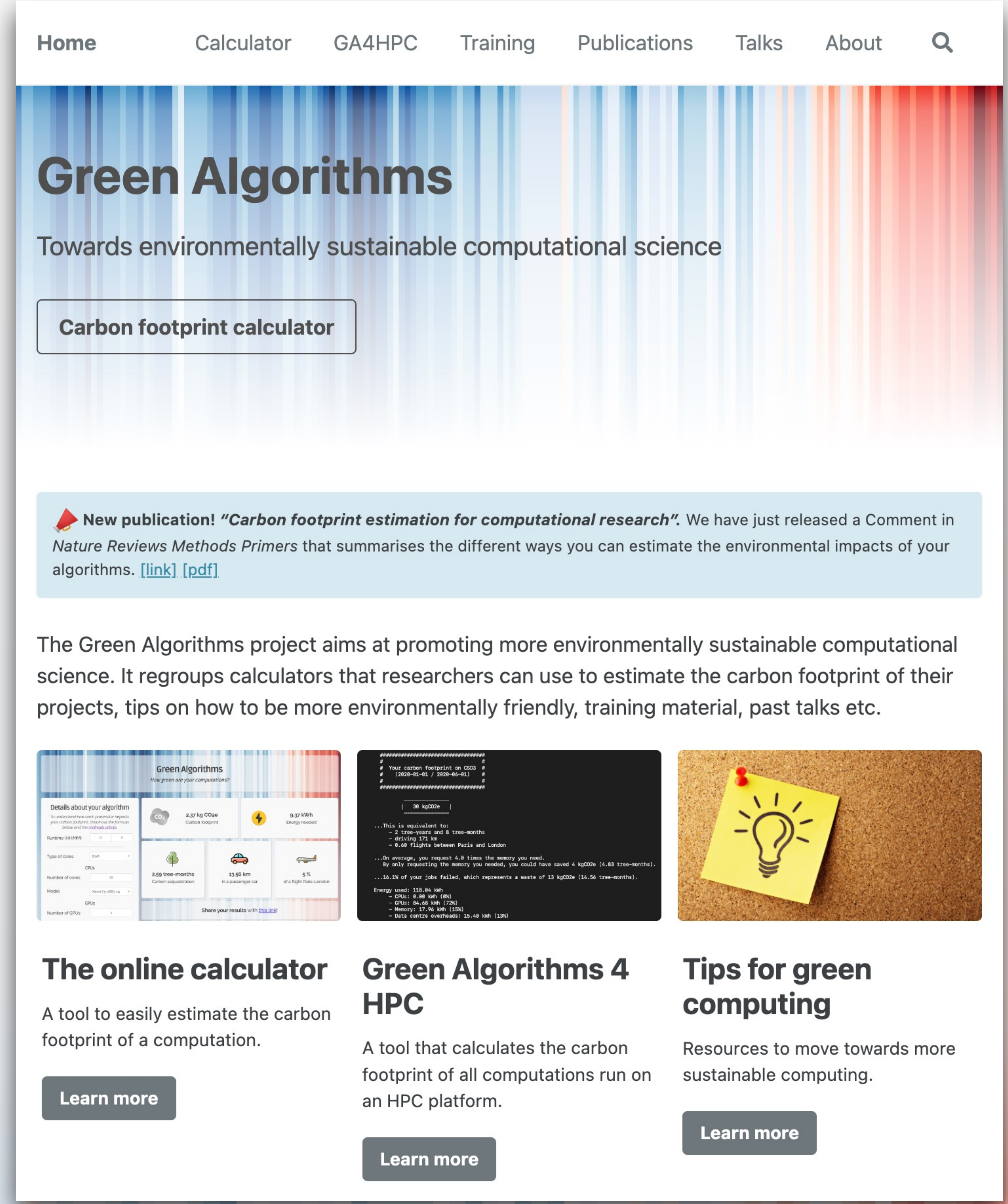
BUILDING THE ESCS COMMUNITY

forum.escs-community.org

To sign up for the newsletter

The Green Algorithms website with all resources

At www.green-algorithms.org



ACKNOWLEDGMENTS

UNIVERSITY OF CAMBRIDGE
Department of Public Health and Primary Care

Cambridge-Baker System Genomics Initiative

Michael Inouye
Jason Grealey
Woei-Yuh Saw
Jonathan Marten
Guillaume Meric
Sergio Ruiz-Carmona
Elodie Persyn
Yang Liu

EMBL-EBI

Alex Bateman
Ewan Birney
Johanna McIntyre
Anna Swan
Sarah Morgan
Matthias Blum

Uni of Sussex

Charlotte Rae
Nick Souther

HDR-UK

Andrew Morris
Hans-Erik Aronson
Gerry Reilly

UKRI

Martin Farley
Martin Juckes

The CATS team

KCL

Gabrielle Samuel

Oxford

Federica Lucivero

LISN/Paris-Saclay

Anne-Laure Ligozat
Clément Morand

LaBRI

Aurélie Bureau
Lucia Souza

British Heart Foundation
NIHR | Cambridge Biomedical Research Centre
Baker HEART & DIABETES INSTITUTE
HDRUK Health Data Research UK
Software Sustainability Institute
UKRI Medical Research Council
Jesus College Cambridge

How to follow the project and reach out

www.green-algorithms.org

@Loic_Lnlg LL582@medschl.cam.ac.uk

The ESCS Community of Practice

Not all compute has large environmental impacts, but enough do for it to be an issue.

Similarly to ethics in medical science, sustainability needs to be an integral part of research practices.

As researchers, we should quantify, monitor, report and reduce (when possible) the environmental impacts of our compute.