

# 天气和气候的人工智能基础模型:应用、设计和实现

**摘要:** 机器学习和深度学习方法在理解大气的混沌行为和天气预测方面进行了广泛的探索。科技公司、政府机构和气象机构对构建数字孪生地球越来越感兴趣。最近,使用 Transformer、基于物理机器学习和图神经网络的方法已经在相对较窄的时空尺度和特定任务中展示了最先进的性能。最近,在及时的工程设计和微调下,一种使用预训练的 Transformer 模型对语言和视觉进行建模的生成式人工智能取得了成功,目前在向通用人工智能迈进。更重要的是,我们正在见证人工智能基础模型的崛起,这些模型可以在多个特定领域的下游任务中发挥竞争力。尽管取得了这一进展,但我们仍处于为全球地球系统模型、区域气候模型和中尺度天气模型建立通用人工智能模型的初期。在本文中,我们回顾了当前最先进的人工智能方法,主要来自气象学背景下的 Transformer 和算子学习文献。我们将从成功标准的角度出发,为天气和气候预测提供一系列基础模型。我们还讨论了这些模型如何在下游任务中具有竞争力,例如降尺度(超分辨率),确定有利于山火发生的条件以及预测不同时空尺度上相应的气象现象,如飓风和大气河流。特别地,我们检查了当前的人工智能方法,并认为它们已经足够成熟,可以设计和实现天气基础模型。

## 目录

1 简介及概述 .....	1
2 应用 .....	2
2.1 预测 .....	2
2.2 模型混合和后处理 .....	4
2.3 降尺度 .....	4
2.4 参数化 .....	4
2.5 数据同化 .....	5
2.6 天气模式的检测和预测 .....	5
2.7 将机器学习纳入气候科学 .....	5
2.8 机器学习的影响和应用模型 .....	6
2.9 天气和气候应用基础模型的好处 .....	6
3 设计 .....	7
3.1 基础模型的气象学视角 .....	9
3.1.1 物理尺度 .....	9
3.2 长期预测 .....	10
3.3 数据 .....	10
3.3.1 观察、再分析、预测数据 .....	10
3.3.2 数据量 .....	10
3.4 诊断 .....	11
3.4.1 架构评估 .....	11
3.4.2 预训练评估 .....	11
3.4.3 与物理基准比较 .....	12
3.5 影响、作用 .....	12
3.5.1 机器学习架构的关键特征 .....	12
4 实现 .....	13

4.1 数据表示 .....	13
4.1.1 坐标表示 .....	13
4.2 模型组件 .....	13
4.2.1 Transformers .....	13
4.2.2 图 .....	15
4.2.3 神经算子 .....	15
4.3 预训练机制和损失函数 .....	16
4.3.1 rmse 型损失:重建与预测 .....	16
4.3.2 替代方法 .....	16
4.3.3 扩散 .....	17
5 总结 .....	17

## 1 简介及概述

在大型数据集的可用性和 GPU 驱动计算的快速革新的推动下，深度学习在 2010 年代取得了快速进展和各种突破。在这场革命之后，自然语言处理（NLP）领域利用 Transformer 和自监督学习建立大型语言模型（LLM）的研究工作促使了基础模型范式的出现[1]–[3]。正如文献[4]所定义的，基础模型是一种以自监督（即与任务无关）的方式在大量数据上训练出来的模型，可以有效地对多个下游任务进行微调。一般来说，基础模型由数百万甚至数十亿个参数组成。

大型语言模型最近从这些努力中获益最多，但天气和气候领域也拥有海量数据、超级计算和机器学习方面的经验。此外，尽管数值天气预测（NWP）在模拟大气的物理过程建模方面取得了巨大成功，但早在 20 世纪 70 年代，就有着使用数据驱动的机器学习方法，将 NWP 的输出后处理为模型输出数据[6]。此外，网格化天气和气候模型输出的数量与日俱增，所有主要国家的气象、气候服务和研究实验室都为这一增长做出了贡献。欧洲中期天气预测中心（ECMWF）的档案包含约 450PB 的数据，且每天新增 300TB[7]。本文特别回顾了利用这些数据集解决天气和气候预测问题的基础模型的开发情况。我们主要聚焦于检测和预测极端天气事件（如飓风、大气河流、火灾天气条件）及它们对人工智能与 NWP 混合的影响。

采用基础模型范式可以带来各种优势。许多应用在模型性能和精确度上取得了突破，这也可能产生其他优势（例如，数据效率）。事实上，对于数据有限的许多应用场景（无论是标签还是协变量），基础模型范例为开发健壮有效的模型[1]、[2]、[8]–[10]提供了一条清晰的路径。这是由于广泛的自监督的预训练，它减少了微调特定任务所需的标记数据量[11]、[12]。最后，即使数据充足，训练大型模型的成本也很高。在这篇综述的中，我们会介绍在数十个 GPU 上接受了数周甚至数月训练的模型。遵循基础模型方法意味着只需进行一次大规模训练，而不是承诺每个应用的成本，因为可以在数小时或数天内少数 GPU 上进行微调。目前与 NWP 模型竞争的最先进的人工智能模型正在涌现[13]–[16]。然而，目前这些模型受到物理规模的限制，难以长期进行预测，且不能推广到多个下游任务。

基础模型通常由编码器  $f_\theta$  和解码器  $g_\phi$  组成，其中  $\theta, \phi$  是参数。预训练基础模型后使用一些自监督策略优化  $g_\phi \circ f_\theta(1)$ 。然而，当针对下游任务微调模型时，要保留  $f_\theta$  并舍弃解码器  $g_\phi$ ，并用特定任务的解码器  $h_\psi$  来替换它。也就是说，微调会优化  $h_\psi \circ f_\theta$ 。通常，解码器是非常轻量级的，而编码器可以包括数亿甚至数十亿个参数，因为它们需要理解和表示输入数据的复杂性。因此，基础模型通过仅使用特定任务的解码器来减少训练开销。为了理解这个过程为什么有效，我们可以认为编码器学习预训练期间数据中存在的动态和关系，而微调则教导解码器将这些关系应用于特定任务。当然，这个阐述是不足的，仅有助于概念上的理解。尽管如此，应该清楚的是，基础模型在很大程度上根植于自监督和表征学习。同样，还应该清楚的是，用于预训练

前的数据需要使学习到的分布能够在某种程度上表示实现所有下游任务模型所需的所有变量和内在物理机制。

尽管在天气和气候领域有大量的数据和众多的应用，但仍有必要在这样的背景下进一步探索自监督学习。最近，用于天气预测的大模型取得了重大发展，但这些模型主要依赖于监督学习。不过，该领域也有一些表征学习的例子，如文献[17]–[19]。目前，公开可用的天气和气候基础模型只有[18]、[20]。

本文的其余部分安排如下。第 2 节讨论各种天气或气象应用，特别关注机器学习和深度学习的适用性。考虑到这一系列潜在的下游任务，第 3 节讨论了在建立基础模型时需要做出的选择。虽然期待最终出现几种能模拟地球系统所有方面的基础模型并不现实，但目前还是强烈建议要确定一系列现象和可能的应用作为目标。确定新基础模型的范围和对其实现的一些具体要求将在第 4 节中讨论，我们将在该节中讨论如何权衡各种人工智能主要方法，例如 Transformer 和图神经网络。

本文在撰写时考虑到了多个读者群体。人工智能、天气和气候研究人员和领域专家会对特定应用感兴趣。有些人可能想阅读第 2 节和第 3 节，而跳过第 4 节中深度学习模型的技术细节。另一方面，人工智能研究人员可以略读第 2 节，重点阅读第 3 和第 4 节。

## 2 应用

我们将研究潜在的应用，以开始我们对天气和气候基础模型的讨论。在这里，我们关注的是天气和气候领域可能的下游机器学习任务意义上的应用，包括将天气/气候信息与其他数据相结合的应用（例如，预测影响）。我们回顾了这些模型，以了解当前机器学习模型的优点和缺点，使用基础模型方法对每个应用的潜在优势，以及基础模型的设计成果。

虽然以下回顾并不详尽，但我们借鉴了几十年的经验，重点关注了对科学界和商业界都具有重要意义的少数应用。本节分为三个部分：天气预测（第 2.1 节），改进 NWP 模型及其输出（第 2.2、2.4、2.5、2.3 节），天气模式检测（第 2.6 节），模拟天气/气候影响和其他天气驱动的应用（第 2.8 节），以及气候科学中的应用（第 2.7 节）。

### 2.1 预测

在考虑天气和气候的机器学习应用时，预测是首先想到的问题之一。从人工智能研究人员的角度来看，这可能是因为人工智能中最直接的问题往往是那些标签和既定基准分数都很丰富的问题，就像 NWP 中的情况一样。然而，即使人们忽略了没有一个单一的分数的分数可以准确地反映给定天气预测模型的性能，人们也需要记住，衡量模型的基准不是其他人工智能模型，而是高度复杂的模拟系统，这些系统是数十年来在通常是大规模和昂贵的高性能计算 (HPC) 系统上运行的研究结果 [5]。

在不同的预测时间尺度上，天气预测需要不同的方法和模型。根据最近的观测或对未来几小时的预测建立当前的天气状况被称为短期预测，它通常利用同化观测的数值大气模型。区域 NWP 模型用于预测未来几天的情况。例如高分辨率快速刷新模型 (HRRR) [21] 和高分辨率确定性预测系统模型 (HRDPS) [22]，它们覆盖北美不同地区，分别以 3 公里和 2.5 公里的分辨率提前 48 小时进行预测。覆盖全球的模型包括 ECMWF 的高分辨率模型 (HRES) 和 NCEP 的全球预测系统 (GFS)，前者以  $0.1^\circ$  分辨率（约 10 公里）预测 10 天，后者以  $13^\circ$ （约  $0.125^\circ$ ）预测 16 天。由于较高的不确定性，次季节性和季节性预测模型更多地依赖于较长时间范围（未来数周至数月）的集成成员。集成成员是通过在扰动初始条件下多次运行模型生成的。例如，第二版气候预测系统 (CFSv2) 包括四个集成成员 [23]。CFSv2 是一个完全耦合模型，代表地球大气、海洋、陆地和海冰之间的相互作用。同样令人感兴趣的是大气再分析数据集，其中使用数值模型来吸收观测并创建对地球-大气系统状态的长期一致的最佳估计。

正如我们稍后在 3.1.1 节中讨论的那样,预测时间尺度的选择对建模方法和设计具有潜在的深远影响。在最短的时间尺度上,传统的 NWP 很难提供有用的提前时间,并且可以使用所谓的欧拉和拉格朗日持久性来从雷达图像中进行高性能的降水实时预报[24];也就是说,没有解决大气过程的全部复杂性,更不用说大气和地球表面之间的相互作用了。

考虑到这些,使用人工智能模型进行预测的第一次成功集中在短期预测上也就不足为奇了。在这里,来自计算机视觉的技术,如光流和相对简单的深度学习模型,允许将气象模式(由图像定义)传播到未来几个像素,而无需解决底层系统的全部复杂性[25]–[32]。

虽然物理系统和涉及的大量数据可能看起来很复杂,但人工智能在解决预测挑战方面的潜力有一个令人信服的论点。运行预测 NWP 模型通过求解一系列微分方程来吸收观测和先前预测,同时优化成本函数,为连续天气预测生成新的初始条件。这个过程本质上可以提炼成一个简单的函数,表示为  $x_{t+1} = f(x_t)$ 。鉴于神经网络几乎有能力近似任何表现良好的函数,一个适当的模型至少应该有能力处理将状态预测到未来的任务,只要它能访问足够多样化的训练数据集。(更复杂的数据同化问题将在 2.5 节讨论)

事实上,过去几年已经出现了 NWP 系统的 AI 模拟器[13]–[15], [33]。非常大的深度学习系统将时刻  $t$  的大气状态作为输入,并将其传播到未来的一步。通常,这些深度学习系统是在重新分析数据上进行训练的,其中  $x_t$  是 ERA5 参数的一个子集[34]。从技术上讲,这些模型不学习模拟大气过程。他们接受再分析系统的网格输出训练,学习模拟模型流体动力学和其他物理过程。因此,使用 ERA5 数据上的 RMSE 损失进行训练的性能最优的 AI 模拟器将重现 ERA5 数据及其所有偏差。

人工智能预测领域取得了重大进展,如[13]–[15], [33]等显著工作,引起了气象界的关注。虽然一些研究人员认为他们的模型优于现有的 NWP 系统,如 ECMWF 的综合预测系统(IFS),也称为 HRES,但我们可以自信地说,这些模型的表现与 IFS 和类似的系统相当。评估天气预测是一项复杂的任务,人工智能模拟器有几个优势,使它们具有吸引力。最重要的是它们的速度。尽管需要在多个 gpu 上进行数周或数月的训练,但人工智能模拟器可以在几秒钟内使用一个或几个 gpu 进行预测。这样的优势使得人工智能插值比高性能计算系统上的 NWP 预测速度快三个数量级。

然而,与任何新兴研究领域一样,所有现有的人工智能模拟器都有明显的局限性。首先,空间分辨率受限于训练数据,由于具有较长的全球覆盖时间序列(1940 年至今)和良好的空间分辨率(25km),上述所有模型的训练数据都是 ERA5 数据。因此,与 HRES 的 0.1 度相比,这些模型的分辨率最多为 0.25 度,更不用说像 HRRR 和 HRDPS 这样的区域模型的千米尺度分辨率了。(参见 2.3 节,了解降尺度的潜在应用)

此外,人工智能仿真器通常只对参数、时间步长和垂直水平的子集进行建模。究其原因,不外乎以下几个方面。ERA5 数据分布在 37 个气压级别上(注 1),代表每个网格点的 6 个主要参数(注 2),这意味着要在 32 位分辨率下提交大约 900MB 的内存。通常,这是通过控制压力水平、时间步长和参数来解决的。这就导致了上面提到的局限性。

(注 1:虽然 NWP 通常使用地形跟踪垂直坐标系统,但将气象场插值到气压水平更适合预测员进行分析。为了我们讨论的目的,意识到一般使用的 NWP 模型输出通常分布在固定的气压水平上就足够了)

(注 2:东风、北风和垂直风以及温度、湿度和位势。)

除了空间或时间分辨率的丰富性以及参数数量的问题之外,人工智能模拟器目前在较长的交付时间内表现出较差的性能。随着提前期的增加,预测变得越来越“模糊”,而且它们无法像集合 NWP 预测那样捕捉到极端情况。这主要是由于均方根误差(RMSE)训练标准造成的,我们将在 3.2 节中回到这个问题。最后,对一些模拟器进行训练,使不同的交货时间被分开处理,这可能导致时间不一致。

如上所述,像 ERA5 这样的数据集是使用同化观测数据的 NWP 模型创建的,以产生物理上一致的最佳大气状态估计。因此,这些数据集包含观测偏差和模型偏差。然而,请注意,有一些深度学习模型直接使用观察结果,正如我们之前在短期预测和较短交货时间的背景下提到的那样。一个例子是利用雷达数据[29],[31]对降水进行基于学习的短期预测,[30]的作者通过将这些数据与 NWP 分析相融合对其进行了扩展。最后,[32]在使用稀疏观测的预测中包含了更多的天气变量。这些模型在空间尺度上与对流分辨 NWP 模型相同或更精细,尽管后者[32]将密集网格数据与来自气象站的稀疏点数据结合在一起。关于气候模型的进一步讨论见第 2.7 章。

## 2.2 模型混合和后处理

虽然基于人工智能的熟练预测是最近发展起来的,但通过将数据驱动技术应用于 NWP 系统的输出来实现卓越性能的方法已有很长的历史。实际上,我们可以把它们看作是站在巨人肩膀上的轻量级方法。

各种模型的输出统计可以追溯到 20 世纪 70 年代[6]。这些统计数据将模型的预测和其他协变量作为输入,然后训练机器学习模型来预测观察结果。该方法的直接扩展是模型混合,其中模型考虑多个 NWP 模型的输出。详情参见[35]。

无论考虑单个 NWP 模型还是多个 NWP 模型,这些模型通常都比它们的输入表现出更好的性能分数。由于机器学习模型完全不知道 NWP 模型中编码的各种约束来进行预测,因此它们在追求优化的过程中可以自由地违反这些约束(尽管如果可以适当定义的话,确实存在强制执行硬约束的技术;参见[36])。当事情进展顺利时,模型就能够修正模型偏差。同时,这些模型可以自由地做出与大气动力学和物理学不相容的预测。

## 2.3 降尺度

在 2.1 节讨论人工智能仿真器时,我们指出当前的仿真器在空间分辨率和训练数据方面受到限制。解决这个问题的一种方法是降尺度,也称为超分辨率。主要是用来提高一些网格化天气数据的分辨率,使其超出单纯插值所能达到的水平。这项任务具有挑战性,部分原因在于不同 NWP 模型之间的实现差异。因为,例如, HRRR 和 ERA5 吸收不同的观测结果,并以不同的方式实现大气的控制方程,所以训练人工智能来预测 ERA5 的 HRRR 是不够的。该任务不仅需要学习超分辨率关系,还需要将一组模型偏差和误差转换为另一组模型偏差和误差。需要创建源数据一致的超分辨率训练数据集,通常是通过将单个 NWP 模型插值到其他网格。

最基本的深度学习方法使用卷积神经网络,通常是 U-Nets[37],从低分辨率输入生成高分辨率输出。后来,生成对抗网络被发现是一种很好的途径,可以增强更精细的结构和细节的表示[38],[39]。最近的工作利用了各种各样的人工智能概念,如归一化流[40]和神经算子[41]。[17]的作者使用自监督学习获得的表征进行降尺度,[20]也探索了这一途径。

## 2.4 参数化

在前两节中,我们的讨论集中在改进预测模型上。这是通过提高分辨率或在观察的背景下放置输出来实现的。在这个简短的部分和随后的 2.5 节中,我们将通过讨论 AI 如何成为模型本身的一部分来结束关于如何改进 NWP 模型的讨论。这与我们的讨论形成了对比,在我们的讨论中,NWP 模型在很大程度上被视为黑盒,输入和输出被输入到 AI 系统中。

一个关键的研究领域是参数化,这意味着用人工智能模型来增强和支持或加速 NWP 系统[42]。参数化是使用算法或统计方法来表示物理过程,其中子网格规模的过程、复杂性或缺少低调说明阻碍了模型中的显式表示[43]。考虑到它并不总是依赖于第一原则,NWP 参数化对于数据驱动的方法来说是一个成熟的领域。例如,[44]使用在高分辨率云分辨模拟模型上训练的深度学习模拟器来取代气候模型中云的参数化。此外,正在开发诸如 ClimSim 之类的基准数据

集来辅助这一领域的研究[45]。更近期的研究见[46]–[48]。

## 2.5 数据同化

正如前面提到的，一个成熟的 NWP 系统不仅将一个状态传播到未来；它还吸收观测数据来建立初始状态，并将模型的演化约束到未来状态。这个过程被称为数据同化。数据分析过程复杂、昂贵，因此是一个正在进行的研究领域。此外，流行的方法依赖于简化假设，如线性。虽然这是一个具有挑战性的领域，但在这方面利用深度学习的研究越来越有前景。看到[49]–[51]。

## 2.6 天气模式的检测和预测

在再分析数据和气候预估中自动检测极端天气特征有助于更好地量化未来气候中风暴和复合灾害造成的极端气候影响的风险[52]，[53]。传统的方法是在再分析或气候模型预测数据中使用系统检测算法和经验启发式方法来检测锋面、热带气旋(tc)、热带外气旋(ETCs)和大气河流(ARs)等特征(例如[54])。此外，用于预测特定特征(如龙卷风或飓风)的专门系统可能比一般的再分析训练模型更有优势，这些模型可以预测整个大气状态，例如获得像雷达这样的高分辨率观测。

最近，深度学习在天气特征检测中的应用已经被探索，最终提供了一个更通用的解决方案，避免了启发式方法对特定变量和阈值的依赖，这些变量和阈值通常是特定于数据集的。[55]提出了一个专家标记的 25km 的 CAM5.1 模型的热带气旋(tc)数据集，并演示了使用计算机视觉应用中为空间分割而开发的架构对 tc 和 ar 进行快速分割。[56]实现了 tc、etc、ar 和热带低气压的检测和定位。[56]使用 3D 卷积模型，将时间作为其中一个维度，[55]在静态 2D 图像上训练，但在模型预测中报告了合理的时间一致性。天气特征检测的其他例子包括使用 CNN 和逻辑回归的对流解析动态预测中的龙卷风诊断[57]，以及区域或高分辨率天气预测中的湍流条件诊断[58]。在后一种应用中，作者将预先计算的湍流指数与传统的机器学习方法(例如 XGBoost)结合使用，因此使用能够进行无监督特征提取的更复杂的机器学习方法有可能进行改进。

除了简单地检测天气特征外，人工智能模型还被应用于预测天气特征的演变。例如，利用 CNN 在复合雷达[59]和多模态数据[60]上完成了龙卷风的短期预测。飓风或热带气旋(TC)预测主要侧重于预测风暴的轨迹和强度(例如，最大风速)。机器学习已被用于单独或组合为这些模型建模，作为元多模型框架的一部分[61]。这些模型使用的数据集包括再分析数据，辅以 TC 观测和实际飓风预测(包括统计和动力模型，分辨率高达约 3 公里)。因此，由于再分析数据集中 TC 的代表性较差，任何与飓风或 TC 预测有关的下游任务都需要消耗更高分辨率的动态模型数据[62]。再次提到的挑战包括监督学习的极端事件数量较少，以及需要整合更广泛的数据集。

训练监督检测模型的一个挑战是标记数据的数量有限，特别是对于极端事件。与特定任务的监督深度学习模型相比，任务独立的基础模型方法可以减少有限标记数据的挑战。我们还注意到，以前的研究倾向于在固定分辨率下对单个模型的输出进行训练和测试。多数据集、多尺度基础模型方法也可以实现更好的泛化(微调可以学习特定数据集的特定偏差和重要变量)。

需要解决的一个问题是，设计用于将预测模拟作为下游任务的基础模型是否足够广泛，以实现天气特征分割，例如大气河流识别。我们还注意到，只要根据需要将模型应用于 NWP 或气候模型预测数据，天气/气候特征检测在预测时间尺度上就具有潜在的价值。

## 2.7 将机器学习纳入气候科学

机器学习正在成为气候科学的关键工具，增强我们对气候系统的理解，促进更快、更准确的气候预测。机器学习技术在分析广泛的气候数据集以确定模型和趋势、提高气候模型的精度、预测不同排放情景下的潜在气候变化以及评估气候变化对不同地区和部门的影响方面发挥着重要作用。CMIP 项目的最新出版物，包括[45]，[63]，[64]的作品，强调了机器学习在解决

气候科学各个方面的日益整合，从不确定性量化，通过基于机器学习的参数化加速模拟，设计模拟器，开发混合机器学习-物理研究的数据集和工具，不确定性量化到使用机器学习识别气候变率模式的可预测性来源。

尽管前景光明，但在气候科学中使用机器学习并非没有挑战。气候系统的复杂性和通常有限或不完整的气候数据对开发能够恰当地表示气候系统行为的机器学习模型构成了重大障碍。此外，气候预估中固有的不确定性仍然是一个关键的限制，尽管机器学习可以通过减少与排放情景相关的不确定性在一定程度上缓解这一限制[65]。

然而，随着不断的进步，机器学习有望进一步扩大其在气候科学中的作用。随着机器学习技术的不断发展和更全面的气候数据的预期可用性，机器学习将巩固其在增强我们对气候变化的理解和做出更精确的未来气候变化预测方面的重要性。

CMIP6 项目产生的丰富的气候数据库是增强机器学习模型在各个领域的有效性的重要资源，包括预测即将到来的气候变化、识别极端天气事件和评估气候变化的影响。

利用 CMIP6 作为基准，可以对机器学习模型的性能进行细致的评估，从而促进持续的改进和创新。此外，CMIP6 促进了合成数据的创建，这是在实际数据不足或不可用的情况下训练机器学习模型的关键资产。这种共生关系也激发了新型机器学习技术的发展，增强了理解和预测气候等复杂系统的能力。这一整合标志着在利用计算智能增强我们对气候动力学的理解和制定数据驱动的气候适应战略方面的飞跃。最终，这可以为气候适应战略[66]、影响评估[67]和应对气候变化的解决方案[65]提供信息。

## 2.8 机器学习的影响和应用模型

为天气和气候的影响和应用开发机器学习模型涉及将天气和气候数据转换为适合特定应用系统的变量，这些变量虽然响应大气，但本身不是大气变量。这还可能包括整合额外的非气候驱动因素和相关数据集。这些系统的复杂性可能会有很大的变化，导致后续建模任务的复杂性发生变化，并影响将基础模型中的预训练权重纳入这些特定应用的潜在好处和相关性。

对洪水和野火等复杂灾害进行建模通常涉及使用具有多变量输入的特定复杂数值模型。例如，机器学习作为洪水预测的工具已经很好地建立起来，但通常被用作多阶段混合建模框架的一个或多个组件(例如[68])，或作为水文模拟的原位替代品[69]。对流程(如洪水)进行建模本身可能需要一个单独的基础模型。因此，这些应用并不是下游基础模型任务的明显候选者，尽管它们仍然可以从基础模型衍生的预测或缩小的数据输入中受益。

机器学习已广泛用于太阳辐射、风和功率输出预测，其时间尺度主要集中在几分钟到几小时之前[70]。然而，使用中尺度高分辨率模型(例如 NCAR 的天气研究和预测(WRF)模型)进行太阳预测，其中云的演变通常超过 6 小时，并且需要能量调度来捕捉峰值电价[71]，可以使用大气动力学和辐射传输的机器学习替代品来加速。此外，基于季节和气候变量(如气溶胶和尘埃)的太阳能资源评估[72]可以受益于基于机器学习的方法[73]。对于有限的太阳光度计和气溶胶观测网络，使用基础模型的预训练可以实现有限标签的推理。

## 2.9 天气和气候应用基础模型的好处

基础模型正在成为一种关键工具，具有显著提高天气和气候应用的效率和精度的潜力。下面，我们概述了我们认为对天气和气候领域很重要的潜在好处：

1. 提高准确性和效率:基础模型在预测不同时间尺度的大气变量方面提供了更高的准确性和计算效率。我们认为，与传统的数值模型相比，这提供了显著的竞争优势。
2. 缓解监督瓶颈:由于在广泛的数据集上进行了预训练，这些模型通过缓解监督瓶颈来增强其鲁棒性。减少对监督学习的依赖不仅减少了对大量标记数据的需求，而且还提高了它们在实际部署中的性能。

3. 多功能性和适应性:基础模型可以很容易地进行微调,以解决各种任务,包括天气预测、气候建模、危害分析等。它们可以处理地球大气研究中的各种场景和情况,提供了宝贵的多功能性。特别有利的是,与传统的机器学习模型相比,它们需要的标记数据要少得多,并且减少了额外的监督。这种适应性和多功能性使基础模型对地球大气研究的研究人员和实践者很有价值,简化了将人工智能解决方案应用于该领域的复杂挑战。

4. 改进的泛化:这些模型显示出优越的泛化能力,可以管理训练分布之外的变化,例如,提高预测极端天气事件和长期气候预测的可靠性。

5. 丰富和多模态的数据利用和处理:基础模型可以整合基于物理的气候模拟模型进行预训练,提供丰富的数据来源,增强学习过程并实现更明智的预测。他们通过创新技术整合不同数据类型和不规则数据集,熟练地处理气候数据的高度多模态性质,从而提高模型性能。

6. 可扩展性和创新:基础模型展示了良好的可扩展性,可以提供更大、更高分辨率的数据集的增强,预示着地球系统科学中新一代多尺度和多物理场数据驱动模型的出现,以及最近的创新,如 FourCastNet[74]、Panguweather[15]和 GraphCast[13],描绘了该领域的快速发展。

7. 更广泛的影响和潜力:将现有的气候科学知识 with 机器学习模型相结合,有可能实现更快的推理和数据同化能力,可能对气候科学和政策制定产生更广泛的影响,包括促进政府规划和社会相关的减灾工作。

8. 与多感官数据整合的潜力:基础模型的新兴趋势暗示了天气和气候科学中与多感官数据整合的前景,有望扩大未来预测和分析的范围和准确性。

### 3 设计

设计天气和气候的基础模型需要考虑第2节中讨论的相关尺度。地球系统在从几秒到几个世纪的各种时间尺度上运行,并涵盖从微观到全球行星尺度的不同空间范围。考虑与太阳系相关的影响地球天气的外部循环也很重要。因此,通过基础模型确定机器学习可以增加最大价值的领域,并考虑可以在社会和商业影响方面进行可行微调的下游任务至关重要。同样,我们需要提供足够的预训练数据,以显示足够丰富的现象,以满足下游任务的所有要求。通常,这意味着涵盖数十年的更长时间的时间序列,例如 ERA5、MERRA2 甚至 CMIP6。展望未来,我们可能会为地球的子系统(大气层、水圈、冰冻圈、生物圈和地圈)配备单独的基础模型系统。

在第2.1节中,我们深入研究了与仅适用于 CONUS 的 HRRR 相比,在全球范围内获取 ERA5 和 MERRA-2 等高时空分辨率数据集的挑战。CMIP6 为预训练大型模型提供了绝佳的机会,但处理多个源之间不一致的变量具有挑战性。因此,社区技术最先进的方法是在 10-14 天内(小于次季节)对从几小时到几周的时间尺度进行建模。图1说明了地球系统模型错综复杂的多尺度复杂性。为了开始我们的讨论,我们检查[15]、[33]、[75]-[78]中提出的模型以确定其设计标准。

Pangu-Weather[15]引入了一种基于人工智能的新型中期全球天气预测方法,其性能与传统方法相比具有竞争力。NWP 方法在准确性方面显著优于 NWP 在推理计算效率方面。关键的创新是 3D 地球专用 Transformer (3DEST),这是一种深度学习架构,旨在处理大气数据的三维性质。通过将高度制定为单独的维度,3D 模型可以有效地捕捉天气数据中的复杂模式以及不同压力水平下大气状态之间的关系。该模型的地球特定位置偏差增强了其处理天气数据中地球特定特征的能力。分层时间聚合策略还减少了中期天气预测所需的迭代次数,从而实现更快、更准确的预测。盘古天气展示了稳健的确定性预测结果,与业务数值天气预测系统的性能相媲美。集成 3D 深度神经网络、地球特定先验和分层时间聚合,使盘古天气能够实现准确的全球天气预测,同时显著降低计算成本。

[75]中的方法利用数据中编码的时间动态来训练用于动力系统预测的扩散模型。该方法通过将时间动态与网络中的扩散步骤直接耦合来实现多步骤和远程预测能力。这种设计选择增强了灵



活性和连续时间采样轨迹，并允许在推理过程中权衡性能与加速采样。此外，与传统的基于高斯噪声的扩散模型相比，动态信息扩散过程提高了计算效率。该方法展示了各种系统复杂动态

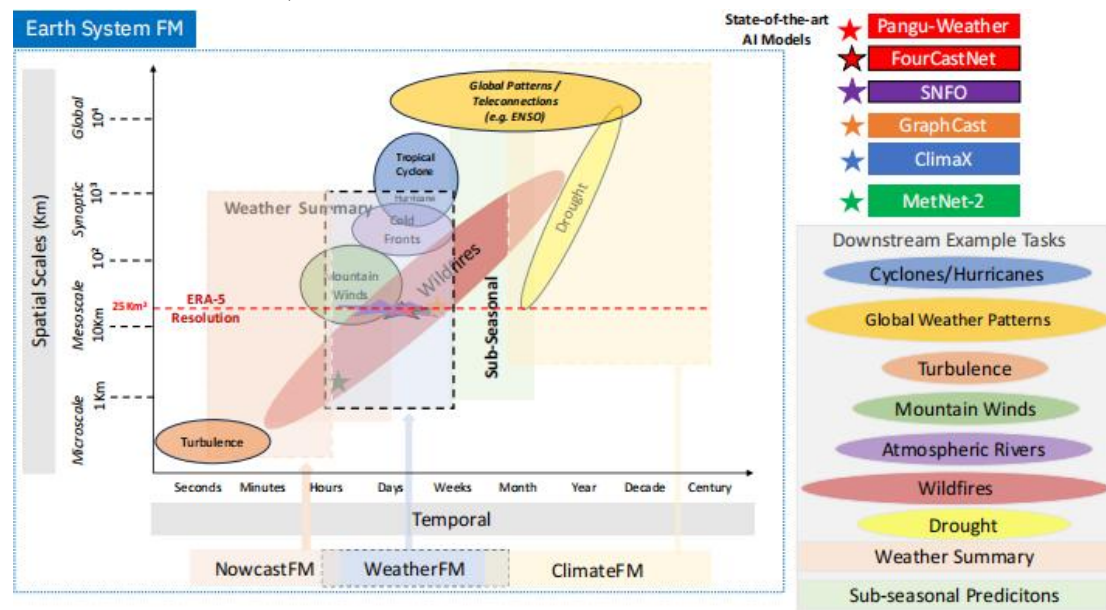


图 1：气象学和气候科学对基础模型的看法，具有时空焦点窗口。天气基础模型是最成熟的，但正如所指出的，当前使用算子学习、Transformer 和图表的最先进的人工智能模型大多围绕更广泛的窗口内的红色 ERA5 时空线。基础模型范围：地球系统（蓝色点/框）；短期预测（橙色破折号/框）；天气（黑色破折号，蓝色框）；气候（黄色破折号/框）。

预测的竞争性能，包括海面温度、纳维斯托克斯流和弹簧网格系统。这凸显了将大气动力学纳入扩散模型以提高预测精度和效率的有效性。几何克利福德代数网络（GCAN）[76]旨在通过利用几何代数的对称群变换来对动态系统进行建模。该方法引入了组动作层，通过将等距编码为  $\text{Pin}(p, q, r)$  组元素，使用预先指定的组动作线性组合对象变换。这些层以及独特的激活函数和归一化方案是可调整的几何模板，可以通过梯度下降进行细化。该方法在三维刚体变换建模和大规模流体动力学模拟方面显示出显著的优势，优于传统方法。通过结合几何代数和对称群变换，GCAN 提供了一种有前途的编码大气动力学和模型动力学的方法，从而从气象角度提高天气预测的性能。GCAN 的性能优于复杂动力系统的传统模型，例如三维刚体变换和大规模流体动力学模拟。它们利用对称群变换和几何代数，提供理论优势并提高性能。引入组动作层作为可调整的几何模板增强了它们在表示动态方面的灵活性和表现力。然而，尽管克利福德代数的前景广阔，但这项工作仍处于起步阶段，在没有进一步研究的情况下尚未为现实世界的操作系统做好准备。

[77]中的方法使用自监督学习（SSL）和联合嵌入方法从异构数据中学习偏微分方程（PDE）的通用表示。该方法可以使用 SSL 从各种来源提取有用信息，包括对真实动态系统的混乱或不完整的观察。学习到的表示在不变任务（例如回归偏微分方程系数）中优于基线方法，并提高了神经求解器的时间步进性能。这种方法显示出为偏微分方程开发通用基础人工智能模型的前景，通过为数值求解器提供计算效率高的替代方案，可能对包括天气预测在内的科学和工程产生广泛影响。

最近关于基于 Transformer 的天气预测架构的文献提出了几种开发和集成注意力机制的方法。已成功应用于天气领域的方法主要可分为 3D 自注意力[15]和 2D 交叉注意力[20]。这些方法特别成功，因为它们的目的减少与数据密集型领域中的注意力机制相关的计算成本。通过采用基于窗口的注意力（Swin Transformer[9]）或使用变量聚合和跨空间位置的交叉注意

力[20]可以减少计算量。对于基于窗口的方法，特征被划分为窗口，并且注意力机制（即自注意力）应用于每个窗口。这种方法已成功应用于非球形[15]和球形[79]设置，并且通常表现出有希望的性能。利用变量聚合和交叉注意力来降低普通视觉 Transformer（ViT）架构注意力机制的复杂性。值得一提的是，除了计算复杂性之外，注意力层还注入了高模型精度，捕获长期关系，可以处理不同类型的数据，可以包含不同类型的附加偏差，设计上是并行的，并且主要不需要傅立叶变换，允许模型处理高分辨率数据[79]。

球形傅里叶神经算子（SFNO）是傅里叶神经算子（FNO）的推广，用于在球形几何上模拟大气动力学[33]。由于 FNO 能够捕获远程依赖关系，因此它们已成功地有效地学习了时空数据的运算符。然而，它们在球坐标方面受到限制，导致视觉伪影和耗散。SFNO 通过在紧凑黎曼流形上使用广义傅里叶变换（在本例中为球体上的动力系统）来解决这些问题，并展示了大气动力学的稳定自回归预测。从气象角度利用 SFNO 预训练基础人工智能模型可以增强大气动力学和模型动力学的表征，从而改善天气预测和气候模拟。这一进步可以加速气候模型的开发，并有助于更有效地理解和应对气候变化。

卷积神经算子（CNO）提供了一种新颖的 CNN 架构改编方法来学习偏微分方程的解算子[78]。与传统神经网络不同，CNO 可以将函数作为输入和输出进行处理，同时保持其连续性，即使是离散形式。当使用多尺度解决方案对各种 PDE 基准进行测试时，CNO 的性能显著优于基线方法。这表明 CNO 为稳健且准确的算子学习提供了一种有前途的替代框架，在特定应用中比 SFNO 具有优势。与 FNO 相比，CNO 通过保留函数的连续性质（即使是离散形式）来最大限度地减少混叠错误，从而减少伪影。虽然 CNO 在大多数指标上都优于 DeepONets 和 FNO，但 CNO 和 SFNO 之间的基准测试尚未进行。

### 3.1 基础模型的气象学视角

地球系统基础模型空间有一个相对开放的运动场，技术成熟可行，可以建立图 1 所示尺度的气象基础模型。此外，在多尺度气象基础模型方面也存在可行性和高影响力的创新空间。

#### 3.1.1 物理尺度

考虑我们是否可以训练一个可以处理不同分辨率和时间间隔的多尺度模型。一些深度学习模型，如 GraphCast[13]，已经证明了在不同时间范围内进行预测的能力，但需要进一步的研究来探索更复杂的多尺度架构。

目前，大多数人工智能模拟器，包括盘古[15]、FourCastNet[33]和 GraphCast[13]等，主要在单一空间尺度上运行，通常倾向于用 ERA-5 易于处理其的大量训练数据。

尽管花费巨大，至少是训练成本的两倍，并且可能由于分辨率差异的 20 倍而更高，但单个基础模型应该能够用 MERRA2, ERA5 和 HRRR 作为微调来模拟多个数据集。这种能力意味着基础模型需要处理不同的分辨率（如 HRRR 所述）和时间步骤，并提供多区域覆盖。值得注意的是，HRRR 仅限于北美，而其他网站则覆盖全球。

这种情况需要一个关键的决策过程，人们需要选择特定的规模来集中努力。此外，虽然跨多个尺度建模的可能性不能完全排除，但它似乎只在一定的有限范围内可行。目前，为 ERA5 和 HRRR 部署独立监督模型的策略是必要的。然而，多分辨率和多物理场基础模型应该利用 HRRR 提供的高时空分辨率和 ERA5 数据的全球可用性。

总之，这一领域的创新将受益于多尺度和多物理场基础模型的发展，这种模型能够有效地近似多个 PDEs 的解，而不需要严格的时空边界条件，分辨率提高 25 倍（例如从 ERA5 到 HRRR）就意味着需要严格的时空边界条件。这种多尺度和多物理场基础模型可以超越现有先进模型的能力。

此外，与 ECMWF 的 IFS 相比，在多个下游任务（如热带气旋）上的推理速度有机会提高几

个数量级。

### 3.2 长期预测

基于人工智能的预测模拟器[15]和[13]表明，它们的预测在较长的时间范围内会变得有些模糊。评估人工智能天气模型的长期性能和稳定性对其实际适用性至关重要。长期预测的稳定性和对极端情况的更好的保真度是实时预测、天气和气候基础模型需要解决的关键问题。迄今为止训练的全天候人工智能模型的缺点是长时间范围的行为和捕捉极端情况的能力。

文献[80]中的证据表明，铺展不稳定性可能是由频率空间中次导模的误差引起的。这意味着可以通过扩散[80]或合适的傅里叶空间算子[33]来提高稳定性。或者，我们可以通过合适的损失函数或训练程序来解决这个问题。SFNO[33]和DLWPU-Net[81]通过对抗极奇点表现出自回归的1年预测无不稳定性。SFNO以牺牲空间局部性为代价使用球谐变换来减少伪影。同时，DLWPU-Net将数据表示从经纬度转换为等角标尺立方球，并使用CNN架构来保持空间局域性。

然而，没有人为因素的自回归预测与实现长期预测稳定性是不一样的，在长期预测稳定性中，模型不会简单地崩溃为气候学和模糊的极端。后者作为人工智能模型预测的长期稳定预测尚不能与物理模型竞争。

PDE-Refiner的论文[80]研究了一维和二维的多个pde问题。他们的方法使用细化步骤作为扩散模型的去噪目标，将焦点从仅高振幅转移到变化振幅。每个细化步骤添加高斯噪声，模型预测该噪声以对其输入进行去噪/细化。

未来的模型创新需要提高预测范围，且预测稳定性从7天扩展到更长的时间尺度。

### 3.3 数据

#### 3.3.1 观察、再分析、预测数据

人们需要决定是使用网格化的再分析产品还是使用观测数据。后者意味着模型需要能够处理来自自动气象站和卫星地球观测的稀疏和异质。大多数现有的人工智能天气预测模型都是基于网格分析的，不能直接利用观测数据。

观测数据直接测量了特定地点的大气变量，但覆盖范围可能很窄，特别是在偏远或海洋地区。网格数据产品，在覆盖整个地球的规则网格上提供插值和同化的数据，但可能受到同化过程中引入的插值误差和偏差的影响。目前，稀疏数据的合并还没有通过最先进的人工智能模型来解决，因为它将需要一个能够执行数据同化的基础建模系统。

#### 3.3.2 数据量

目前最先进的人工智能天气预测模型利用网格分析来确定参数和压力水平的固定子集，有时还包括时间步长。天气和气候领域每个时间步的相对数据量比大多数深度学习模型通常训练的数据量要高几个数量级。数据集的时间分辨率也有很大差异。CMIP6侧重于多年代际气候预测；而MERRA-2提供历年趋势监测的日尺度数据；ERA5和HRRR提供了更精细的时间分辨率，为每小时提供数据。

从CMIP6到HRRR，计算量和数据量逐渐增加。虽然CMIP6需要温和的计算资源，但MERRA-2和ERA5的需求会增加，需要更复杂的降尺度方法和额外的数据集成。向HRRR过渡，复杂性进一步加剧，需要高保真模型和大量的计算资源。此外，这些数据集的数据可访问性各不相同，数据管理和存储方面的挑战越来越大，特别是对于HRRR生成的高频数据。因此，从CMIP6向HRRR的过渡涉及到从宏观层面的气候建模到颗粒级实时天气预测的转变，以及越来越多的计算和数据管理挑战。

ERA5数据集由ECMWF在哥白尼气候变化服务项目下生成，是一个涵盖几十年天气和气候数据的综合存储库。它每小时更新一次，为分析历史天气模式和全球气候趋势提供了丰富的资

料来源。用户可以访问气象变量，如温度、降水、风速和风向、海平面压力和湿度。该数据集的格式通常为  $30 \times 30\text{km}$  的网格，但为了更深入地检查较小的特征，也可以访问更精细的分辨率数据 ( $10 \times 10\text{km}$ )。最新的先进机器学习天气预测模型经常使用该数据集。

高分辨率快速刷新 (HRRR) 数据集是由美国国家环境预测中心 (NCEP) 运行的北美地区最高分辨率预测模型生成的。HRRR 模型采用水平单元尺寸为  $3 \times 3\text{km}$  的网格，每小时更新一次以纳入最新观测。HRRR 使用 WRF 模型来同化各种观测，它的高时间更新频率允许观察和预测快速天气发展，使其适用于监测和预测恶劣天气事件。

### 3.4 诊断

为了从设计角度评估所提出的方法和模型，我们可以进一步寻求回答以下诊断问题：我们如何评估独立于下游任务的架构选择？是否有一套在预训练时进行评估的标准来表明良好的下游性能？这些诊断可以帮助将 AI 天气 FM 模型与 ECMWF 的 IFS 等物理模型、使用 ERA5 和 MERRA-2 等再分析数据集训练的 HRES 等操作模型以及 HRRR 等高分辨率区域模型进行比较。

#### 3.4.1 架构评估

1. 气象相关性：根据模型与气象过程的相关性评估其架构选择。验证模型是否有效地捕捉大气的垂直结构，处理复杂的天气模式，并编码重要的气象特征。

2. 网格表示：评估模型处理全球尺度气象数据的能力，以及适应气象模拟中常用的不同空间分辨率的能力。

3. 可解释特征：研究模型是否能够生成与已知气象现象（如气旋、锋面和大气稳定性）一致的可解释特征。

4. 球面与笛卡尔网格表示：使用机器学习的地球系统模型中网格表示的选择深刻影响模型的性能和解释，无论是球形、笛卡尔还是其他。球形网格与地球的几何形状对齐，提高了全球现象的准确性并确保更好的物理一致性，尤其是在两极。相比之下，笛卡尔网格计算简单，更适合区域模型，尽管它们可能在大规模应用中引入误差。物理原理，例如质量和能量守恒，也许在全球球形表示中得到更好的维护。机器学习模型可能需要基于所选网格的不同特征工程。由于数据表示差异，网格之间的迁移学习具有挑战性。其他网格，如六角形或二十面体 [33][15][13]，通过避免经纬度网格中存在的极奇点等问题来达到平衡。不同的网格可以更好地代表具体的现象，例如，大规模环流适合球形网格，而局部事件可能更适合笛卡尔网格。最优网格的选择取决于应用、规模和所需的模型属性。网格选择对于集成机器学习的地球系统模型的准确性、效率和一致性至关重要。

#### 3.4.2 预训练评估

1. 迁移学习：分析预训练模型如何将知识转移到下游气象应用。评估他们在各种气象预测任务中的表现，如温度、降水、风和压力预测。

2. 鲁棒性：检验模型对不同气候区域、极端天气事件和气象数据集中常见的数据异常值的稳健性。

3. 时间尺度依赖性：研究预训练模型如何处理短期和长期天气预测。评估它们捕捉每日变化和季节变化的能力。

4. 预测任务与通常前置任务：虽然天气预测是最终目标，但研究人员也探索了使用通用前置任务来预训练气象模型。前置任务，如对比学习和遮罩，有助于从未标记的数据中学习通用表示，这可以转移到下游预测任务中 [17], [20], [84], [85]。这些前置任务在提高天气预测能力方面的有效性与传统预测方法中进行广泛的评估和比较。

### 3.4.3 与物理基准比较

1. 准确性和性能:将建议模型的预测准确性与 ECMWFIFS 模型进行比较, IFS 模型是广泛使用的全球天气预测基准。评估他们在不同预测提前期表示气象变量的能力。

2. 计算效率:与 IFS 相比,评估拟议模型的计算效率,这对于实时天气预测应用至关重要。

3. 数据表示:验证所提出的模型是否能够有效地表示 MERRA-2 和 ERA5 数据集的独特特征,这些特征可能在数据质量、空间分辨率和时间覆盖方面存在差异。

4. 泛化:评估模型跨不同数据集和气候区域的泛化能力,主要针对观测数据有限的地区。

5. 跨任意长度和时间尺度的信息流:在天气预测中,跨越任意长度和时间尺度的信息流对于捕捉大气的全部动态至关重要。

这些评估和比较使我们能够从气象和气候科学的角度确定最合适的预训练基础人工智能模型的策略。所选择的策略应有效和稳健地编码地球大气中发生的内在行为和过程(大气动力学),以及用于在天气模型中模拟这些行为和过程的计算方法和算法(模型动力学)。这在理想中可以增强天气预报能力,通过促进各种尺度的更广泛推广,有可能超越 IFS 等传统的 NWP 方法。

## 3.5 影响、作用

### 3.5.1 机器学习架构的关键特征

总之,根据上述诊断,我们建议天气基础模型在后续第 4 节中需要考虑以下机器学习架构实现的关键特征:

1. 多尺度建模
2. 长期预测的稳定性
3. 增加价值的能力(降低微调的复杂性和所需资源)
4. 考虑稀疏输入约束的数据可用性

在前面的章节中,我们建议社区关注与天气基础模型相关的空间和时间尺度;完整的地球系统基础模型超出了范围。除了图 1 中建议的空间尺度之外,使用短期预测方法处理几秒到不到 1 小时之间的微尺度状况将需要付出巨大的努力。微尺度可能会破坏所提出的天气基础模型的稳定性,从而影响其他时空条件中的下游任务。

除了超分辨率微尺度之外,短期预测所需的另一项创新是将稀疏观测数据同化纳入基础模型。虽然这是可行的,但在获得构建短期预测天气基础模型的经验后,可能需要在专注于该主题的第一年之外进行更多的项目工作。因此,在接下来的部分中,我们考虑小于 1 小时且小于 1 公里的分辨率的短期天气预测,我们对具有高影响力的创新更有信心。但对于超过 14 天的中期天气预测和次季节预测,机器学习和基于物理的模型尚未表现出比气候学更好的结果。这些情景是否可以产生重大创新还具有高度不确定性。

#### 天气基础模型之外的未来创新机会

在考虑扩展到多年、长期预测的范围时,两个基本方面受到关注:气候预测和气候预测。前者涉及利用 IPCC 的 CMIP6 数据集等广泛资源对全球气候振荡(包括 ENSO)等自然现象进行建模和预测。

另一方面,气候预测更深入,涉及考虑人类应对气候危机的潜在情景。这就需要开发能够吸收特定参数(例如全球二氧化碳浓度)的变化如何影响气候随时间演变的机器学习模型。

然而,在预测包括 ENSO 在内的全球气候振荡等复杂现象以及了解全球碳循环和范围 1-3 排放动态变化的不确定性方面仍然存在重大障碍。人类对气候危机的反应以及相关的反馈和临界点的不确定性带来了额外的复杂性。尽管面临挑战,但仍有广阔的创新空间,气候科学家将继续解决这些复杂的问题,目标是促进更准确、更全面的气候预测和预测。

这个范围的另一个需要改进的领域是，人们只能往回测试预测结果，并且需要等待几年才能在数据可用时确保其影响。然而，这可能不会成为采用气候变化模型进行充分回测的情景分析工具的限制。这就给了一个机会研究加速模型执行数百年的模拟并提高其多尺度保真度。

## 4 实现

本节讨论实现天气基础模型的具体需求，并比较在模型中使用不同组件(如 Transformer 块、数据表示和神经操作符)之间的权衡。

如 3.5.1 节所述，在开发天气模型时，关注两个关键目标是至关重要的：多尺度稳健性和长期稳定性。第一个特性使模型能够有效地处理可能在距离和时间上发生变化的各种数据源。同时，长期的稳定性保证了模型的可靠性，即使预测时间超过了设计初期的考虑范围。最后，在处理涉及天气和气候的大量数据时，需要解决代码中可能影响计算性能的任何差距。

### 4.1 数据表示

数据表示是第一步，是设计任何机器学习解决方案的关键组成部分。虽然天气和气候数据可能与图像和视频数据有一些相似之处，但由于前者的独特性，利用现成的系统可能是不够的。本节讨论有关天气和气候数据的一些重要注意事项。

#### 4.1.1 坐标表示

利用笛卡尔网格来表示天气和气候数据的原因很明显。这些数据与互联网上丰富的 2D/3D 图像和视频具有共同的特征，它们的应用也更加流行。当我们查看有关将机器学习应用于天气和气候数据的文献时，它们的基础通常来自 U-Net[37]、ResNet[86]、ViT[8]和 VideoMAE[87]等工作，这些工作最初是用于图像和视频的。

通常，这些模型使用均匀的矩形网格模型的气候数据。一些例子包括 Climax[20]、Pangu[15]、SwinRDM[16]、DYffusion[75]、FourCastNet[14]和 MetNet-2[30]。虽然这种方法可能适用于局部和区域建模属性，但它无法捕捉地球的物理和测地线属性，特别是当焦点从赤道转移时[88]。

在另一个方向上，一些工作探索了球面表示，如等矩形(ERP)和立方体[89]，二十面体网格[13]，[88]，[89]和分层等面积等纬度像素化(HEALPix)[90]。在[33]中，作者讨论了球面傅里叶神经算子的一个新版本。他们引入了球面卷积的概念使傅里叶层线性和等变。heali-Swin[79]将球体的 HEALPix 像素化与 Swin Transformer 的分层结构相结合。HEALPix 特别适合这种方法，因为它定义了嵌套像素。不幸的是，目前似乎还没有争论，当从笛卡尔网格移动到球面表示时，额外的工程工作的影响是什么。需要通过消融研究做出决定。

### 4.2 模型组件

在接下来的章节中，我们将讨论天气和气候基础模型的主要候选组件。例如 Transformer，图神经网络，和不同形式的神经算子。大多数关于大型人工智能天气模型的现有出版物都集中在其中一个方面。然而，原则上，它们也可以结合在一起。

#### 4.2.1 Transformers

为了克服固定长度编码向量的瓶颈问题，作者在[91]中引入了注意机制。这种方法对于更长的或更复杂的序列尤其重要，因为它们的表示将被限制在与更短或更简单的序列相同的维度上。作为 Bahdanau 工作的扩展，Transformer 模型用包含查询-键-值(Query-Key-Value, QKV)机制的自注意机制取代了原始注意力机制。等式 2 给出了自注意力机制的矩阵表示，对于查询  $Q \in \mathbb{R}^{N \times D_k}$ 、键  $K \in \mathbb{R}^{M \times D_k}$  和值  $V \in \mathbb{R}^{M \times D_v}$ ，

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)V = AV, (2)$$

其中  $N$  代表查询的长度,  $M$  代表键(或值)的长度。 $D_k$ 和 $D_v$ 分别表示键(或查询)和值的维度。 $A = \text{softmax}\left(\frac{QK^T}{\sqrt{D_k}}\right)$ 为注意矩阵, 按行使用  $\text{softmax}$  激活函数。查询和键的点积除以 $\sqrt{D_k}$ 来解决梯度消失的问题。Transformer 使用多头注意功能而不是单个注意功能。具有 $D_m$ 维的原始查询、键和值通过  $H$  组不同的学习投影投影到 $D_k$ 、 $D_k$ 和 $D_v$ 维。

目前自然语言、图像、视频和时间序列处理的基础模型依赖于 Transformer, 并代表了每个领域的最新技术 (AI Index, 2023)。此外, 正如前面提到的, Transformer 已经在天气预测方面显示出了令人鼓舞的结果。因此, 基于 Transformer 创建天气和气候基础模型的可能性是一个令人兴奋的前景。

**Token 化策略** Transformer 使用离散块中的数据, 称为 Token。将原始输入数据(如文本、图像或时间序列)转换为 Token 称为 Token 化。对于语言来说, 这个过程相对简单, 因为单词是自然的信息单位。但对于其他类型的数据没有自然的定义, 不同的选择严重影响模型的性能。

在将输入数据分解为 Token 之后, 有必要将它们嵌入到度量空间中。正如我们所讨论的, 天气和气候变量与 3D 视频有相似之处。然而, 大气资料往往具有更多具有不同相关性的变量(通道)。例如, 与风和地势相比, 温度和湿度的相互作用不同。在 Pangu[15]中, 变量被编码为三维卷积中的通道, 这使得不同含义的变量共享相同的滤波器。

另一方面, ClimaX[20]对每个变量进行了单独的修补, 但这增加了模型的复杂性。当然, 盘古的选择降低了模型的复杂性, 使其训练和发球更快。然而, 这样的选择限制了模型成为基础模型, 因为下游任务可能没有在训练期间使用的所有变量, 这将使模型失败。

气候和天气的基础模型需要能够灵活地处理多种输入配置, 包括不同的变量集或分辨率, 但不能过多地增加复杂性。此外, 当存在稀疏输入、数据是 2D 或 3D 或分辨率变化时, Token 化可能需要具有鲁棒性。如果 Token 化方案突出一个好的词汇表(“气候词”), 它可以更好地解释模型输出, 这可能是大气变量的可识别模式。最后, Token 化可以允许模型学习空间和时间上的远程连接, 例如远程连接。

**位置编码** Token 化之后, 最好向每个 Token 添加一个信号, 告知模型该 Token 的来源。这被称为位置编码。这个信号可以表示绝对位置或相对位置, 但它也可以区分不同的物理参数或时间步长。接下来, 我们将讨论三个关键属性。

第一, 位置 Token 有助于直接告知模型剩余 Token 的内容。通过位置 Token, 我们可以将信息单位告知模型(例如, “嵌入的 Token 指的是 100hPa 的湿度”), 或者使用位置 Token 来传递地形信息(例如, “当前感兴趣的区域在陆地、海洋、城市等上方”)。特别是在多区域、多模式设置中, 位置 Token 可以帮助我们告知模型嵌入的输入 Token 的差异。

第二, 位置 Token 可以硬编码(位置编码)或学习(位置嵌入)。对于位置编码, 我们区分绝对编码、相对编码和条件编码。绝对位置编码与输入 Token 的嵌入具有相同的维度, 并确定地引用数据中的特定索引, 例如纬度。相对位置编码考虑 Token 之间的距离, 条件位置编码是动态生成的, 依赖于输入 Token 的周围区域。位置嵌入提供了另一种表示位置的方法: 为每个位置学习一组位置 Token[1], [92]。学习嵌入提供了比编码位置表示更大的灵活性, 因为它们可以通过反向传播适应特定的任务。

第三, 我们可以利用位置嵌入在推理过程中跨压力水平进行泛化。在训练期间, 可以获得一些特定压力水平的数据。当将位置嵌入作为压力水平的函数时, 我们可以在训练中看不到的推理中进行不同层次的推广。

总的来说, 位置 Token 提供编码或嵌入附加信息, 处理任意序列长度, 并在推理期间跨压力级别进行泛化。与基于图的方法相比, 这些属性在许多天气建模场景中都是有益的, 并且是

特定于 Transformer 体系结构的。最近用于天气预测的 Transformer 为例如，利用地球表面可学习的绝对位置符号嵌入变量信息和值的位置信息[15]。

**序列长度** 处理任意序列长度是气象建模的重要前提。处理比在训练时间中看到的序列更长的序列(即“归纳”方法)是特别相关的。例如，当训练 Transformer 在最高分辨率(即  $0.25^\circ \times 0.25^\circ$ )的 ERA5 再分析数据上进行每小时预测时，我们的 2D 输入分辨率为  $1440 \times 721$ ，另外还有 37 个压力级别用于第三维度。

当我们考虑湿度、温度、位势、东风、北风和垂直风速时，我们在每个时间步长总共得到  $1440 \times 721 \times 37 \times 6 = 230489280$  个数据点(见上文，c. f. [15])。对于模型训练来说，序列长度通常相对较短，以保持相关的计算工作可管理(例如，在计算机视觉或自然语言处理中，每个序列的数十万到数百万个 Token)。这指的是大约  $100 \times 100$  的二维空间分辨率(例如， $100 \times 100 \times 37 \times 6 = 2220000$ )。然而，对于长期预测或全球预测，模型需要能够消化更多的数据。更长的序列。有可能通过位置嵌入(即，学习到的位置 Token)和相对和条件位置编码(即，手工制作的位置 Token)实现这样的能力。

**稀疏注意** 前面描述的注意机制允许每个查询节点从所有内存节点收集信息，创建一个完全连接的二部图。虽然这种方法可以改善表示，但公式(2)中矩阵 A 的大小可能会变得很大，这取决于输入，并显著影响复杂性和内存使用。相反，稀疏注意模型消除了节点  $A_{ij} = q_i k_j^T$  之间的特定连接，产生更少的内存使用和更低的复杂性。此外，在此过程中可能会增加结构性偏差。

大气变量的表示发生在三维空间中。流体运动的有限速度意味着空间中的一个点影响另一个点需要时间。因此，模型不需要在任何时候都关注所有的位置，从而减少了  $k_i$  和  $q_j$  参加的数量。该模型如何减少资源使用的一个例子是在提高学习能力的同时利用带状注意力[93]或扩张性注意力[94]。此外，根据输入信号的冗余程度，可以忽略部分输入空间并应用随机稀疏注意，类似于 BigBird[95]。

#### 4.2.2 图

在 4.1 节中，我们探讨了 Transformer 模型中数据的典型表示，它类似于密集张量，例如图像。此外，我们注意到典型的注意机制可以被视为一个完全连接的二部图。在本小节中，我们将讨论直接使用图形结构表示天气数据和不同像素和区域之间的信息流。

图神经网络(GNN)已成为天气预测的强大工具。图卷积网络(GCNs)、图注意网络(GATs)、图自动编码器(GAEs)和图扩散模型已被证明在学习表征、预测气象要素的空间分布、关注特定节点和推导潜在表征方面是有效的。正在进行的研究，如[13]、[96]–[98]表明，基于图表的方法在天气预测中有更创新和有效的应用。此外，最近对具有消息传递的新架构的 GNN 的研究表明，在求解网格上的偏微分方程方面取得了很好的结果[99]，[100]。本文提出了一种编码器-处理器-解码器架构，其中编码器将现有网格或网格编码成多图，处理器使用 N 个相同的消息传递块，将 GraphNet 块推广到多个边缘集，然后解码器从输入时间步长  $t$  预测状态  $t+1$  并将转换后的网格解码到原始空间。

请注意，天气基础模型的一个关键特征是对不同输入配置的灵活性，如 4.1 节所述。像在一个卷积中编码多个变量的方法一样，GNN 编码器可能需要对不同的输入集进行重新训练，从而影响其运行多个下游任务的能力。

#### 4.2.3 神经算子

在过去的几年中，科学家和工程师一直在努力解决复杂的计算问题，并且在求解偏微分方程(PDEs)的传统方法中需要更多的数据特异性，这些方法描述了不同领域的各种现象。标准数值方法，如有限差分和有限元技术，以计算量大而闻名，特别是在高维情况下，寻找有效的解



决方案具有挑战性。

数据驱动的方法，如 DeepONets[101]和傅立叶神经算子[102]有望解决偏微分方程，因为它们可以直接从数据中学习算子。它们能有效地处理复杂的算子学习任务，并具有良好的泛化能力。在天气预测方面，有两种模型特别引人注目，分别是 FourCastNet 和 SFNO。FourCastNet 将自适应傅立叶神经算子 (AFNO) 模型与 Transformer 相结合，而 SFNO 则将 FNO 推广到球面上，以学习球面几何上的算子。特别是，FourCastNet 模型利用 AFNO 的注意机制与 ViT 主干进行有效的 Token 混合。SFNO 基于广义黎曼流形，在这种情况下是一个球。

在天气预测中，有一种特殊的方法似乎更有效、更稳健，那就是卷积神经算子 (CNO) [78]。第 4.2.1 节讨论了使用 Transformer 主干时 Token 化的基本步骤。对于图像，卷积层是实现这一目标的首选。然而，CNN 通常具有有限维的输入输出映射，并且在直接应用时显示出对网格分辨率的依赖性到 PDE 问题。通过引入卷积神经算子 (CNOs)，作者克服了这些障碍，创建了一个能够学习算子的系统，该系统可以将输入函数映射到输出函数。

CNOs 不受网格分辨率的严格限制，这是其他网络的一个重要瓶颈。通过将输入函数直接映射到输出函数，它们可以处理不同的数据集，而不需要特定的网格依赖关系。这样的特性使它们更具适应性和灵活性，特别是在天气数据方面。此外，以其并行化和注意力机制而闻名的 Transformer 可以与 CNOs 无缝集成，从而创建利用两种体系结构优势的协同作用。与 DeepONets 和 FourCastNet 等其他网络的集成相比，这种集成可以使模型能够以更高的技能和准确性处理天气预测中普遍存在的高维数据。

### 4.3 预训练机制和损失函数

在 4.1 节和 4.2 节分别讨论了数据表示和模型主干的可能选择之后，现在是时候转向实现基础模型时的最后一个组成部分：预训练机制的选择。

#### 4.3.1 rmse 型损失：重建与预测

在本文中，我们多次提到网格化天气和气候数据与图像或视频数据的相似性。在预训练的背景下，目前的标准方法是遮罩 [85]。当训练一个视觉 Transformer 时，移除一个随机的 Token 子集。其余的 Token 是编码的。解码器被给予潜在空间表示以及关于缺失 Token 的信息。接下来的任务是重建整个图像。通常这是用 RMSE 损失来训练的。

对于天气和气候数据，人们将各种代理任务重新表述为遮罩主题的许多变体的变体。例如，[20]的作者使用预测作为代理任务。也就是说，给定时刻  $t$  的某个输入，要求模型在时刻  $t + \Delta t$  产生相同的大气场。解码器没有被告知缺失贴图的空间位置，而是得到一个反映  $\Delta t$  的时间位置编码。虽然没有在文献中实现，但可以自然地使用类似的模式从输入中删除整个字段或压力级别并重建它们。我们可以把所有这些情况——甚至是预测的情况——看作是通过优化 RMSE 损失来重建缺失的信息。

最近发布的基于 Transformer 的 AtmoRep [18] 也使用遮罩。然而，重建是不确定的，或者通过简单的逐像素 RMSE 损失来训练。相反，AtmoRep 生成了一个预测集合，允许进行概率处理。损失函数是自定义的，涉及预测集合的前两个统计矩。

#### 4.3.2 替代方法

**对比学习** 上述重建任务的关键特性是损失是通过逐像素比较计算的。在某些情况下，这可能不是理想的。其中一个例子就是降水或降雪等稀疏区域。对于像素级损失，一个公认的替代方法是对比学习。给定数据输入  $x$ 、模型  $f$  和一组随机变换  $T$ ，[103] 的 SimCLR 方法是创建两个增广  $x' = T(x)$  和  $x'' = T(x)$ 。然后训练模型  $f$ ，使  $f(x')$  和  $f(x'')$  在潜在空间中接近，但远离其他数据。这是通过优化“归一化温标交叉熵” (NT-Xent) 来实现的为了确保来自不同数据的嵌

入在潜在空间中被很好地分离，NT-Xent 在训练模型时涉及当前批次的所有样本。在[103]的情况下，这会导致巨大的内存和资源需求，因为批大小可以高达 4096。然而，在原始论文的基础上有一些改进，如[104]，设法规避了这个问题。

对比方法的一个关键方面是变换  $t$  的设计和选择。常见的选择有裁剪、模糊、遮罩、颜色抖动和索贝尔滤镜。有人努力定义特定于大气动力学的自定义转换。然而，[19]表明，至少对于天气和气候，不需要复杂或定制的代理任务。相反，使用子抽样作为唯一的增强是完全足够的：给定一个数据立方体  $x$ ，它已经在 4.2.1 节中进行了 Token 化，只需对 Token 的一小分子集进行采样，并丢弃其余部分。这让人想起遮罩，但它绕过了将整个多维数据集保存在内存中的需要。最近另一种不依赖于手工制作的数据增强的方法是基于图像的关节嵌入预测架构 (I-JEPA) [84]。

**特定领域的代理任务** 虽然预训练是任务独立的，但它不一定是领域独立的。换句话说，虽然到目前为止讨论的训练方式最初是在计算机视觉中开发的，但我们现在将转向特定于天气和气候数据的方法。

一个典型的例子是 AtmoDist[17]。作者提出了一种新的自监督学习策略，用于大气动力学中的表示学习，该策略基于地球物理流体动力学原理中特定领域的代理任务。这项任务要求模型预测时间间隔，时间间隔的定义是在给定的时间序列中两个相邻的大气状态之间的时间步数。使用分类交叉熵损失函数来评估学习的有效性，该函数测量预测和实际时间间隔之间的差异。此外，本文还提出了其他潜在的领域特定的代理任务表征学习在这一领域，包括预测空间分离，时间演变，国家间的关系，以及外部力量对大气状态的影响。

### 4.3.3 扩散

一个标准的概率扩散机制包括一个正向去噪过程和一个反向去噪过程，以了解数据的分布。在天气预测方面，[75][105]将扩散用于概率预测，[75][80]将扩散用于长期推广。

如[105][75]所示，扩散模型具有生成概率预测的能力。Dyffusion[75]使用预测目标代替标准去噪目标。内插器利用蒙特卡罗 Dropout 在反向过程中生成概率预测来创建随机输出。SEEDS[105]使用 128 个经过训练的集成模拟器来生成类似于推理过程中提供的“种子”（由两个不同的集成模型生成的事件的两个预测）天气状态的概率预测。

自回归模型中的误差传播使得长期部署不稳定和不准确。为了提高长期预测水平，Dyffusion[75]使用了清洁和噪声初始条件。PDE-Refiner[80]使用扩散过程来改进预测，通过迭代地查看它们来捕获数据中经常被忽略的低幅度信息。

在[75][106]中，扩散激发了时间条件插值，以利用数据中的时间动态。DiTTO[106]仅使用扩散模型的正演过程，在初始和最终时间步之间创建了连续插值。它利用时间进化代替了前向过程中的增量噪声。在 Dyffusion[75]中也使用了类似的思想，用于前向过程，为中间动态步骤创建一个时间条件插值器。

当使用扩散模型时，我们需要考虑：

1. 特定于任务的噪声策略。例如，[80]使用了一种指数递减的噪声调度程序，其最小噪声方差比常规扩散模型下降得更快，以满足确定性解决方案的低误差标准。基于经验证据[75]使用余弦（线性）调度器进行海温预测，从扩散映射到插值时间步长

2. 用于去噪目标的评分函数。文献中使用的评分函数有 ViT[105]、U-Net[80][75]、CNN[75]、UNet-Transformer[106]。对于我们的模型，骨干网络的选择将决定这个函数。

虽然扩散模型具有上述优点，但往往需要在计算资源和预测精度之间进行权衡。

## 5 总结

在大型语言模型领域的取得巨大进展的推动下，以及相关聊天机器人在学术界之外的影响

的催化下, 去年我们看到了 NLP 领域之外的基础模型范式的采用。与此同时, 利用大型监督人工智能模型解决数值天气预测的努力已经达到了引起气象界关注的程度[13]–[16]。这两种趋势的几乎合乎逻辑的交集是天气和气候基础模型的发展, 始于[18], [20]。正如我们在 2.9 节中所讨论的, 基础模型为天气和气候带来了一系列好处。其中最值得注意的是, 与传统方法相比, 效率提高了, 同时也减少了对标签的依赖。

尽管该方法取得了显著的成功并且前景广阔, 但开发仍处于起步阶段。对于作为这种模型基础的设计原则, 没有明确建立或主导的范例, 也没有对其最突出的应用及其影响的清晰理解。事实上, 虽然[20]和[18]都是基于 Transformer 的架构, 但[13]–[16]的预测仿真器利用了 Transformer、图神经网络和神经算子。考虑到基础模型训练所涉及的相当大的计算成本, 人们不能仅仅基于实验来做出架构选择。考虑到这一点, 我们的工作旨在为任何对利用、设计或实现天气和气候领域的基础模型感兴趣的人提供一些指导。

总的来说, 我们的方法是直接的, 但在方法上是结构化的: 从可能的应用开始(参见第 2 节), 我们的目标是确定基础模型设计的差距和范围。在固定的范围内, 气象问题确实对基础模型设计提出了一定的要求(第 3 节)。最后, 可以使用这些约束来选择合适的构建块、数据集、预训练程序和预训练程序以及微调策略(第 4 节)。

沿着这条道路, 我们提出了几项建议。首先, 应该清楚的是, 当前的基础模型会在空间和时间尺度方面受到限制。话虽如此, 考虑到我们希望捕获尽可能多的下游应用, 单个模型覆盖空间和时间尺度的范围是有益的。在具体应用方面, 预测和缩减规模几乎是它们接近主要的预训练范式而出现的; 然而, 许多其他应用可以很容易地纳入范围。

然而, 尺度的选择不仅受期望的下游任务驱动, 还受可用的预训练数据驱动。在这里, 应该清楚的是, 更长的时间序列将捕捉到更丰富和更广泛的气象现象。实现这一目标的一个直接方法是使用 80 年的可用 ERA5 数据进行训练。然而, 这将模型的规模固定在 ERA5 的规模上。另一种方法是利用雷达观测的短期预测模型或 CMIP6 或类似的面向气候尺度的东西。以 ERA5 数据为基础, 一个模型可能会一直延伸到 km 尺度区域预测模型, 如 RDPS 或 HRRR。

前面的考虑指向一个人工智能系统, 该系统能够使用不同分辨率的网格 NWP 输出, 并可能使用不同的参数定义作为输入。此外, 同时使用 ERA5 和 HRRR 意味着还要考虑到不同的空间覆盖范围。如果做出了这些选择, 那么实现选择就会产生直接的后果。首先, 关于网格和空间覆盖的灵活性强烈建议使用 Transformer 而不是图。虽然有人可能会说, 在 Token 化过程中将空间信息分解成块有点不自然, 但这个过程产生了相当大的灵活性, 因为可以通过位置编码定义每个 Token 的空间覆盖范围、位置和分辨率的具体含义。与此同时, 这意味着要制定明确的计划来处理用于注意力计算的序列长度。我们的讨论概述了可能的选择。

最后, 虽然预测远不是唯一感兴趣的下游任务, 但它可能是具有最明确确定的基准和明确影响的任务。然而, 一旦可以进行预测, 人们就不得不处理预测的稳定性问题, 以及对极端情况的保真度问题。我们的讨论再次强调了几个选择。

## 参考文献

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.
- [2] T. B. Brown, B. Mann, N. Ryder, et al., “Language models are few-shot learners,” arXiv, p. 2005.14165v4, 2020.
- [3] A. Radford, J. W. Kim, C. Hallacy, et al., “Learning transferable visual models from natural language supervision,” in International conference on machine

- learning, PMLR, 2021, pp. 8748 – 8763.
- [4] R. Bommasani, D. A. Hudson, E. Adeli, et al., “On the opportunities and risks of foundation models,” arXiv preprint arXiv:2108.07258, 2021.
  - [5] P. Bauer, A. Thorpe, and G. Brunet, “The quiet revolution of numerical weather prediction,” *Nature*, vol. 525, no. 7567, pp. 47 – 55, 2015.
  - [6] H. R. Glahn and D. A. Lowry, “The use of model output statistics (mos) in objective weather forecasting,” *Journal of Applied Meteorology and Climatology*, vol. 11, no. 8, pp. 1203 – 1211, 1972.
  - [7] ECMWF, Key facts and figures | ECMWF, Sep. 2023. [Online]. Available: <https://www.ecmwf.int/en/about/media-centre/key-factsand-figures> (visited on 09/09/2023).
  - [8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.
  - [9] Z. Liu, Y. Lin, Y. Gao, et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012 – 10 022.
  - [10] Z. Liu, J. Ning, Y. Cao, et al., “Video swin transformer,” arXiv, p. 2106.13230v1, 2021.
  - [11] D. Hendrycks, X. Liu, E. Wallace, A. Dziedzic, R. Krishnan, and D. Song, “Pretrained transformers improve out-of-distribution robustness,” arXiv preprint arXiv:2004.06100, 2020.
  - [12] C. Zhang, M. Zhang, S. Zhang, et al., “Delving deep into the generalization of vision transformers under distribution shifts,” in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2022, pp. 7277 – 7286.
  - [13] R. Lam, A. Sanchez-Gonzalez, M. Willson, et al., “Graphcast: Learning skillful medium-range global weather forecasting,” arXiv preprint arXiv:2212.12794, 2022.
  - [14] J. Pathak, S. Subramanian, P. Harrington, et al., “Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators,” arXiv preprint arXiv:2202.11214, 2022.
  - [15] K. Bi, L. Xie, H. Zhang, X. Chen, X. Gu, and Q. Tian, “Accurate medium-range global weather forecasting with 3d neural networks,” *Nature*, pp. 1 – 6, 2023.
  - [16] L. Chen, F. Du, Y. Hu, Z. Wang, and F. Wang, “Swinrdm: Integrate swinrnn with diffusion model towards high-resolution and high-quality weather forecasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 2023, pp. 322 – 330.
  - [17] S. Hoffmann and C. Lessig, “Atmodist: Self-supervised representation learning for atmospheric dynamics,” *Environmental Data Science*, vol. 2, e6, 2023.
  - [18] C. Lessig, I. Luise, B. Gong, M. Langguth, S. Stadler, and M. Schultz, “Atmorep: A stochastic model of atmosphere dynamics using large scale representation learning,” arXiv preprint arXiv:2308.13280, 2023.
  - [19] J. Nathaniel, M. Freitag, P. Curran, I. Ruddick, and J. Schmude, “Title here,” To appear, 2023.
  - [20] T. Nguyen, J. Brandstetter, A. Kapoor, J. K. Gupta, and A. Grover, “Climax: A

- foundation model for weather and climate,” arXiv preprint arXiv:2301.10343, 2023.
- [21] D. C. Dowell, C. R. Alexander, E. P. James, et al., “The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. part i: Motivation and system description,” *Weather and Forecasting*, vol. 37, no. 8, pp. 1371 – 1395, 2022.
  - [22] J. A. Milbrandt, S. B  elair, M. Faucher, M. Vall  e, M. L. Carrera, and A. Glazer, “The pan-canadian high resolution (2.5 km) deterministic prediction system,” *Weather and Forecasting*, vol. 31, no. 6, pp. 1791 – 1816, 2016.
  - [23] S. Saha, S. Moorthi, X. Wu, et al., “The ncep climate forecast system version 2,” *Journal of climate*, vol. 27, no. 6, pp. 2185 – 2208, 2014.
  - [24] U. Germann and I. Zawadzki, “Scale-dependence of the predictability of precipitation from continental radar images. part i: Description of the methodology,” *Monthly Weather Review*, vol. 130, no. 12, pp. 2859 – 2873, 2002.
  - [25] G. Ayzel, M. Heistermann, and T. Winterrath, “Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0. 1),” *Geoscientific Model Development*, vol. 12, no. 4, pp. 1387 – 1402, 2019.
  - [26] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, “Machine learning for precipitation nowcasting from radar images,” arXiv preprint arXiv:1912.12132, 2019.
  - [27] C. K. S  nderby, L. Espeholt, J. Heek, et al., “Metnet: A neural weather model for precipitation forecasting,” arXiv preprint arXiv:2003.12140, 2020.
  - [28] R. Prudden, S. Adams, D. Kangin, et al., “A review of radar-based nowcasting of precipitation and applicable machine learning techniques,” arXiv preprint arXiv:2005.04988, 2020.
  - [29] S. Ravuri, K. Lenc, M. Willson, et al., “Skilful precipitation nowcasting using deep generative models of radar,” *Nature*, vol. 597, no. 7878, pp. 672 – 677, Sep. 2021, Number: 7878 Publisher: Nature Publishing Group, issn: 1476-4687. doi: 10.1038/s41586-021-03854-z. [Online]. Available: <https://www.nature.com/articles/s41586-021-03854-z> (visited on 08/16/2023).
  - [30] L. Espeholt, S. Agrawal, C. S  nderby, et al., “Deep learning for twelve hour precipitation forecasts,” *Nature communications*, vol. 13, no. 1, pp. 1 – 10, 2022.
  - [31] Y. Zhang, M. Long, K. Chen, et al., “Skilful nowcasting of extreme precipitation with NowcastNet,” *Nature*, vol. 619, no. 7970, pp. 526 – 532, Jul. 2023, issn: 1476-4687. doi: 10.1038/s41586-023-06184-4. [Online]. Available: <https://doi.org/10.1038/s41586-023-06184-4>.
  - [32] M. Andrychowicz, L. Espeholt, D. Li, et al., “Deep learning for day forecasts from sparse observations,” arXiv preprint arXiv:2306.06079, 2023.
  - [33] B. Bonev, T. Kurth, C. Hundt, et al., “Spherical fourier neural operators: Learning stable dynamics on the sphere,” arXiv preprint arXiv:2306.03838, 2023.
  - [34] H. Hersbach, B. Bell, P. Berrisford, et al., “The era5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999 – 2049, 2020.
  - [35] S. Lu, Y. Hwang, I. Khabibrakhmanov, et al., “Machine learning based multi-physical-model blending for enhancing renewable energy forecastimprovement

- via situation dependent error correction,” in 2015 European control conference (ECC), IEEE, 2015, pp. 283 – 290.
- [36] P. Harder, Q. Yang, V. Ramesh, et al., “Generating physically-consistent high-resolution climate data with hard-constrained neural networks,” arXiv preprint arXiv:2208.05424, 2022.
- [37] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234 – 241.
- [38] K. Stengel, A. Glaws, D. Hettinger, and R. N. King, “Adversarial superresolution of climatological wind and solar data,” Proceedings of the National Academy of Sciences, vol. 117, no. 29, pp. 16 805 – 16 815, 2020.
- [39] J. Wang, Z. Liu, I. Foster, W. Chang, R. Kettimuthu, and V. R. Kotamarthi, “Fast and accurate learned multiresolution dynamical downscaling for precipitation,” Geoscientific Model Development, vol. 14, no. 10, pp. 6355 – 6372, 2021.
- [40] B. Groenke, L. Madaus, and C. Monteleoni, “Climalign: Unsupervised statistical downscaling of climate variables via normalizing flows,” in Proceedings of the 10th International Conference on Climate Informatics, 2020, pp. 60 – 66.
- [41] Q. Yang, A. Hernandez-Garcia, P. Harder, et al., “Fourier neural operators for arbitrary resolution climate data downscaling,” arXiv preprint arXiv:2305.14452, 2023.
- [42] R. Pielke Sr., T. Matsui, G. Leoncini, et al., “A new paradigm for parameterizations in numerical weather prediction and other atmospheric models,” National Weather Digest, vol. 30, pp. 93 – 99, 2006. [Online]. Available: INSERT\_URL\_HERE\_IF\_AVAILABLE.
- [43] T. T. Warner, Numerical Weather and Climate Prediction. Cambridge University Press, 2010. doi: 10.1017/CB09780511763243.
- [44] S. Rasp, M. S. Pritchard, and P. Gentine, “Deep learning to represent subgrid processes in climate models,” Proceedings of the National Academy of Sciences, vol. 115, no. 39, pp. 9684 – 9689, Sep. 2018, Publisher: Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1810286115. [Online]. Available: <https://www.pnas.org/doi/10.1073/pnas.1810286115> (visited on 08/14/2023).
- [45] S. Yu, W. M. Hannah, L. Peng, et al., “ClimSim: An open large-scale dataset for training high-resolution physics emulators in hybrid multiscale climate simulators,” arXiv, 2023. doi: 10.48550/arxiv.2306.08754. eprint: 2306.08754.
- [46] J. Yuval and P. A. O’ Gorman, “Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions,” Nature communications, vol. 11, no. 1, p. 3295, 2020.
- [47] J. Brajard, A. Garrassi, M. Bocquet, and L. Bertino, “Combining data assimilation and machine learning to infer unresolved scale parametrization,” Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, p. 20 200 086, 2021.
- [48] B. L’utjens, C. Crawford, C. Watson, C. Hill, and D. Newman, “Multiscale neural operator: Learning fast and grid-independent pde solvers,” arXiv preprint

- arXiv:2207.11417, 2022.
- [49] J. Mack, R. Arcucci, M. Molina-Solana, and Y.-K. Guo, "Attention-based convolutional autoencoders for 3d-variational data assimilation," *Computer Methods in Applied Mechanics and Engineering*, vol. 372, p. 113 291, 2020.
  - [50] R. Arcucci, J. Zhu, S. Hu, and Y.-K. Guo, "Deep data assimilation: Integrating deep learning with data assimilation," *Applied Sciences*, vol. 11, no. 3, p. 1114, 2021.
  - [51] A. Farchi, M. Chrust, M. Bocquet, P. Laloyaux, and M. Bonavita, "Online model error correction with neural networks in the incremental 4dvar framework," *arXiv preprint arXiv:2210.13817*, 2022.
  - [52] A. J. Dowdy and J. L. Catto, "Extreme weather caused by concurrent cyclone, front and thunderstorm occurrences," *en, Scientific Reports*, vol. 7, no. 1, p. 40 359, Jan. 2017, Number: 1 Publisher: Nature Publishing Group, issn: 2045-2322. doi: 10.1038/srep40359. [Online]. Available: <https://www.nature.com/articles/srep40359> (visited on 08/16/2023).
  - [53] J. L. Catto and A. Dowdy, "Understanding compound hazards from a weather system perspective," *Weather and Climate Extremes*, vol. 32, p. 100 313, Jun. 2021, issn: 2212-0947. doi: 10.1016/j.wace.2021. 100313. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2212094721000116>.
  - [54] Prabhat, S. Byna, V. Vishwanath, E. Dart, M. Wehner, and W. D. Collins, "TECA: Petascale Pattern Recognition for Climate Science," *en, in Computer Analysis of Images and Patterns*, G. Azzopardi and N. Petkov, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 426 - 436, isbn: 978-3-319-23117-4. doi: 10.1007/978-3-319-23117-4\_37.
  - [55] Prabhat, K. Kashinath, M. Mudigonda, et al., "ClimateNet: An expertlabeled open dataset and deep learning architecture for enabling highprecision analyses of extreme weather," *Geoscientific Model Development*, vol. 14, no. 1, pp. 107 - 124, 2021. doi: 10.5194/gmd- 14- 107- 2021. [Online]. Available: <https://gmd.copernicus.org/articles/ 14/107/2021/>.
  - [56] E. Racah, C. Beckham, T. Maharaj, S. Ebrahimi Kahou, M. Prabhat, and C. Pal, "ExtremeWeather: A large-scale climate dataset for semisupervised detection, localization, and understanding of extreme weather events," *in Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, et al., Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/519c84155964659375821f7ca576f095-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/519c84155964659375821f7ca576f095-Paper.pdf).
  - [57] R. A. Sobash, D. J. Gagne, C. L. Becker, D. Ahijevych, G. N. Gantos, and C. S. Schwartz, "Diagnosing Storm Mode with Deep Learning in Convection-Allowing Models," *EN, Monthly Weather Review*, vol. 151, no. 8, pp. 2009 - 2027, Jul. 2023, Publisher: American Meteorological Society Section: Monthly Weather Review, issn: 1520-0493, 0027-0644. doi: 10.1175/MWR-D-22-0342.1. [Online]. Available: <https://journals.ametsoc.org/view/journals/mwre/151/8/MWR-D- 22- 0342.1.xml> (visited on 08/21/2023).
  - [58] D. Mu~noz-Esparza, R. D. Sharman, and W. Deierling, "Aviation turbulence

- forecasting at upper levels with machine learning techniques based on regression trees,” *Journal of Applied Meteorology and Climatology*, vol. 59, no. 11, pp. 1883–1899, 2020. doi: <https://doi.org/10.1175/JAMC-D-20-0116.1>. [Online]. Available: <https://journals.ametsoc.org/view/journals/apme/59/11/JAMC-D-20-0116.1.xml>.
- [59] R. Lagerquist, A. McGovern, C. R. Homeyer, D. J. G. Li, and T. Smith, “Deep Learning on Three-Dimensional Multiscale Data for Next-Hour Tornado Prediction,” *EN, Monthly Weather Review*, vol. 148, no. 7, pp. 2837–2861, Jun. 2020, Publisher: American Meteorological Society Section: Monthly Weather Review, issn: 1520-0493, 0027-0644. doi: 10.1175/MWR-D-19-0372.1. [Online]. Available: <https://journals.ametsoc.org/view/journals/mwre/148/7/mwrD190372.xml> (visited on 08/21/2023).
- [60] C. A. Barajas, M. K. Gobbert, and J. Wang, “Performance Benchmarking of Data Augmentation and Deep Learning for Tornado Prediction,” in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 3607–3615. doi: 10.1109/BigData47090.2019.9006531.
- [61] L. Boussioux, C. Zeng, T. Gu’enaïs, and D. Bertsimas, “Hurricane Forecasting: A Novel Multimodal Machine Learning Framework,” *EN, Weather and Forecasting*, vol. 37, no. 6, pp. 817–831, Jun. 2022, Publisher: American Meteorological Society Section: Weather and Forecasting, issn: 1520-0434, 0882-8156. doi: 10.1175/WAF-D-21-0091.1. [Online]. Available: <https://journals.ametsoc.org/view/journals/wefo/37/6/WAF-D-21-0091.1.xml> (visited on 08/21/2023).
- [62] G.-F. Bian, G.-Z. Nie, and X. Qiu, “How well is outer tropical cyclone size represented in the ERA5 reanalysis dataset?” *Atmospheric Research*, vol. 249, p. 105339, 2021, issn: 0169-8095. doi: <https://doi.org/10.1016/j.atmosres.2020.105339>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S016980952031276X>.
- [63] T. Schneider, S. Behera, G. Boccaletti, et al., “Esa-ecmwf report on recent progress and research directions in machine learning for earth system observation and prediction,” *Nature Climate Change*, vol. 13, pp. 887–889, 2023.
- [64] M. J. Molina, T. A. O’ Brien, G. Anderson, et al., “A review of recent and emerging machine learning applications for climate variability and weather phenomena,” *Artificial Intelligence for the Earth Systems*, vol. 2, no. 4, p. 220086, 2023. doi: <https://doi.org/10.1175/AIES-D-22-0086.1>. [Online]. Available: <https://journals.ametsoc.org/view/journals/aies/2/4/AIES-D-22-0086.1.xml>.
- [65] D. Rolnick, P. L. Donti, L. H. Kaack, et al., “Tackling climate change with machine learning,” *ACM Comput. Surv.*, vol. 55, no. 2, Feb. 2022, issn: 0360-0300. doi: 10.1145/3485128. [Online]. Available: <https://doi.org/10.1145/3485128>.
- [66] H. Jain, R. Dhupper, A. Shrivastava, D. Kumar, and M. Kumari, “Aienabled strategies for climate change adaptation: Protecting communities, infrastructure, and businesses from the impacts of climate change,” *Computational Urban Science*, vol. 3, no. 1, p. 25, 2023.
- [67] A. Jones, J. Kuehnert, P. Fraccaro, et al., “Ai for climate impacts: Applications



- in flood risk,” *npj Climate and Atmospheric Science*, vol. 6, p. 63, 2023. doi: 10.1038/s41612-023-00388-1. [Online]. Available: <https://doi.org/10.1038/s41612-023-00388-1>.
- [68] S. Nevo, E. Morin, A. Gerzi Rosenthal, et al., “Flood forecasting with machine learning models in an operational framework,” *Hydrology and Earth System Sciences*, vol. 26, no. 15, pp. 4013–4032, 2022. doi: 10.5194/hess-26-4013-2022. [Online]. Available: <https://hess.copernicus.org/articles/26/4013/2022/>.
- [69] Z. Guo, J. P. Leit˜ao, N. E. Sim˜oes, and V. Moosavi, “Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks,” *en, Journal of Flood Risk Management*, vol. 14, no. 1, e12684, 2021, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jfr3.12684>, issn: 1753-318X. doi: 10.1111/jfr3.12684. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12684> (visited on 08/21/2023).
- [70] G. Alkhatyat and R. Mehmood, “A review and taxonomy of wind and solar energy forecasting methods based on deep learning,” *Energy and AI*, vol. 4, p. 100 060, Jun. 2021, issn: 2666-5468. doi: 10.1016/j.egyai. 2021.100060. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666546821000148> (visited on 08/17/2023).
- [71] S. K. Mukkavilli, A. A. Prasad, R. A. Taylor, A. Troccoli, and M. J. Kay, “Mesoscale simulations of australian direct normal irradiance, featuring an extreme dust event,” *Journal of Applied Meteorology and Climatology*, vol. 57, no. 3, pp. 493–515, 2018. doi: <https://doi.org/10.1175/JAMCD-17-0091.1>. [Online]. Available: <https://journals.ametsoc.org/view/journals/apme/57/3/jamc-d-17-0091.1.xml>.
- [72] S. Mukkavilli, A. Prasad, R. Taylor, et al., “Assessment of atmospheric aerosols from two reanalysis products over australia,” *Atmospheric Research*, vol. 215, pp. 149–164, 2019, issn: 0169-8095. doi: <https://doi.org/10.1016/j.atmosres.2018.08.026>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169809518301236>.
- [73] C. Hoyne, S. K. Mukkavilli, and D. Meger, “Deep learning for aerosol forecasting,” *arXiv preprint arXiv:1910.06789*, 2019. [Online]. Available: <https://arxiv.org/abs/1910.06789>.
- [74] T. Kurth, S. Subramanian, P. Harrington, et al., “Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators,” in *Proceedings of the Platform for Advanced Scientific Computing Conference*, Jun. 2023, pp. 1–11.
- [75] S. R. Cachay, B. Zhao, H. James, and R. Yu, “Dyffusion: A dynamicsinformed diffusion model for spatiotemporal forecasting,” *arXiv preprint arXiv:2306.01984*, 2023.
- [76] D. Ruhe, J. K. Gupta, S. De Keninck, M. Welling, and J. Brandstetter, “Geometric clifford algebra networks,” *arXiv preprint arXiv:2302.06594*, 2023.
- [77] G. Mialon, Q. Garrido, H. Lawrence, D. Rehman, Y. LeCun, and B. T. Kiani, “Self-supervised learning with lie symmetries for partial differential equations,” *arXiv preprint arXiv:2307.05432*, 2023.

- [78] B. Raonić, R. Molinaro, T. De Ryck, et al., “Convolutional neural operators for robust and accurate learning of pdes,” arXiv preprint arXiv:2302.01178, 2023.
- [79] O. Carlsson, J. E. Gerken, H. Linander, et al., “Heal-swin: A vision transformer on the sphere,” arXiv preprint arXiv:2307.07313, 2023.
- [80] P. Lippe, B. S. Veeling, P. Perdikaris, R. E. Turner, and J. Brandstetter, “Pde-refiner: Achieving accurate long rollouts with neural pde solvers,” arXiv preprint arXiv:2308.05732, 2023.
- [81] J. A. Weyn, D. R. Durran, and R. Caruana, “Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere,” *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 9, e2020MS002109, 2020.
- [82] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, IEEE, Jun. 2009, pp. 248–255. doi: INSERT \_ DOI \_ HERE \_ IF \_ AVAILABLE. [Online]. Available: INSERT \_ URL\_HERE\_IF\_AVAILABLE.
- [83] Y. LeCun, The mnist database of handwritten digits, [http : / / yann . lecun.com/exdb/mnist/](http://yann.lecun.com/exdb/mnist/), 1998.
- [84] M. Assran, Q. Duval, I. Misra, et al., “Self-supervised learning from images with a joint-embedding predictive architecture,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 15 619–15 629.
- [85] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, “Masked autoencoders are scalable vision learners,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.
- [86] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [87] Z. Tong, Y. Song, J. Wang, and L. Wang, “Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training,” *Advances in neural information processing systems*, vol. 35, pp. 10 078–10 093, 2022.
- [88] ClimFormer - A Spherical Transformer Model for Long-term Climate Projections, vol. *Proceedings of the Machine Learning and the Physical Sciences Workshop, NeurIPS 2022*, 2022.
- [89] S. Cho, R. Jung, and J. Kwon, “Spherical transformer,” arXiv, p. 2202.04942v2, 2022.
- [90] K. M. Gorski, E. Hivon, A. J. Banday, et al., “Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere,” *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.
- [91] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” arXiv preprint arXiv:1409.0473, 2014.
- [92] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, “Convolutional sequence to sequence learning,” in *International conference on machine learning*, PMLR, 2017, pp. 1243–1252.
- [93] Q. Guo, X. Qiu, P. Liu, Y. Shao, X. Xue, and Z. Zhang, “Star-transformer,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and*

- Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1315 – 1325. doi: 10.18653/v1/N19-1133. [Online]. Available: <https://aclanthology.org/N19-1133>.
- [94] A. v. d. Oord, Y. Li, and O. Vinyals, “Representation learning with contrastive predictive coding,” arXiv preprint arXiv:1807.03748, 2018.
- [95] M. Zaheer, G. Guruganesh, K. A. Dubey, et al., “Big bird: Transformers for longer sequences,” *Advances in neural information processing systems*, vol. 33, pp. 17 283 – 17 297, 2020.
- [96] Z. Chen, Z. Wang, Y. Yang, and J. Gao, “Resgraphnet: Graphsage with embedded residual module for prediction of global monthly mean temperature,” *Artificial Intelligence in Geosciences*, vol. 3, pp. 148 – 156, 2022.
- [97] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, “Graph wavenet for deep spatial-temporal graph modeling,” arXiv preprint arXiv:1906.00121, 2019.
- [98] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, “Hierarchical graph representation learning with differentiable pooling,” in *Advances in neural information processing systems*, vol. 31, 2018.
- [99] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. Battaglia, “Learning to simulate complex physics with graph networks,” in *International conference on machine learning*, PMLR, 2020, pp. 8459 – 8468.
- [100] T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. W. Battaglia, “Learning mesh-based simulation with graph networks,” arXiv preprint arXiv:2010.03409, 2020.
- [101] L. Lu, P. Jin, G. Pang, et al., “Learning nonlinear operators via deepnet based on the universal approximation theorem of operators,” *Nature Machine Intelligence*, vol. 3, pp. 218 – 229, 2021. doi: 10.1038/s42256-021-00302-5. [Online]. Available: <https://doi.org/10.1038/s42256-021-00302-5>.
- [102] Z. Li, N. Kovachki, K. Azizzadenesheli, et al., “Fourier neural operator for parametric partial differential equations,” arXiv preprint arXiv:2010.08895, 2020.
- [103] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *International conference on machine learning*, PMLR, 2020, pp. 1597 – 1607.
- [104] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729 – 9738.
- [105] L. Li, R. Carver, I. Lopez-Gomez, F. Sha, and J. Anderson, Seeds: Emulation of weather forecast ensembles with diffusion models, 2023. arXiv: 2306.14066 [cs.LG].
- [106] O. Ovadia, E. Turkel, A. Kahana, and G. E. Karniadakis, Ditto: Diffusioninspired temporal transformer operator, 2023. arXiv: 2307.09072 [cs.LG].