

数据驱动：理查孙手工NWP实践百年后的新引擎

■ 贾明群 唐伟 张萌

大约100年前，现代动力气象的鼻祖，巧妙绕过了难以得到解析解的偏微分方程组，通过向观测数据学习，用天气图上不断丰富的各自物理量的观测和计算（如通量）、借助模型对复杂问题的简化（如涡度守恒、正压近似等），让现代动力气象学，在大气运动偏微分方程数值解被得到前数十年，就被有效用于天气预报。

DOI: 10.3969/j.issn.2095-1973.2022.06.001

即将过去的2022年，是理查孙最早探索数值天气预报（NWP）的著作《数值过程天气预报》（Weather Prediction by Numerical Process，图1）发表100年。1922年，英国人理查孙通过这部著作的发表，总结了他在第一次世界大战的硝烟中，如何通过计算，尝试预报天气变化的可行之路。



图1 《数值过程天气预报》封面

这次没有成功的探索，却在科学和应用上，具有划时代的意义。NWP从那时起步，逐渐进入了人们的视野并成为很多气象学者终生探索的目标。

我们纪念理查孙在百年前开创性的工作，回顾自20世纪50年代初NWP借助电子计算机的出现获得成功，并在随后70余年持续进步的历程。面对未来，审视以人工智能（AI）为代表的新技术“是否”和“怎样”为NWP的发展带来新的机会。这样的回顾和审视具有特别的意义。首先，NWP在过去数十年里，实现以大约每10年有效预报天数延长1天的速度不断改进，在20世纪后期全面超过预报员的预报能力，在世纪之交已经接近可预报性的天花板，同样的改进速度在未来难以为继。再者，NWP模式已经成为高维、高度复杂的系统，模式运行对高速计算资源的要求，如果没有量子计算等实质性的突破，在未来同样难以为继。第三，以数据驱动和机器学习（ML）为代表的、瞄准复杂过程给出预测的新技术，有可能在使得NWP在动力驱动确定性预报发展走向“内卷”或“锢囚”时，

打开“数据驱动”的新窗户和新路径。上述演变和趋势，都在不同侧面孕育着气象和大气科学研究和发展范式的改变，不仅与快速发展的智能技术快速融入科学研究相呼应，更是具有“统计”和“混沌”视角和传统的现代气象科学，实现快速华丽转身的优势所在。

1 1920年代气象先贤靠向稀疏的观测数据学习提出天气模型并开始科学预报

100年前，在理查孙尝试数值求解大气运动偏微分方程组的同时，包括现代气象学理论的提出者皮叶克尼斯在内的早期气象学家，却另辟蹊径，尝试基于物理合理性的简化，使用绘图工具进行“图形演算”。例如，1908年准平衡近似被提出，随后，大气正压假设等框架和针对大气点和线的物理近似，也相继被应用。这些起到简化作用的猜想或定律，如果从大气运动偏微分方程组的视角考量，在某种意义上具有了“过滤”方程的效果，并支持在获得方程组的解析解或数值解之前，以简驭繁实现“图解”天气过程，这与1950年查尼等基于正压涡度方程这一简化方程的数值解法首获NWP成功具有异曲同工之效。

气象观测数据首先带来的最珍贵的信息，是系统演变背后对其发展规律的感悟、认知和提出模拟模型，这样的初步“数据训练”或“数据驱动”的过程，在针对北半球主要天气系统——温带气旋的认识过程中具有很好的代表性。百年前，气象预报员的先驱基于稀疏的地面气象观测和更少的利用热气球和风筝等得到的高空气象观测，从这些数据的蛛丝马迹中不断挖掘，尝试提出和不断改进了温带气旋模型，开启了科学意义上的经验预报，为后续包括NWP的发展打下了基础（图2）。

善于向观测数据不断学习的卑尔根学派，在历史上几乎与理查孙发表手工数值天气预报著作的同时，

页注：资助信息：国家自然科学基金（42142009）

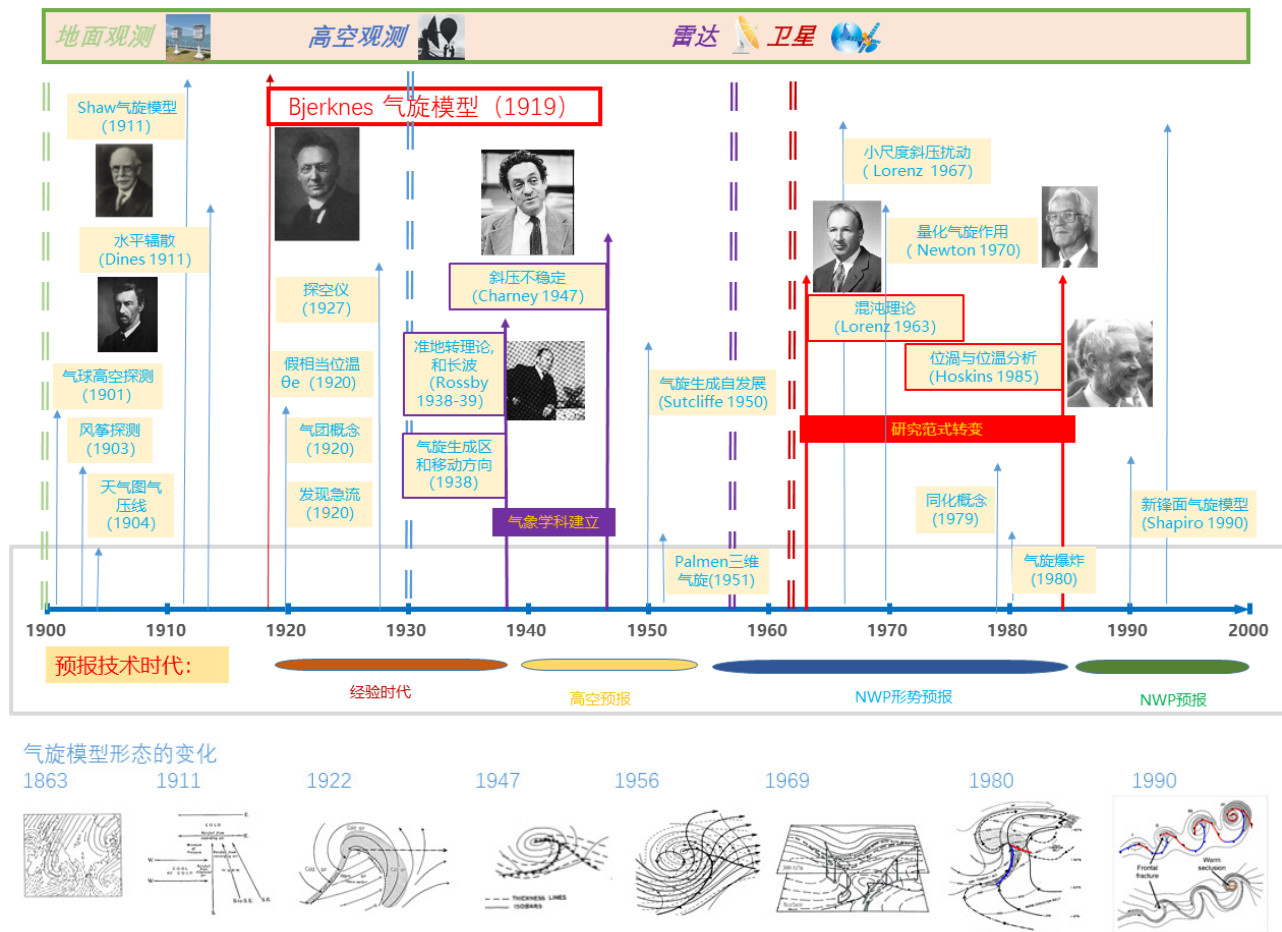


图2 现代天气学伴随对温度气旋的认知发展历程回顾 (1900—2000年)

于1920年代提出和不断完善了极锋理论，该理论基于20世纪初大气运动为物理初值问题的理论，定义了大气运动物质面上的主要特征，如冷暖峰、槽脊线等。当这些特征在不断完善的天气图上，与温度、气压等气象变量等值线碰撞，就构建了最初科学意义上的天气预报的基础。这样的基础，既有动力气象方程原理的神韵，而以数据驱动的观点看，更是基于对（按照科学原理）大气运动的观测结果，构建天气系统模型并利用模型预报天气的过程。在这里，大约100年前，现代动力气象的鼻祖，巧妙绕过了难以得到解析解的偏微分方程组，通过向观测数据学习，用天气图上不断丰富各种物理量的观测和计算（如通量）、借助模型对复杂问题的简化（如涡度守恒、正压近似等），让现代动力气象学，在大气运动偏微分方程数值解被得到前数十年，就被有效用于天气预报。

2 数据驱动或将带来动力系统的另类通解

可以说，百年前现代气象科学建立之初的两个发展方向，即理查孙数值求解方程组和卑尔根学派基于物理定律向观测学习开展天气诊断预报，或许已经再

次上演。两位主角，一方是理查孙的后辈，在查尼等人借助电子计算机使得NWP于1950年获得成功后，全球气象界一支最强大的天气气候模拟，并且开展走向地球系统模拟的研究和业务力量；另一方则是刚刚兴起，借助AI技术试图再创以数据驱动为特征的全新NWP。再次的碰撞或许更加精彩，这是因为，不仅前次双箭齐发的共同“底蕴”犹在，而且，一方面，借助于大量数据而存在的ML等技术，“吃掉”的数据中大量依靠来自动力模式的输出数据；另一方面，ML等新技术作为一种工具，还可以在完善动力模式，特别是在提高参数化精确度等方面带来突破。两者你中有我、我中有你的格局业已形成。

即将在2023年初举办的美国气象学会（AMS）年会，主题为“数据：驾驭科学、信息决策、充实人文（Data: Driving Science. Informing Decision. Enriching Humanity）”。这一主题的选择具有多方面的意义，其中的一个重点就是数据驱动科学的探究。今天，我们具有了百年前理查孙时代难以想象的数据，从数量质量、到数据的实时性、到数据的快速处理和应用。丰

富的数据本身，已经可以通过挖掘数据，对其表征主体的演化规律加以认识的阶段，而当这种认识处于目前科学研究的空白或准空白时，这种挖掘和规律认知的新范式，无疑具有了更加重要和新方向的意义。

2022年，著名科学期刊《哲学学报A-数学、物理和工程科学》出版了“动力系统中的数据驱动预测”。该专辑的内容以及该刊2021年出版的其他相关主题专辑，如“物理流体动力学中的数学问题”“复杂系统中的传输现象”“天气和气候模拟的机器学习”等，从动力学、物理学、复杂系统、计算数学和气象科学等多个视角，全面介绍了ML为代表的AI技术对包括天气气候预测等不同领域，数据驱动模拟技术革命性的促

进所涉及的各种科学问题。

数据驱动在过去10年里受到越来越多的重视。一是因为传感器、信息存储等技术进步带来大量可用数据；二是对更多复杂系统的认知过程中，迫切需要从数据中揭示潜在的物理原理或规则，特别是针对尚无法进行定量分析描述，甚至还没有找到描述的变量的系统，后者几乎成了唯一可能的手段。数据驱动的一些方法，如无方程模拟、人工神经网络、非线性回归、经验动力学模拟、范式识别、非线性拉普拉斯分析和动力学自动推理等。我们在表1中列举了与NWP有关、以AI技术为特点的一般性动力系统解决方案的相关领域。

表1 动力系统的数据驱动相关领域

技术领域	描述	面对的问题或性能	解决方案例子
降维技术	高维度复杂系统分析的重要组成部分	传统线性降维方法不适用于基本控制方程具有随时间演化的局部解特征的问题	基于定位的降阶模拟及其与模型缩减公式结合的局部降维，本征正交分解(POD)降阶模拟，平衡截断方法，非侵入式数据驱动
物理信息识别(施加物理约束)	将物理规律与数据一起嵌入神经网络建立模型	学习模型不能很好地扩展到缺乏训练数据的区域	没有优化情况下遵循基本原则的架构，基于径向基函数配置法，噪音观测中发现贝叶斯模型，短脉冲串模拟数据中提取随机支配定律
深度学习广义解决方案		无物理知识，需要大量数据；从嘈杂和不规则的观测数据中学习	基于偏微分方程无网格解的径向基函数(RBF)配置法；GPNODE机器学习方法，用于从局部、有噪声和不规则观测中发现贝叶斯模型；从模拟数据的短爆发中提取具有(高斯)布朗运动和(非高斯)Lévy运动的随机控制律
闭包问题	计算统计学的关键领域	缺乏训练数据	非马尔可夫闭包模拟框架用于准确预测受外部强迫的湍流动力系统的统计响应
动力系统稳定性分析	是预测复杂系统动力学的重要要求，对表现出分岔不稳定性的系统提供了一般和约简形式	传统的方法识别中心流形上的动力学需要一个精确的系统模型和大量的代数使用非线性系统方程	用在失稳发生前系统动力学记录的随机快照识别系统是否存在失稳风险并在中心流形上返回系统动力学的封闭形式模型；利用数据流形中的拓扑不变量，实现动态系统中的稳健模式识别和异常检测
小颗粒在流体中分散动力学	这些粒子的动力学可以通过跟踪每个粒子直接模拟特定的(采样的)色散	可以减少传统方法的计算成本	用一个统计矩的数据信息条件双曲求积方法训练一个RNN，调整基于正交的弯矩方法积分，得到更高精度的评估未闭合力矩
工程性问题		数据驱动技术为许多具有挑战性的工程问题提供了解决方案	两级、数据驱动的飞机自主着陆数字孪生
用于近似非线性微分方程解的数值时间步进算法。	时间步进方案通常基于泰勒级数展开，这些展开在时间上是局部的，并且具有由步长确定的数字累积。	许多以多尺度物理为特征的系统动力学时间尺度很宽广，这使得数值积分很昂贵。	一个基于合成多个DNN的数据时间步进器框架，用于在多个时间尺度上训练的分层时间步进器(HITS)

数据驱动预测研究在气象领域取得进步，由于物理气象预报的目标系统，向着越来越复杂、规模或体量越来越大的方向快速转变，这样的系统受到更多的高维度、高复杂性和混沌行为的限制。此外，天气气候模式能力数十年能力的持续提升，让模式输出数据比肩甚至超越观测数据，可以方便地利用各种AI技术从数据中提取动力规则、知识打下了基础。这时，AI技术在面对复杂问题的优势，如降维等技术就排上了用场。除了气象预报以外，还有数据驱动用于预测生态系统、流行病学系统、交通系统等各种复杂系统动力学中的关键转变和极端事件的例子。

这些通用的技术在大量动力系统的研究和应用中落地，涵盖了气象、生态、生物、交通和金融等大量看似形态各异的领域。其中，针对气象尤其关键的技

术，如物理约束等，我们在下节展开讨论。

3 天气气候模拟的数据驱动：硬AI、中度AI和软AI的3个融入机制

沿着理查孙开辟的道路，借助现代计算机的出现，由冯·诺伊曼和查尼等于1950年最终让NWP取得成功。NWP技术随后先后“入驻”各国气象中心，逐步成为天气预报的主要工具。不仅如此，随着这类被称为基于第一原理(first-principle)的确定性模式预报能力持续提升，导致气象统计经验模型预报日渐衰落。

神经网络(NN)等ML技术的兴起，让气象统计经验模型又见生机。借助AI技术的统计经验预报是否会取代目前的NWP动力模式，或者二者之间更多地融合、促进，携手发展，成为21世纪第3个十年到来时，理查孙开创的事业百年是否有了新拐点的学科提问。

当神经网络等AI技术面对天气气候模拟和预报问题时，预测的时空尺度依然是绕不开的问题，尤其是时间尺度，反映了面对的天气气候系统，其复杂性的主要来源和应对的主要方向。对比相对成熟的NWP，目前AI技术渗透到NWP乃至地球系统预测领域，可以划分为硬AI、中度AI和软AI的三种介入方式（表2）。不同方式即表明了传统预报被“取代”的程度，也划分出AI技术和动力模式各自优劣的领域。未来NWP的“硬核”技术是否改弦虽然还不得而知，但这些都与AI捆绑的新技术，每项都已经有很多应用的例子，诠释着数据驱动如何从3个层面，改变和改进当

前的预报。例如，中等AI通过学习观测数据和模式输出数据，尤其是高分辨率的模式结果数据，一旦能够得到更好的模式组件，就有可能将目前还知之甚少的更多复杂物理过程，例如云物理过程等给出更好的表达（算法），从而减少模式系统中参数化代表的“灰色地带”，实现对尚无物理规律可循或半经验的大气过程相对准确的模拟。

串联NWP动力模式与数据驱动的一个关键要素，是与物理约束相关的问题。确信使用的AI技术值得信任，这在气象预报相关研究和应用中尤其重要。针对更广泛问题时提出的PIML（physical-informed ML，有

表2 天气气候模拟中的硬AI、中度AI和软AI的介入

介入程度	描述	可能领域	AI面对的问题或优势
硬AI	任何尺度预测均被AI取代	临近预报	1) 可忽略物理约束; 2) 海量物联网数据可用
		季节预报	动力模式预报偏差大
中度AI	进入预报部分环节，为应用提供足够丰富的独立数据点	作为后处理工具	1) 网络映射、偏差校正、不确定性表示; 2) 生成对抗网络(GAN)提升模式或卫星数据空间分辨率
		学习观测和模式数据得到替换模型的组件	1) 大气过程和机理尚没有物理框架及解析; 2) 半经验过程算法的改进
软AI	降低计算成本等	AI方案替代参数化	节省的计算成本可用于模型复杂性和分辨率提升

物理依据的ML)和其他类似概念，如PCNN（physics-constrained neural networks，物理约束NN）等AI技术，是气象领域引入智能技术时的首选。这些概念中的PI或PC，是指对物理方程进行编码以得到残差，即动力系统的控制方程，也被AI吸纳和参与运算。而控制方程转化为损失函数，就让物理要素抵达AI核心。

在具体AI技术方案中，物理约束的程度，一般可以从NN的选择、物理知识的表达方式和物理信息的整合方法等划分，但添加了物理约束后，“硬AI”相对难以企及，这也是为什么在表2中硬AI涉及的领域，或者是“一瞬间”短临预报，或者是预报机理尚不清晰的季节预报的原因。而在地球系统中，人类尚没有掌握动力机理的过程，或大或小还有很多，因此硬AI的作用，依然不可限量。

4 结语：AI已呈现超越当年理查孙，或引领气象发展的新范式

我们今天纪念理查孙百年前完成的探索，其中一个重要的原因，就是当天气预报问题是物理初值问题已经被学界接受，但面对一组动力学和热力学非线性方程组而无能为力时，他巧用空间和时间分离散化的方法，化解了困难，从而完成了气象学历史上一次最重要的、将气象科学带入现代化的研究范式的改变：即从理论走向数值模拟。

近年来，AI技术实际上已经开始为气象科学发展范式的下一个改变做好了准备。从ECMWF当前业务

链已经和将要融入的ML要素（图3）来看，现代天气气候预报系统，尤其是包括多圈层耦合的地球系统预测，AI化已经呈势不可挡之势。特别是，地球系统中除了表征其动力核心的大气环流模式以外，更多的过程是无法用物理定律描述，甚至是随机发生的（如火山喷发等），带来更高复杂性的耦合，也正是AI技术的优势领域。

在20世纪初给出完整大气运动偏微分方程组的皮耶克尼斯，虽然没有实际尝试解这组方程来预报天气，但他在1914年还看不到高性能数值计算的曙光时就颇具前瞻性地指出：“只要（基于大气运动方程）计算与事实相符，就可以赢得科学的胜利”。他进一步用挖掘隧道进行比喻：人类挖掘穿越大山的隧道，许多工人可能无法活着见到隧道贯通，然而，这不会阻止后来者以特快列车的速度穿越隧道。如果说理查孙当年面对NWP发展隧道挖下第一铲，那么，今天更多后来者或许借助AI技术，改变隧道挖掘和加快进程的方式。虽然开始时可能存在诸多疑问，但面对AI技术在分析复杂数据、提取海量数据中的非线性关系和建立预测模型方面的强大优势，加上以数据再分析为代表的气象大数据方面的优势和考虑动力预报传统方法未来面临的后劲不足等，当前需要保持当年皮耶克尼斯的态度和前瞻视野，学习理查孙有了信念就动手实践，借AI技术曲尽原始动力方程之幽深，气象和大气科学以新范式迎接获得更大发展的未来，完全可期！

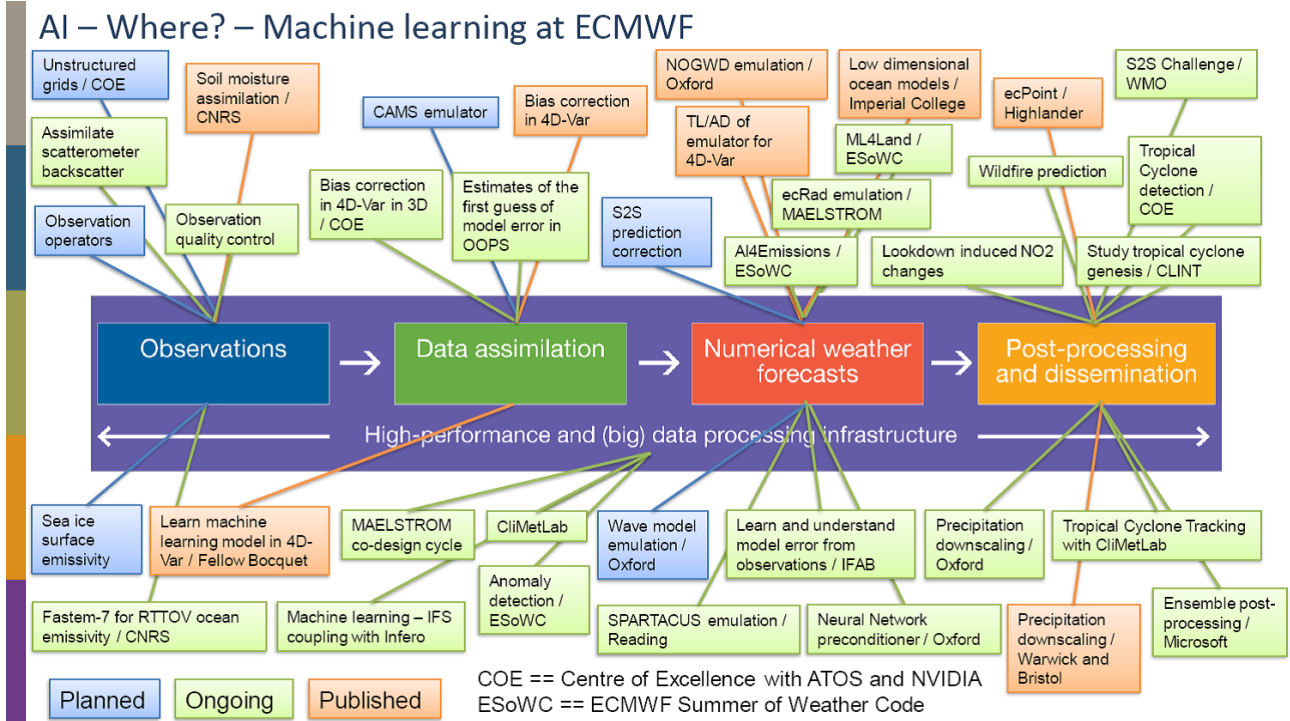


图3 在ECMWF的观测、同化、预报和后处理环节计划、进行中和已经发布的融入了AI技术的内容

深入阅读

贾朋群, 2018. “他乡”或许不再是“异乡”——ECMWF“揭开灰色地带的谜底”学术会评介. 气象科技进展, 8(1): 299-302, 304.

许小峰, 2018. 从物理模型到智能分析——降低天气预报不确定性的新探索. 气象, 44(3): 341-350.

Balaji V, 2021. Climbing down Charney’s ladder: machine learning and the post-Dennard era of computational climate science. Philosophical Transactions of the Royal Society A, 379, <https://doi.org/10.1098/rsta.2020.0085>.

Chantry M, Christensen H, Dueben P, et al, 2021. Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. Philosophical Transactions of the Royal Society A, 379(2194), <https://doi.org/10.1098/rsta.2020.0083>.

Cuomo S, Di Cola V S, Giampaolo F, et al, 2022. Scientific machine learning through physics-informed neural networks: where we are and what’s next. Journal of Scientific Computing, 92: 88.

Dueben P, 2022. Machine learning for weather and climate prediction, 7th ENES HPC Workshop, 9-11 May 2022, Barcelona, Spain.

Ghadami A, Epureanu B I, 2022. Data-driven prediction in dynamical systems: recent developments. Philosophical Transactions of the Royal Society A, 380: 20210213.

Kashinath K, Mustafa M, Albert A, et al, 2021. Physics-informed machine learning: case studies for weather and climate modelling. Philosophical Transactions of the Royal Society A, 379(2194), <https://doi.org/10.1098/rsta.2020.0093>.

Rudy S H, Brunton S L, Proctor J L, et al, 2017. Data-driven discovery of partial differential equations. Science Advances, 3(4): e1602614.

Schultz D M, Lynch P, 2022. 100 Years of L. F. Richardson’s Weather Prediction by Numerical Process, Monthly Weather Review, 150(4):693-695.

(作者单位: 中国气象局气象发展与规划院)