

EOS

VOL. 101 | NO. 8
AUGUST 2020

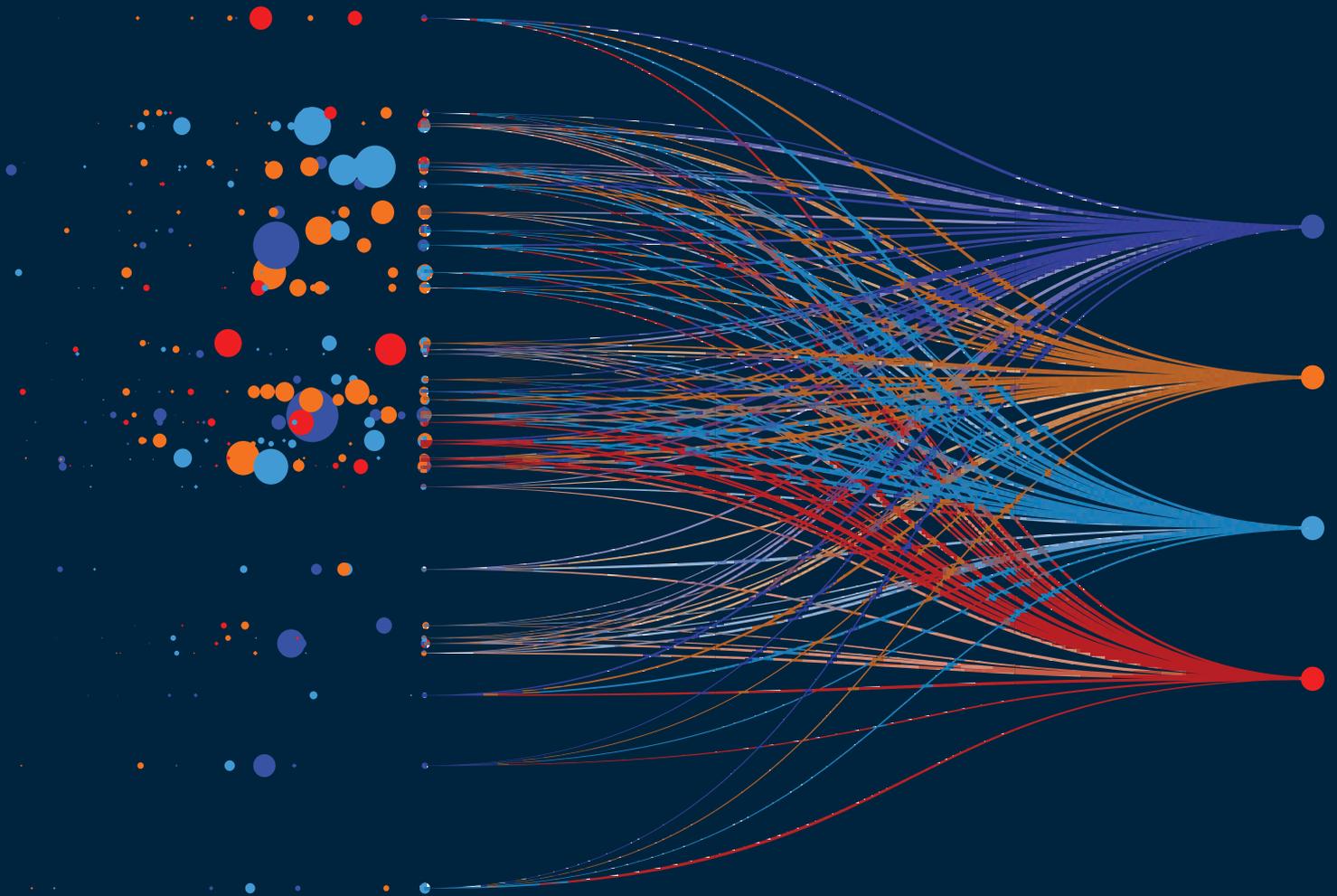
SCIENCE NEWS BY AGU

Europe's Biodiversity Strategy

A Virtual Hackathon
Fights Locusts

MH370's Search
Reveals New Science

INNOVATIONS IN TECHNOLOGY GOT US INTO
THE DATA PROBLEM.
WE NEED AN EVOLUTION IN TECHNOLOGY TO GET US OUT.





A GREENER EARTH MODEL

BY RICHARD LOFT



As weather and climate models grow larger and more data intensive, the amount of energy needed to run them continues to increase. Are researchers doing enough to minimize the carbon footprint of their computing?

SOME PROJECTIONS INDICATE THAT BY 2030, ELECTRICITY USAGE BY COMMUNICATION TECHNOLOGY COULD CONTRIBUTE UP TO 23% OF GLOBAL GREENHOUSE GAS EMISSIONS.

Recently, the National Center for Atmospheric Research (NCAR), where I serve as director of technology development in the Computational and Information Systems Laboratory, conducted a carbon footprint analysis. The organization was quite pleased with the results, until it realized that the analysis neglected to account for carbon dioxide emissions related to the lab's modeling activities. When these emissions were included, the overall picture looked considerably less green.

How could this oversight have happened? In part, it's because modern society is very good at obscuring the environmental impacts of its activities, whether these are impacts from food production, manufacturing, or many other activities. The same is true of the information technology (IT) industry. Carbon dioxide emissions from the IT industry already rival those of the prepandemic airline industry, and some projections indicate that by 2030, electricity usage by communication technology could contribute up to 23% of global greenhouse gas emissions [Andrae and Edler, 2015].

In his book *How Bad Are Bananas? The Carbon Footprint of Everything*, Mike Berners-Lee estimates that the energy required to transmit a typical email generates 4 grams of

carbon dioxide equivalents. That number may give pause to some, while for others it may represent an acceptable cost of doing business in the modern world. Regardless, unlike a gasoline-powered car with an exhaust pipe, there's nothing about the act of sending an email that makes it obvious that we're causing carbon dioxide emissions, so the environmental cost is easy to overlook. With the continuing growth of cloud computing and the Internet of Things, more emissions, like those from computing and the communication of data, will be further virtualized.

Within the Earth system sciences (ESS) community, many initiatives around the world are planning the next generation of global weather and climate models that will be capable of resolving storms and, ultimately, clouds. Examples of these include the Climate Modeling Alliance (CliMA) model, the Energy-efficient Scalable Algorithms for Weather and Climate Prediction at Exascale (ESCAPE-2) model, and the Energy Exascale Earth System Model (E³SM). The repeated calls from the community [Shukla et al., 2010; Palmer, 2014] to

build powerful supercomputing machines to tackle long-standing model biases in the water cycle and improve predictions seem to be coming to fruition. All of these long-running, high-resolution models, and the big computers needed to run them, will require investments from governments and philanthropic organizations. For example, in February, the United Kingdom announced £854 million in funding to develop its next-generation computer, and the U.S. National Oceanic and Atmospheric Administration announced that it was tripling the size of its investment for weather- and climate-related computing.

In this context, the ESS community should consider the question, What is our collective responsibility to reduce carbon emissions related to these large-scale modeling activities?

Leading by Example

In discussions with colleagues, three counterarguments have been offered: first, that weather and climate modeling activities are only a small contributor to societal emissions overall; second, that ESS research is too important to let it be slowed by these considerations; and third, that even raising the subject of emissions from ESS computing provides ammunition to both ends of the political spectrum to attack the research goals.



The central processing unit-based Cheyenne supercomputer, went into service in 2017 at the National Center for Atmospheric Research (NCAR)-Wyoming Supercomputing Center. Credit: University Corporation for Atmospheric Research (UCAR), Carlye Calvin, CC BY-NC 4.0 (bit.ly/ccbync4-0)

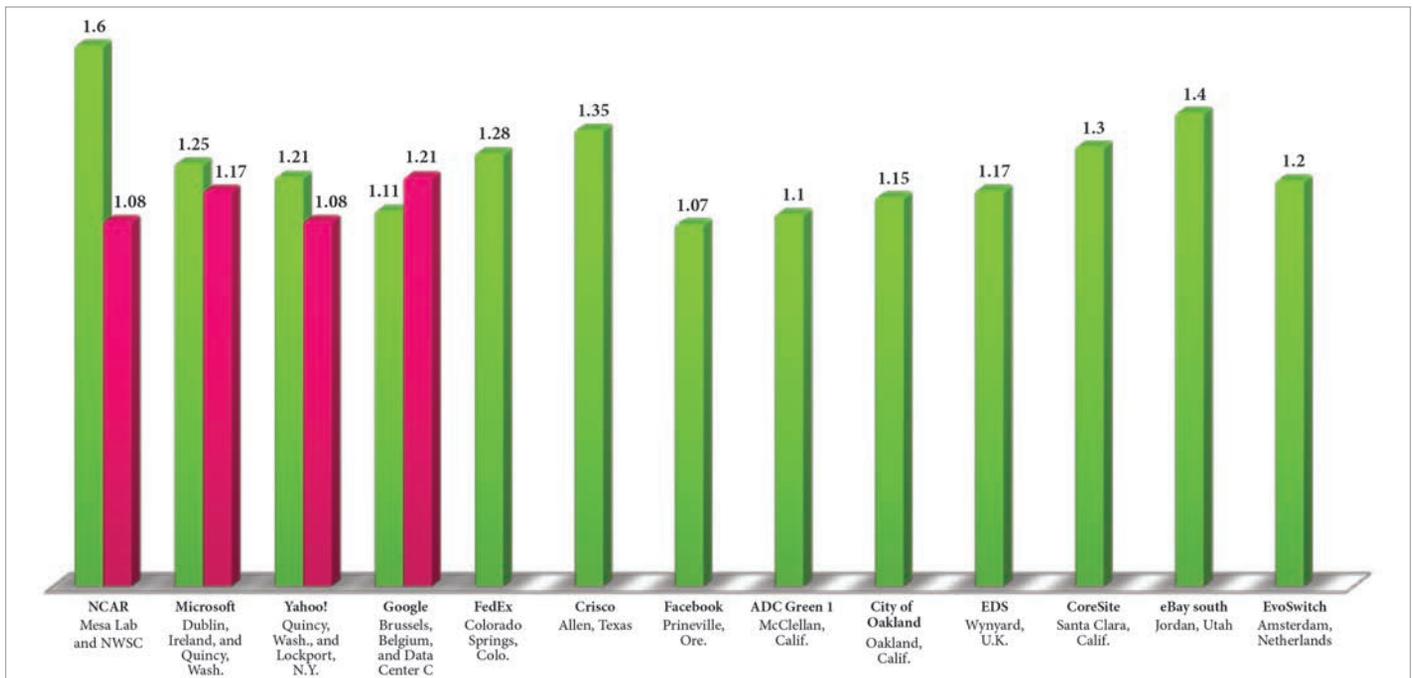


Fig. 1. A comparison of power usage efficiency (PUE) at a variety of energy-efficient data centers. All of these data centers were built within the past few years, with the exception of NCAR's Mesa Laboratory, which was built in 1967. The average PUE in 2019 among more than 600 data centers surveyed by the Uptime Institute was 1.67.

To be sure, studying the impacts of human activity on our planet is both an important and, of late, a risky area of research: Scientists studying climate, for example, are frequently caught in political crosshairs and are also subjected to trolling and other forms of harassment. But the first two arguments above sound like rationalizations that might be made by anyone when they are first confronted with their carbon footprint.

Certainly, it seems this is an opportunity for the ESS community to lead by example. So perhaps a more productive question is, Have we done everything we can to minimize the carbon footprint of our computing activities?

In answering this question, it's useful to separate considerations about decarbonizing utilities from those of reducing the "energy to solution," a metric for comparing high-resolution atmospheric models that accounts for the total energy needed to run a model simulation from start to finish [Fuhrer et al., 2018]. In this way, questions about the merits of carbon offset schemes or the locations of facilities, for example, can be considered apart from questions about how to reduce the energy consumed by our research activities.

Components of Computing's Carbon Footprint

Energy sources: Switching to renewable energy sources like wind, solar, and biogas must be part of the solution to mitigate cli-

mate change, and we should laud and try to emulate organizations that do so. It is worth considering, however, that the environmental side effects of a future decarbonized energy portfolio are not well understood [Luderer et al., 2019], and switching to renewable energy sources often means buying credits, which can be traded, thereby obfuscating the actual source of the energy powering a computing facility.

THIS IS AN OPPORTUNITY FOR THE ESS COMMUNITY TO LEAD BY EXAMPLE. HAVE WE DONE EVERYTHING WE CAN TO MINIMIZE THE CARBON FOOTPRINT OF OUR COMPUTING ACTIVITIES?

In the meantime, one way to address the problem without creating more problems or sacrificing transparency is to improve the efficiency of the modeling enterprise. This idea is perhaps best expressed by the Japanese expression "mottainai," often interpreted to mean "waste not, want not."

We can consider the challenge of improving the energy efficiency of modeling as an infrastructure stack to be optimized, where the components are the computing facilities (the data center), the computing infrastruc-

ture (the computer and data system), the software (the model), and the data motion (the workflow).

Computing facilities: According to a recent report from the Uptime Institute, "Efforts to improve the energy efficiency of the mechanical and electrical infrastructure of the data center are now producing only marginal improvements. The focus needs to move to IT." The commonly used measure of facility efficiency is power usage efficiency (PUE), which is basically the total power a facility uses divided by the power used to run the infrastructure within the facility. Whereas older computing facilities still have room for improvement in PUE, recently built data centers are, for the most part, already quite efficient (Figure 1). For example, the NCAR-Wyoming Supercomputing Center (NWSC), which my organization operates for the ESS research community on behalf of the National Science Foundation, has a PUE of 1.08, meaning that if the facility were perfect, with no energy usage overhead, it would yield only an 8% improvement in PUE compared with its present state. Still, as efficient as it is, since the facility's commissioning in 2012, the computers that NWSC houses have emitted roughly 100,000 metric tons of carbon dioxide—roughly the mass of a modern aircraft carrier. In such cases, the best path to further optimization isn't the facility.

Computing infrastructure: Supercomputers on the scale that will be required for high-resolution Earth system models are being deployed by the U.S. Department of Energy.

The Summit supercomputer at Oak Ridge National Laboratory and the planned Aurora system at Argonne National Laboratory, for example, will each consume on the order of 10 megawatts, annually producing carbon dioxide emissions equivalent to those from more than 30,000 round-trip flights between Washington, D.C., and London (calculated using the Avoided Emissions and Regeneration Tool (AVERT) and the MyClimate initiative). Using values of standard system performance benchmarks—specifically the High Performance Conjugate Gradients (HPCG) and High Performance Linpack (HPL) benchmarks—from published sources, such as the top500.org list of the 500 fastest supercomputers, we find, for example, that the graphics processing unit (GPU)-based Summit supercomputer system is 5.7 and 7.2 times more power efficient on HPCG and HPL, respectively, than NCAR’s roughly contemporaneous central processing unit (CPU)-based Cheyenne system at NWS. GPUs look like an energy-efficient alternative to CPUs for reducing energy to solution.

The software: In its report, the Uptime Institute notes that optimizing energy-efficient software is the least frequently used best practice suggested by the European Code of Conduct (CoC) for Data Centre Energy Efficiency. Could Earth system models, which have been designed to run on CPUs predominantly, be optimized for GPUs? Experience with atmospheric models that have been adapted, or ported, to run on GPUs, like the Consortium for Small-scale Modeling (COSMO) model and most recently the Model for Prediction Across Scales (MPAS), has shown that significant energy savings can be achieved. However, refactoring legacy models for GPUs is labor intensive, creating an obstacle to this potential energy savings. And because of the substantial complexity of ESS code, Earth system modelers have not shown much appetite for adopting these power-efficient devices.

Data motion: Also relevant to the carbon footprint of ESS modeling is data motion. Moving a petabyte of data around a data center requires roughly the amount of energy contained in half a barrel of oil; burning this much oil produces about 215 kilograms of carbon dioxide emissions. These figures explain the economics of Amazon’s Snowmobile, a low-tech but effective “sneakernet” solution to the big-data problem in which a data center on wheels—in the form of a semitruck—plugs in to your data center, transfers data to

onboard storage devices, and drives them to Amazon for storage in the cloud. Yes, it is actually more efficient and faster to load data into a truck and drive them across country than to transfer them via the Internet. Given that a single 1-kilometer 3D atmospheric field will produce roughly half a terabyte in single precision, data movements of this magnitude surely will be required routinely in an exascale computing complex.

The Path to Efficiency

I propose five steps the ESS modeling community should take to move toward a more energy efficient and lower-emissions future.

First, although many modern computing centers are already very efficient, there may still be ways to improve PUE of older facilities that should be examined. Also, when possible, use renewable energy sources to power facilities.

Second, regarding modeling software, you can’t improve what you don’t measure. Publishing the energy consumed per simulated day when reporting benchmarking results, particularly for high-resolution

EXPERIENCE WITH ATMOSPHERIC MODELS THAT HAVE BEEN ADAPTED, OR PORTED, TO RUN ON GPUS HAS SHOWN THAT SIGNIFICANT ENERGY SAVINGS CAN BE ACHIEVED.

models requiring large-scale computing resources, would help raise awareness about energy efficiency in ESS modeling, and through competition could foster developments leading to further energy savings. This approach should be integrated into existing model intercomparison projects, such as the Dynamics of the Atmospheric General Circulation Modeled on Non-hydrostatic Domains (DYAMOND) initiative from the Centre of Excellence in Simulation of Weather and Climate in Europe.

Third, porting existing models to, and developing models for, new computing architectures must be made easier. Researchers seeking to program the current generation of energy-efficient GPUs can attest to the difficulty of achieving this feat. This difficulty comes from multiple sources:

the inherent complexity of Earth system model codes, the lack of workforce trained in programming new technology, the underresourcing of such programming efforts, and the inherent architectural complexity of the heterogeneous computing systems currently available.

Fourth, chip manufacturers and supercomputer vendors should make it easier for users to measure the actual amount of power drawn during execution of models. For some models, the actual power draw can be as much as 40% less than the “nameplate” power of the systems on which they’re run. Currently, the tools to make these measurements are often poorly documented and are architecture dependent. They are also typically used only by computer science teams.

Fifth, the ESS community should increase research into techniques that could lead to energy savings, including reducing floating point precision computation; developing machine learning algorithms to avoid unnecessary computations; and creating a new generation of scalable numerical algorithms that would enable higher throughput in terms of simulated years per wall clock day.

Achieving higher scientific throughput with reduced energy consumption and reduced emissions is a mottainai approach that is not only defensible but is also one of which everyone in the ESS community could be proud.

References

- Andrae, A. S. G., and T. Edler (2015), On global electricity usage of communication technology: Trends to 2030, *Challenges*, 6(1), 117–157, <https://doi.org/10.3390/challe6010117>.
- Fuhrer, O., et al. (2018), Near-global climate simulation at 1 km resolution: Establishing a performance baseline on 4888 GPUs with COSMO 5.0, *Geosci. Model Dev.*, 11, 1665–1681, <https://doi.org/10.5194/gmd-11-1665-2018>.
- Luderer, G., et al. (2019), Environmental co-benefits and adverse side-effects of alternative power sector decarbonization strategies, *Nat. Commun.*, 10, 5229, <https://doi.org/10.1038/s41467-019-13067-8>.
- Palmer, T. (2014), Climate forecasting: Build high-resolution global climate models, *Nature*, 515, 338–339, <https://doi.org/10.1038/515338a>.
- Shukla, J., et al. (2010), Toward a new generation of world climate research and computing facilities, *Bull. Am. Meteorol. Soc.*, 91(10), 1,407–1,412, <https://doi.org/10.1175/2010BAMS2900.1>.

Author Information

Richard Loft (loft@ucar.edu), Computational and Information Systems Laboratory, National Center for Atmospheric Research, Boulder, Colo.

The article represents the personal views and opinions of the author only and does not reflect the opinions of NCAR/UCAR.

► Read the article at bit.ly/Eos-greener-Earth-model