

大规模计算产生巨量碳足迹

(中国科学院高能物理研究所 刘珺怡 陈刚 编译自 Michael Allen. *Physics World*, 2022, (3): 46)

荷兰莱顿大学的天体物理学家 Simon Portegies Zwart 富有生态意识。他几乎不再因为工作缘故坐飞机，而是选择乘火车出行。“我喜欢成为素食环保主义者，尽量减少自己的碳足迹，同时也告诫孩子们避免洗澡时间过长，并尽可能使用可再生资源。”在他决定做出这些生活上的改变时，也在思考着其他方面带来的碳足迹。

“我经常使用大型计算机，它们消耗的能源相当于一座小城市，”他说道，“我可能是这条街上污染最严重的人了。假如使用一台超级计算机耗费的能源相当于10000户家庭，那么我有何权利去跟我的孩子或者别人说，他们不该洗20分钟的澡？”

在全球为解决气候变化问题而努力的同时，许多科学家开始正视自己碳排放量过大的现实。

巨大的计算成本

除学术旅行影响气候变化外，过去几年中许多物理学家还发现，使用计算机造成的碳足迹数量巨大——有时甚至超过航空旅行。

Adam Stevens 是西澳大利亚大学的一名天体物理学家，他和同事们对2018—2019年期间澳大利亚天文学家因旅行、使用超级计算机以及在大型观测站工作等“常规活动”产生的温室气体排放总量进行了分析。他们的研究发现，平均每位澳大利亚天文学家产生约37吨二氧化碳当量，超出澳洲人的平均水平40%，是全球平均水平的5倍。

其主要原因在于天文学家需要使用超级计算机来处理望远镜收集的大量数据，并进行宇宙学模拟。每位天文学家在此项工作上的排放量约为15吨，几乎是年飞行排放量的4倍(图1)。

另一个例子是即将开展的大型中微子探测阵列(GRAND)项目，该项目计划利用分布在世界各地山区的20万根天线，探测来自深空的超高能量中微子。2021年，该项目背后的团队估算了三个不同实验阶段的温室气体排放，分别是：原型实验、中规模实验以及将于2030年进行的全面实验。他们把模拟和数据分析、数据传输和存储以及计算机和其他电子设备称为“数字技术”，这些技术将在碳足迹中占据很大比重。

预计原型实验阶段数字技术产生的排放将占69%，相比之下旅行仅占27%，4%来自“硬件设备”，如制造无线电天线。在中期实验阶段，数字技术将占总排放量的40%，剩余排放中旅行和硬件各占一半。当整个实验完成并投入使用后，主要排放量将由硬件(48%)和数字技术(45%)分担。

超级计算机的环境成本很大程度上取决于为设备供电的能源来源。2020年，

荷兰天文学委员会邀请 Portegies Zwart 和另一支研究团队分析其6个成员机构的碳足迹。据估计，2019年平均每位荷兰天文学家排放4.7吨二氧化碳当量，远低于澳大利亚，而其中仅有4%来自超级计算。

荷兰天体物理学家 Floris van der Tak 主持了该项研究，他认为荷兰的天文学家不会比澳大利亚的同行们更少使用超级计算机，因此差异可能源自不同的能源供应。由于荷兰100%使用风能或太阳能产生的可再生能源，国家级超级计算中心 SURF 不产生任何碳排放，少量的排放由国际设备和荷兰的小型超级计算机产生。如今，Portegies Zwart 已经养成查看自己所用的超级计算机是否采用环保能源的习惯，如果不是，他将考虑使用其他设备。

问题根源

德国马克斯·普朗克天文研究所的温室气体排放数据同样显示出国家间的碳排放差异。2018年，该研

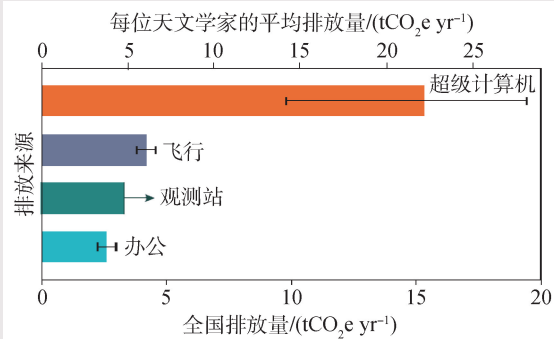


图1 澳大利亚天文学家的四种排放来源，以吨(t)二氧化碳(CO₂)当量(e)每年(yr⁻¹)每人为单位。图中标明了误差线，注意观测站的数值为排放量下限

究所每位研究人员排放了约18吨二氧化碳当量——超过荷兰天文学家，但只有澳大利亚同行的一半(图2)。这一数值比普通德国居民高出60%，是德国2030年减排目标的三倍，而该目标是符合巴黎气候协定的。

在马克斯·普朗克研究所2018年的碳排放中，约有29%来自电力

消耗，其中计算(尤其是超级计算)占75—90%。德国和澳大利亚碳排放差异的关键在于电力的来源。2018年，德国约有一半电力来自太阳能和风能，而在澳大利亚，绝大部分电力来自化石燃料，主要是煤炭。这就意味着在澳大利亚，用于计算的电力每千瓦时产生0.905 kg

二氧化碳，而在马克斯·普朗克研究所只有0.23 kg。

van der Tak同时指出，这些调查工作是在几年前进行的，如今世界已经向前发展，比如现在使用可再生能源的机构越来越多。荷兰的一项研究发现，2019年荷兰天文界的碳足迹中只有不到三分之一(29%)来自电力使用，包括为六家研究机构的本地计算供电。当时就有一半研究所使用绿色电力，随后又有两家开始改用100%可再生能源，van der Tak预计第六家研究所将在未来两年内实现转变。

澳大利亚的状况也在改变。作为该国三大国家级高性能计算设施之一的超级计算机OzSTAR，自2020年7月起已改用从附近的风力发电厂购买的100%可再生能源。超级计算机所在的斯威本科技大学声称，这将大幅度减少其碳足迹，因为电力排放占总排放的70%以上。

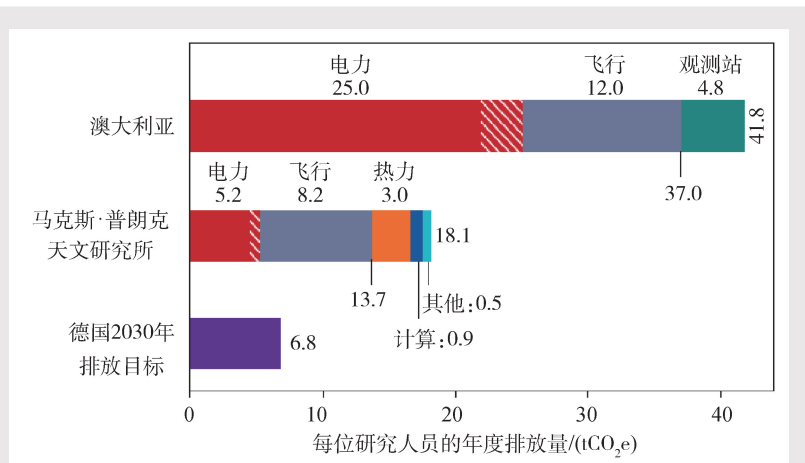


图2 一位澳大利亚天文学家和一位马克斯·普朗克天文研究所的德国研究员2018年的平均排放量，按排放来源分类，并与德国根据《巴黎协定》设置的2030年目标排放比较。与电力相关的排放包括计算和非计算消耗，无论在德国还是澳大利亚，其绝大部分排放都是由计算产生的

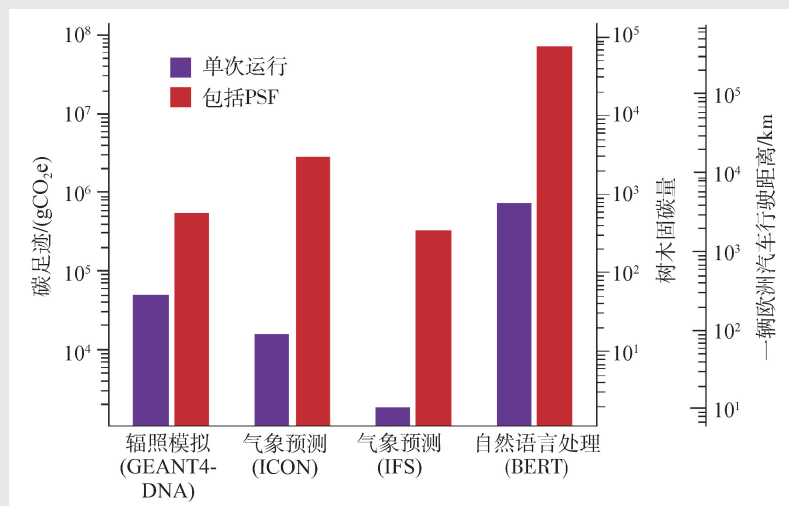


图3 绿色算法是一款免费工具，用于估计算法的碳足迹，估算过程涉及一系列因素，包括硬件要求、运行时间和数据中心位置等。用户可对计算性能进行评估，或者估算在其他架构上重新部署算法节约或消耗的碳。该图对比了不同科学领域算法的碳足迹——从粒子物理模拟和DNA辐射损伤到大气科学再到机器学习——并比较了每个算法只运行一次和同一任务的重复计算(PSF)的结果。上述结果以克(g)二氧化碳(CO₂)当量(e)为单位，并与树木固碳量以及日常活动(如驾驶汽车)的碳排放量进行了比较

地点，地点，还是地点

但是，如何能确切计算出使用超级计算机的碳排放量呢？英国剑桥大学的数学家和物理学家Loïc Lannelongue没有找到简单的方法，于是他开发了一个名为“绿色算法”(green-algorithms.org)的在线工具，估算研究人员的碳足迹。

Lannelongue重申地点是关键。举例来说，在同样的硬件上运行相同任务，澳大利亚排放的二氧化碳大约是瑞士的70倍，因为瑞士的大部分电力来自水电。虽然估算任何一种算法的碳足迹都需要依据硬件、任务所需时间和数据中心或超级计算机的位置等关键因素，但绿色算法还有一个“实际比例因子”(PSF)，用于估计实际计算的次数——这对排放量有着直接影响。

事实上,大多数算法都要运行多次,有时甚至要在不同参数下运行数百次,而且根据任务和研究领域的不同,运行次数会有很大差异(图3)。研究还发现,南非和美国某些州有着与澳大利亚类似的计算碳排放量,而冰岛、挪威和瑞典的电力碳排放则特别低。

如今,随着云计算的出现,研究人员可以更方便地选择所使用的超级计算机。但即使无法更换机器,他们仍有其他方法可以减少碳排放。Lannelongue说,如果无法改变自己的所在地,则可以使用最新版本和优化的软件,因为这将会降低计算要求。

更好的编码

高效的代码对于使计算更环保同样至关重要。正如Portegies Zwart所说,如果你花费更多时间在代码优化上,它会运行得更快,产生的排放也会更少。此外,更换编码语言不失为一种好办法。

为验证这一观点,Portegies Zwart进行了实验,他用十几种不同的编码语言运行同样的算法。没有哪种语言的代码经过特别优化,而且编写每种代码花费的时间相近。与其他编码语言(如C++或Fortran)相比,物理学家常用的Python运行算法时间要长得多,因此会产生更多碳排放。问题在于Python易于使用,却难以优化,而其他语言虽难编码,但更容易优化。

然而,远离Python未必能解决问题。法国格勒诺布尔-阿尔卑斯大学的研究员Pierre Augier称,更好的教育和Python编译器的使用同样有效。他采用更加优化的代码和5种不同的Python实现方式进行了类

似的实验,其中四种实现方式比C++和Fortran更快,产生的排放更少,而且更容易被理解和使用。

Portegies Zwart同意Python可以是高效的,但并不能反映实际情况。他认为天文学家对代码的优化程度并不高,与其让他们学更多计算机知识,物理研究机构或许应该雇佣更多计算机专家。“我们擅长物理,但计算机科学家把我们学习物理的时间都花在了学习计算机上,”他说,“毫无疑问,‘他们’更擅长编程。”

隐性排放

碳密集型工作不止有超级计算机上的仿真。作为GRAND中微子项目的联合发起人,法国索邦大学的Kumiko Kotera和她的同事们发现,在实验的原型阶段,数据存储和传输将占年总排放量的大约一半,中期阶段占四分之一,全面实验阶段占三分之一以上。相比之下,数据分析和仿真在三个阶段产生的碳排放占比分别为16%、13%和7%。

数据存储和传输的碳足迹取决于数据中心的能源需求,使用排放量较低的数据中心可以在一定程度上解决问题。不过,缩减数据量仍有作用,科学家们会对所传输的内容更加谨慎。Kotera表示,GRAND项目将研究如何减少数据量,找到有效清理数据的方法。

粒子物理学家也需要贡献力量。欧洲核子研究中心(CERN)每年产生约100 PB数据。全球LHC计算网格(WLCG)整合全球40多个国家约170个计算中心的计算资源,并对这些数据进行存储、分发和分析。CERN近年来开始发布环境报



图4 为降低整体碳排放,CERN聘请了一名环境工程师,负责监督未来项目的建造工程

告,2021年发布的第二份报告介绍了在LHC上实施的能效改进措施(图4),改进后每单位能源能够采集更多数据。升级后,LHC在20年使用寿命内的能效将是首次启用时的10倍。但该报告也承认其中并没有真正涵盖WLCG的全部排放,仅对CERN拥有或运营的WLCG设备能耗进行了详细说明。

改变心态

Lannelongue希望越来越多的研究人员能够开始考虑计算产生的碳排放,并将其纳入决策之中。一个典型的例子就是从前研究者经常通宵运行那些效率低下的代码和软件,当被告知提高计算效率将减少碳足迹后,他们有了改变的动力。

谈及GRAND项目,Kotera表示,他们计划建立一个仿真库,让用户可以重复使用常用的仿真,而不用自己创建,这样能够避免同样的数据被不断复制。即使在大型合作中,由于没有中央存储,经常会有不同用户反复运行同一个仿真的现象。“只需按下一个按钮,就能进行为期一周的仿真,得到结果然后说‘哦,其实我并不需要它’,这实在太常见了,”Kotera说,“我们的目的在于,鼓励用户在运行之前先思考是否真正需要这次仿真。”