

# 海洋与大气领域大模型研究进展

张子良<sup>1</sup>, 于华明<sup>1,2</sup>, 任姝彤<sup>3,4\*</sup>, 张辰宇<sup>5</sup>

(1. 中国海洋大学海洋与大气学院, 山东 青岛 266100; 2. 中国海洋大学三亚海洋研究院, 海南 三亚 572025; 3. 国家海洋环境预报中心, 北京 100081; 4. 国家海洋环境预报中心 自然资源部海洋灾害预报技术重点实验室, 北京 100081; 5. 青岛埃克曼科技有限公司, 山东 青岛 266100)

**摘 要:** 详细探讨了近年来全球尺度上主要的海洋和大气深度学习预报模型, 尤其是基于Transformer架构的自回归模型。按照模型的提出时间, 重点介绍了FourCastNet、Pangu-Weather、ClimaX、GraphCast、FengWu、FuXi和AI-GOMS模型的基本架构、训练数据以及预报性能。通过比较这些模型, 揭示其发展的主要趋势, 即从简单的Transformer结构到引入AFNO、Swin Transformer、GNN等不同架构, 模型在结构上不断创新; 同时, 模型在输入变量选择、训练策略和计算效率方面也不断优化, 提升了预报的准确性和实用性。展望未来, 随着图形处理器(GPU)技术的进步和跨学科合作的加强, 预计模型将在提高预报技巧、延长预报时效和揭示新物理机制等方面取得进一步突破, 同时模型也将进一步提高计算效率和降低资源需求。

**关键词:** 深度学习; 数值预报; Transformer; 大模型

**中图分类号:** P732.6 **文献标识码:** A **文章编号:** 1003-0239(2025)02-0109-14

## 0 引言

海洋与大气系统在动力学方面展示了许多显著的相似性, 这一认知对沿海地区人民的生活产生了深远影响, 特别是在预报技术层面, 这些相似性对理解气候变化和灾害预防至关重要。传统预报模型, 如集成预报系统(Integrated Forecasting System, IFS)、天气研究与预报模型(Weather Research & Forecasting Model, WRF)和普林斯顿海洋模型(Princeton Ocean Model, POM)在过去取得了显著成就, 但它们在处理复杂气候系统时的速度和精确度受到限制。近年来, 机器学习和深度学习兴起, 尤其是FourCastNet等模型在全球0.25°分辨率气象预报中被广泛应用, 并展示出较传统模型更快的计算速度和更高的准确度。这些深度学习模型的部署成本相对较低, 训练后甚至可以在个人电脑上运行, 应用潜力较大。本文旨在综述全球主要

的海洋和大气深度学习预报模型, 深入探讨它们在现代气象预报中的应用和发展。通过分析这些模型, 旨在为研究者提供背景资料, 并就深度学习模型在未来气象预报中的潜在应用提出见解和建议, 为该领域的进一步研究提供方向和灵感。

需要指出的是, 深度学习领域近期流行的“大模型”一般是指采用大量参数的大型语言模型(Large Language Model, LLM)。早期的深度学习气象模型主要进行单点预报<sup>[1-2]</sup>或者区域平均值预报<sup>[3]</sup>, 而非二维区域预报, 前几年的深度学习海洋模型也专注于单点预报<sup>[4-6]</sup>, 这些预报模型的变量数普遍较小, 基本只能作为传统预报模型在某一需求区域的补充。本文使用“大模型”一词的主要原因在于, 与先前海洋与大气领域的深度学习模型相比, 本文介绍的模型是针对全球范围内多个变量(20个及以上)进行自回归建模, 同时它可以提供主要预报产品, 而非仅仅作为补充。因此, 尽管这些模型

收稿日期: 2024-01-02。

基金项目: 山东省高等学校青年创新团队计划(2024KJN017)。

作者简介: 张子良(1996-), 男, 硕士, 主要从事深度学习在海洋预报方面的应用工作。E-mail: aldrichsmithjp@gmail.com

\*通信作者: 任姝彤(1985-), 女, 高级工程师, 硕士, 主要从事海洋观测预报与防灾减灾战略研究工作。E-mail: evarentong@126.com

在参数规模上可能不及传统深度学习中的“大模型”标准,但由于其研究范围和变量显著增加,我们认为将其称为气象大模型或海洋大模型是恰当的。

## 1 深度学习及Transformer

鉴于目前比较有名的大模型主要基于深度学习以及Transformer在计算机视觉方面的多种拓宽及应用,因此我们先介绍这两个部分。

深度学习是一种基于人工神经网络的计算方法,近年来因其强大的数据处理能力在多个领域获得广泛应用。这种方法的核心在于其多层次网络结构,它可以从数据中自动提取复杂特征,特别适合处理非线性和高维度数据。深度学习在图像识别、语言处理等领域取得了显著成就,主要归功于其出色的模式识别能力和高效处理大规模复杂数据集的能力。然而,这种方法也有局限,例如高度依赖大量高质量数据,在需要高解释性的场景下的应用受限。

在本文中,我们将介绍多个模型的架构、训练数据及输入、输出变量。在这些模型中,由于Pangu-Weather是首个全面超过传统预报模型的深度学习模型,FuXi是本研究进行期间效果最好的模型,AI-GOMS是深度学习在海洋方面的应用,因此我们还将介绍这几个模型的预报效果。

Transformer模型自2017年问世以来<sup>[7]</sup>,已成为深度学习领域的一项重大突破。它最初应用于机器翻译领域,但由于其高效的架构和处理机制,也被广泛应用于自然语言处理和图像识别等多个领域。标准的Transformer模型由编码器和解码器构成,前者将输入序列处理为隐向量,后者则根据编码器的输出和之前的结果生成目标序列。该处理流程的核心是多头自注意力(Multi-head Self-Attention)机制,它允许模型同时处理序列中的所有元素,并且通过计算序列中元素之间的相关性,使得模型在编码或生成每个元素时,都能充分利用序列中其他元素的信息。与传统递归或卷积神经网络(Convolutional Neural Network, CNN)相比,Transformer在处理长序列数据时能更有效地捕捉长距离依赖关系,同时还具有并行计算的优势,将这个特点与图形处理器(Graphics Processing Unit,

GPU)的发展相契合,大大提升了Transformer的运算效率。然而,Transformer的高计算成本也对计算资源提出了更高的要求。

## 2 预报模型介绍

### 2.1 FourCastNet模型

#### 2.1.1 模型架构

FourCastNet模型由Pathak等<sup>[8]</sup>于2022年提出。在短期预报(3天)中,模型对大尺度变量的预报准确度与欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF)综合预报系统(The Integrated Forecast Systems, IFS)的结果相当。该模型是一种先进的数值天气预报模型,被广泛用于全球天气预测和气候研究。

为了能进行有效训练,先前的深度学习预报模型常常采用分辨率较低的数据<sup>[9-10]</sup>,尽管取得了一些成果,但也缺失了许多物理细节。为了能够处理高分辨率数据,FourCastNet结合了Vision Transformer(ViT)和自适应傅里叶神经算子(Adaptive Fourier Neural Operator, AFNO)模型。相较于常处理视觉问题的CNN架构(通常在单一窗口内捕捉数据特征),ViT通过采用多头自注意力机制,能够在全局角度上处理数据间的关系。初步实验结果也显示,ViT在长时间过程中对运动细节的捕捉能力优于CNN<sup>[8]</sup>。然而,多头自注意力机制带来了较大的时间复杂度,即 $O(n^2)$ ,其中 $n$ 为输入的令牌数(token),使得ViT在处理高分辨率数据时十分困难。受到傅里叶神经算子在提升纳维-斯托克斯方程和其他偏微分方程求解速度<sup>[11]</sup>的启发,AFNO被提出并专门用于处理高分辨率数据<sup>[12]</sup>。AFNO通过在傅里叶域中使用快速傅里叶变换,实现了一种覆盖全局并且能够连续处理输入数据的卷积操作,使得时间复杂度降至 $O(n\log(n))$ ,这为模型的构建提供了可能。

FourCastNet架构见图1。输入数据被切割成 $(h \times w)$ 块(patch), $h$ 和 $w$ 分别表示行数和列数,每块的大小为 $p \times p \times c$ , $p$ 表示像素长度, $c$ 表示通道数;随后对这些块进行嵌入和位置编码,形成输入令牌;利用AFNO对令牌在空间上进行混合和处理,整个过程在多个层中重复;最后使用一个线性解码

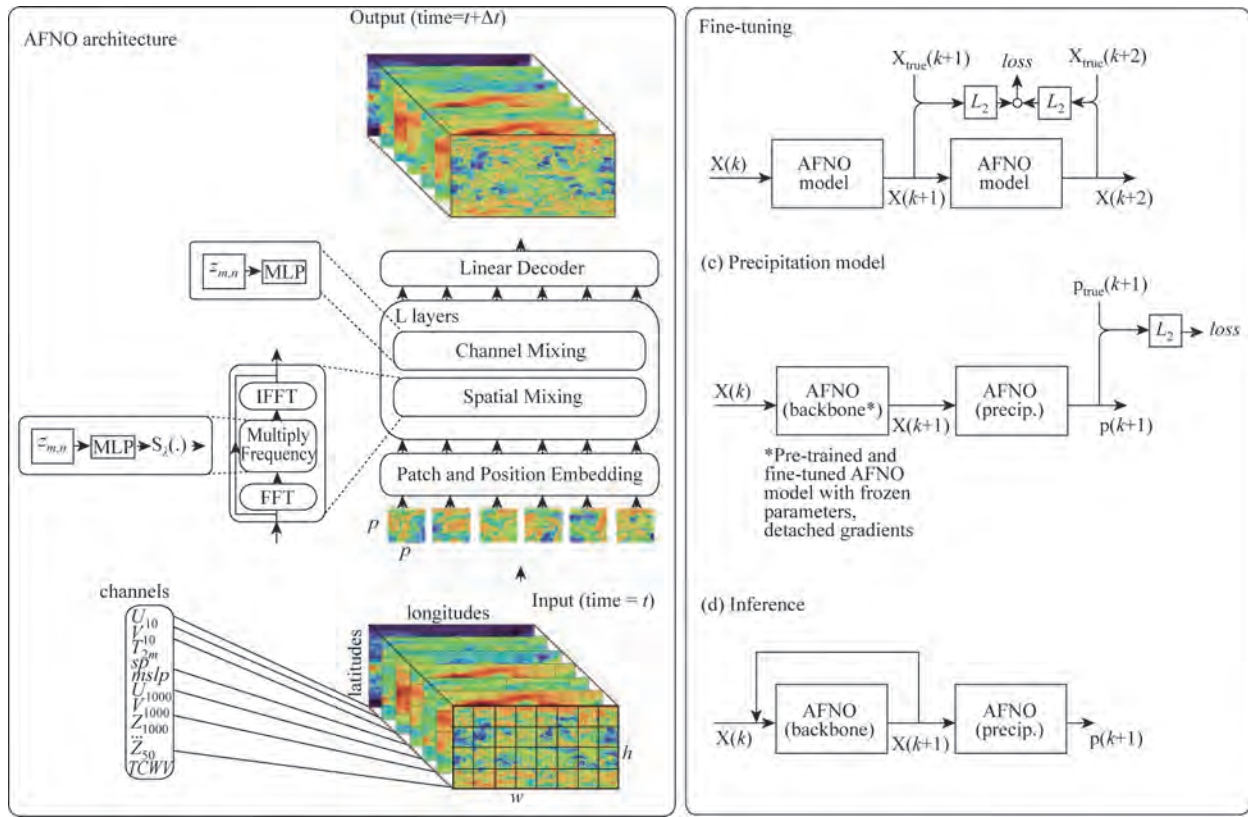


图1 FourCastNet架构示意(引自文献[8])

Fig.1 Architecture schematic of the FourCastNet model (cited from reference [8])

器得到模型输出数据。

### 2.1.2 训练数据及输入、输出变量

FourCastNet模型使用了ECMWF的第5代再分析数据ERA5(ECMWF Reanalysis v5),选取20个变量作为训练数据,包括风速、温度、气压等气象参数(见表1)。这些数据已通过高斯网格转换为分辨率

为 $0.25^\circ$ 的规则欧几里得网格,并形成一个维度为 $721 \times 1440 \times 20$ 的张量。训练数据集涵盖了1979—2015年的数据,验证集为2016—2017年,2018年及后续年份的数据用作测试集。

FourCastNet训练分为两个阶段。预训练阶段,模型学习从时间点 $X(k\Delta t) \rightarrow X[(k+1)\Delta t]$ 的映射,

表1 不同模型的训练变量

Tab.1 Training variables for different models

模型名称	变量
FourCastNet	$U_{10}, V_{10}, T_{2m}, sp, mslp, T \times 2, U \times 3, V \times 3, Z \times 3, RH \times 2, TCWV$
Pangu-Weather	$U_{10}, V_{10}, T_{2m}, mslp, T \times 13, U \times 13, V \times 13, Z \times 13, RH \times 13$
ClimaX	$U_{10}, V_{10}, T_{2m}, T \times 7, U \times 7, V \times 7, Z \times 7, RH \times 7, SH \times 7$
GraphCast	$U_{10}, V_{10}, T_{2m}, mslp, tp, T \times 37, U \times 37, V \times 37, W \times 37, Z \times 37, SH \times 37$
FengWu	$U_{10}, V_{10}, T_{2m}, mslp, T \times 37, U \times 37, V \times 37, Z \times 37, RH \times 37$
FuXi	$U_{10}, V_{10}, T_{2m}, mslp, tp, T \times 13, U \times 13, V \times 17, Z \times 13, RH \times 13$
AI-GOMS	$U_{10}, V_{10}, T_{2m}, sp, mslp, ssh, t \times 15, u \times 15, v \times 15, s \times 15$



其中  $\Delta t$  是固定的 6 h 时间间隔。微调阶段,模型优化从  $X[(k+1)\Delta t] \rightarrow X[(k+2)\Delta t]$  的映射。这种训练方法使 FourCastNet 能够准确预测未来天气状态,提高了其在复杂气候条件下的性能和准确性。

表 1 展示了本节中所有模型使用的变量,其中  $U_{10}$  ( $V_{10}$ ) 为距地表 10 m 处的纬向(经向)风速、 $T_{2m}$  为距地表 2 m 处的温度、sp 为海平面气压、mslp 为平均海平面气压、tp 为 6 h 总降水、ssh 为海表面高度、 $T$  为温度、 $U$  为纬向风速、 $V$  为经向风速、 $W$  为垂向风速、 $Z$  为位势、SH 为比湿、RH 为相对湿度、TCWV 为整柱水汽总量、 $u$  为纬向流速、 $v$  为经向流速、 $s$  为海水盐度,此外,  $x_n$  表示输入  $n$  层该变量。

## 2.2 Pangu-Weather 模型

### 2.2.1 模型架构

Pangu-Weather 模型由 Bi 等<sup>[13]</sup>于 2022 年提出,并于 2023 年正式发表<sup>[14]</sup>。与 IFS 相比, Pangu-Weather 模型在一周内时间尺度的预报中,所有测试变量均获得了更强的确定性预报结果。

ViT 的改进一直是比较热门的问题, AFNO 是其中的一个改进方案,但由于改进的通用性略低,因此风头很快被同一时期通用性更高的 Swin

Transformer 超过<sup>[15]</sup>。Swin Transformer 在对 ViT 的改进中,通过移动窗口以及相对位置偏差编码方案,将自注意力计算限制在非重叠的局部窗口内,同时允许跨窗口连接,从而提高了效率,这一改进也被运用于 Pangu-Weather 模型中。Pangu-Weather 模型以 Swin Transformer 为基础构建 3D Earth-specific transformer (3DEST),将其中的相对位置偏差编码方案替换成地球特定位置偏差编码方案来适应数据结构。

与 FourCastNet 类似,数据被切割成不同的块进行处理(见图 2),不同的是, PanGu-Weather 将含有多个层的变量按照  $2 \times p \times p \times c$  的大小进行立方体切割,即 2 个气压层合在一起切割,缺失层以 0 值填充,而单层变量则仍是按照  $p \times p \times c$  的大小进行切割。这些块随后进行嵌入以及位置编码,形成输入令牌,通过 3DEST 进行处理并输出。

### 2.2.2 训练数据及输入输出变量

Pangu-Weather 模型使用了与 FourCastNet 相同的 ERA5 数据,但变量更多(69 个),包括风速、温度、气压等关键气象参数,具体变量见表 1。这些数据被插值到 0.25° 分辨率的  $721 \times 1\,440$  网格上,数据频率为每小时。

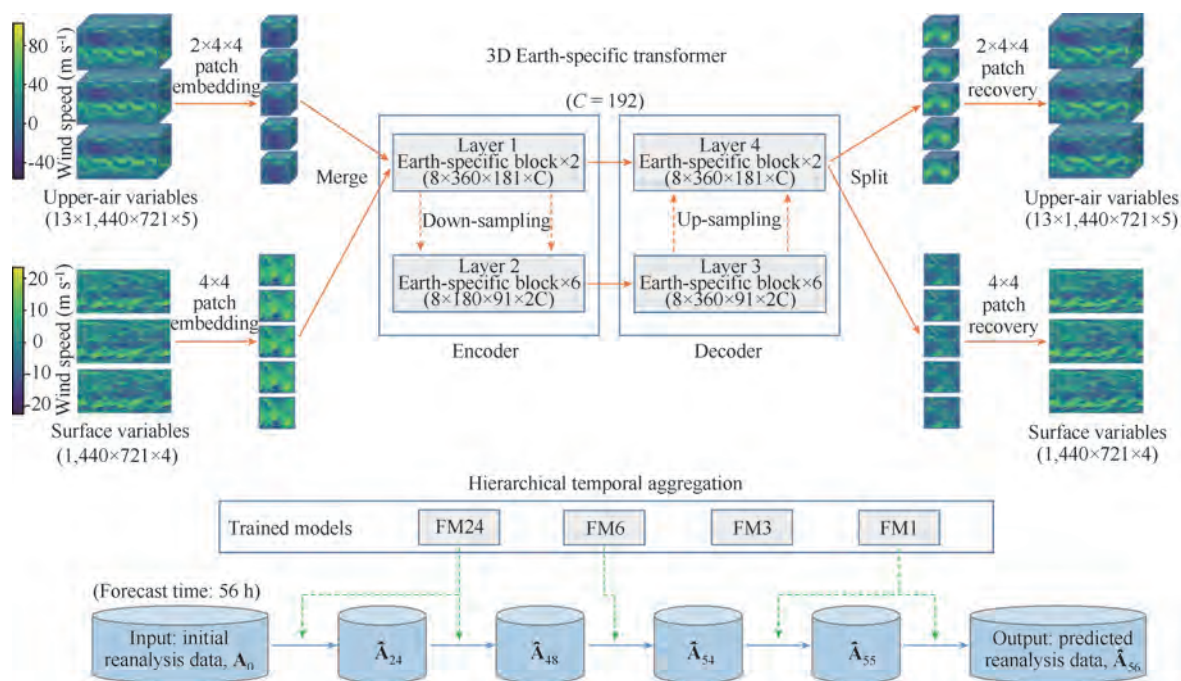


图2 Pangu-Weather 架构示意(引自文献[13])

Fig.2 Architecture schematic of the Pangu-Weather model (cited from reference [13])

Pangu-Weather的训练集包括1979—2017年的数据,验证集为2019年的数据,而2018年、2020年和2021年的数据用作测试集。在预训练过程中,模型主要学习了从时间点 $X(k\Delta t)$ — $X[(k+1)\Delta t]$ 的映射,其中 $\Delta t$ 表示不同的时间间隔。在先前的深度学习预报模型中,常常采用一个模型进行固定时间步长自回归迭代,因此预报时间越长,模型的累计误差越大。为了减少迭代式预测中的误差累积,Pangu-Weather实际上训练了4个不同模型,分别针对1 h、3 h、6 h和24 h时间尺度,在进行预报推理时,采用贪心算法(Greedy Algorithm),即总是优先调用预测时长最长的可用模型。例如,当目标预测时长为56 h时,会按顺序调用24 h模型2次,6 h模型1次,1 h模型2次。由此可以大大减少迭代次数,并减少误差累积。

### 2.2.3 预报效果

在评价模型效果中,常常使用均方根误差(Root Mean Squared Error, RMSE)和异常相关系数(Anomaly Correlation Coefficient, ACC)作为指标。RMSE越低,ACC越高,说明模型效果越好。RMSE的计算公式为:

$$E_{\text{RMS}}(v, t) = \sqrt{\frac{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} L(i) (\hat{A}_{i,j,t}^v - A_{i,j,t}^v)^2}{N_{\text{lat}} \times N_{\text{lon}}}} \quad (1)$$

$$L(i) = N_{\text{lat}} \times \frac{\cos \phi_i}{\sum_{i=1}^{N_{\text{lat}}} \cos \phi_i} \quad (2)$$

式中: $v$ 表示用于比较的变量; $A_{i,j,t}^v$ 是一个标量,表示 $v$ 在 $t$ 时刻在 $(i,j)$ 处的值; $N$ 表示网格数; $\phi$ 表示纬度。ACC的计算公式为:

$$C_{\text{AC}}(v, t) =$$

$$\sqrt{\frac{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} L(i) (\hat{A}_{i,j,t}^v - A'_{i,j,t})^2}{\sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} L(i) (\hat{A}_{i,j,t}^v)^2 \times \sum_{i=1}^{N_{\text{lat}}} \sum_{j=1}^{N_{\text{lon}}} L(i) (A'_{i,j,t})^2}} \quad (3)$$

式中: $A'$ 表示 $A$ 与气候学数值间的差异,气候学数值是指利用39年以上的训练数据估计的天气状态的长期平均值。其余变量同式(1)。

Pangu-Weather在多方面展现了不错的效果,但

我们尤其关心其与传统预报模型的比较。图3显示了168 h纬向平均后Pangu-Weather、IFS和FourCastNet预报效果的ACC和RMSE。从结果可见,在整个预报时间段内,Pangu-Weather的预报效果优于IFS和FourCastNet。

## 2.3 ClimaX模型

### 2.3.1 模型架构

ClimaX模型由Nguyen等<sup>[16]</sup>于2023年提出,是一个面向天气和气候预测的基础模型,特别善于处理多种异质数据源。与IFS相比时,对于14天内空间分辨率为 $1.40625^\circ$ 和 $5.625^\circ$ 的预报,ClimaX模型在所有测试变量上的效果与IFS相当甚至更优。

ClimaX也采用ViT作为基础架构(见图4),但不同于之前模型采用ERA5数据作为单一数据源,ClimaX的目标是制作可以对异源数据进行训练的基础模型,因此其在ViT的架构上做了两个更改,分别称为可变令牌化(Variable Tokenization)和可变聚合(Variable Aggregation),前者通过对每个变量使用单独的嵌入方式,使得模型可以更好地处理每一个变量,但同时计算需求也更大,因此后者通过采用交叉注意力(Cross-attention)方式将同一空间位置的不同变量视为一个向量进行处理,使得最后的序列长度仍为地图尺寸。ClimaX也在可变聚合的过程中加入了可变位置嵌入(Variable Positional Embedding)以标记不同变量,解码后能够输出不同变量。

综上,ClimaX更改了数据的维度排布,输入的数据将以 $c \times p \times p$ 的块进行切割及嵌入,再通过交叉注意力进行嵌入和位置编码,形成输入令牌,之后通过ViT进行处理和输出。

### 2.3.2 训练数据及输入输出变量

ClimaX模型在预训练阶段使用了第六次国际耦合模型比较计划(Coupled Model Intercomparison Project Phase 6, CMIP6)的历史模拟数据,涵盖了MPI-ESM(Max Planck Institute Earth System Model)、TaiESM(Taiwan Earth System Model)、AWI-ESM(Alfred Wegener Institute Earth System Model)、HAMMOZ(Chemistry-Aerosol-Meteorology-Ocean-model system with Zoom capabilities)、CMCC(China Meteorological Administration Climate Change

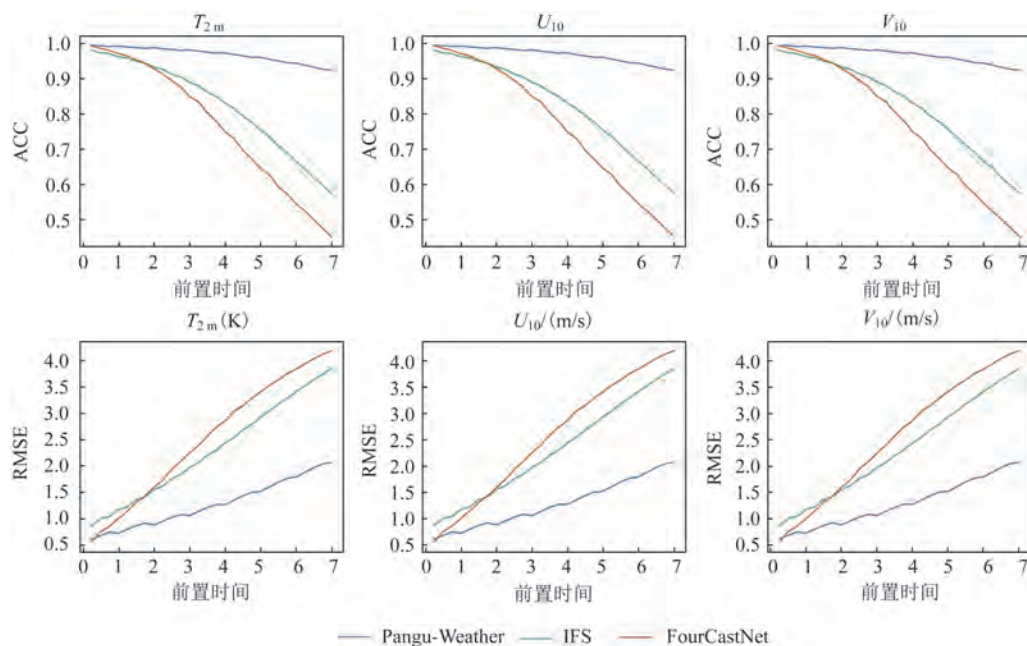


图3 Pangu-Weather、IFS 和 FourCastNet 的  $T_{2m}$ 、 $U_{10}$ 、 $V_{10}$  预报效果评估曲线(引自文献[14])

Fig.3 Evaluation curves for the forecasting performance of Pangu-Weather, IFS, and Four CastNet for  $T_{2m}$ ,  $U_{10}$  and  $V_{10}$  (cited from reference [14])

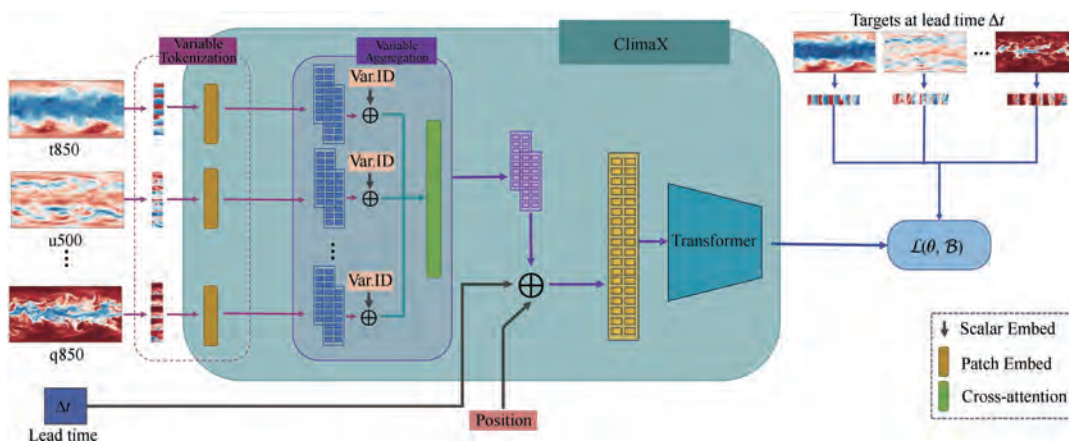


图4 ClimaX 架构示意(引自文献[16])

Fig.4 Architecture schematic of the ClimaX model (cited from reference [16])

Center) 这 5 个数据源,而在微调阶段则转向使用 ERA5 数据<sup>[17]</sup>。此模型涉及的主要变量包括风速、温度、位势、比湿和相对湿度等(见表 1)。CMIP6 的数据时间段为 1850—2015 年,而 ERA5 的数据覆盖了 1979—2018 年。考虑到数据量庞大,ClimaX 将这些数据插值到  $1.40625^\circ$  和  $5.625^\circ$  两种分辨率,时间分辨率为 6 h。

ClimaX 的创新之处在于其能够预测多个时间

尺度的气候。模型将时间间隔  $\Delta t$  作为输入的一部分,随机选择 6~168 h 中的值,并设置目标数据为对应时间点,形成  $X(t, \Delta t) \rightarrow X(t + \Delta t)$  的映射。在微调阶段,当下游任务使用的变量与预训练阶段相同时,对整个模型进行微调;如果出现新变量,则使用新初始化的网络替换嵌入层和预测头,同时微调或冻结变量聚合模块和注意力模块。这种设计使得 ClimaX 能够灵活适应多种时间尺度和不同数据集,



从而提高预测的准确性和适用性。

## 2.4 GraphCast模型

### 2.4.1 模型架构

GraphCast模型由Lam等<sup>[18]</sup>于2022年提出,并于2023年发表<sup>[19]</sup>。在涵盖10天范围的1380个预报变量中,GraphCast模型在90.3%的变量上的表现优于高分辨率的欧洲中期天气预报中心模型(High-Resolution European Centre for Medium-Range Weather Forecasts Model, HRES)。HRES是ECMWF开发的先进的数值天气预报模型,是IFS中的高分辨率确定性预报组件,可提供10天以内的数值天气预报,较IFS更侧重于预报结果的细节和精确度。

GraphCast是一个中期全球天气预报模型,采用图神经网络(Graph Neural Networks, GNN)架构(见图5)。与常见的Transformer结构不同,GNN是一种专门用于处理图结构数据的深度学习架构,如社

交网络、化学分子结构以及推荐系统中的实体(节点)和它们之间的关系(边)。

GraphCast的核心在于有效地在图结构中传递和聚合信息,并利用节点间的相互关系来更新每个节点的特征表示,这种方法使得GNN在图分类、节点分类和链接预测等任务中展现了出色的性能。GraphCast使用1 038 240个网格节点( $721 \times 1\,440$ )处理474个输入变量,包括5个地表变量、37层大气变量的两个时间步数据、3个时间步的5个强迫项和5个常数。模型的整体架构由多层感知机(Multi-Layer Perceptron, MLP)组成,它使用前两个时间步长的数据来预测后一个时间步长的天气,展现了在中期天气预报中的高效性和准确性。

### 2.4.2 训练数据及输入输出变量

GraphCast模型应用了ERA5数据集,共选取了227个变量,包括风速、温度、气压、降水量等多种气象参数(见表1)。变量的空间分辨率为 $0.25^\circ$

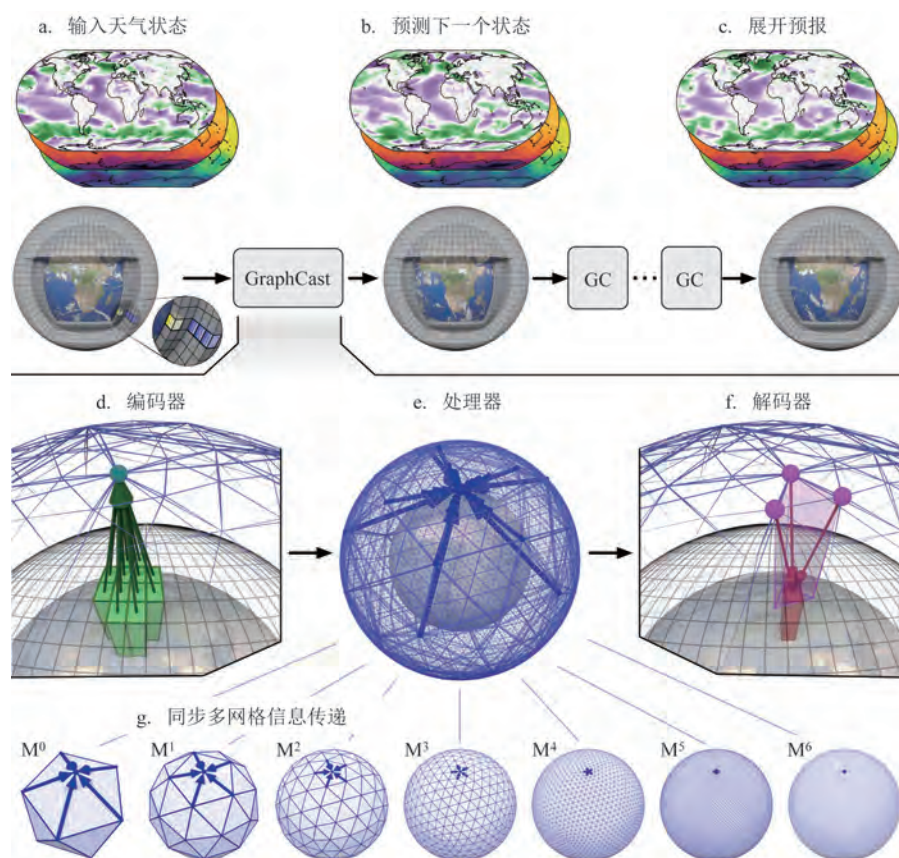


图5 GraphCast架构示意(引自文献[19])

Fig.5 Architecture schematic of the GraphCast mode (cited from reference [19])

( $721 \times 1\,440$ ), 时间分辨率为 6 h。训练数据集涵盖 1979—2015 年, 验证集为 2016—2017 年, 测试集为 2018—2021 年。

GraphCast 的训练过程特别注重学习率的调整。最初, 模型将学习率从 0 线性提升到  $1e-3$ , 完成单步预测模型的训练; 随后, 模型采用半余弦衰减函数来降低学习率, 最终在固定学习率  $3e-7$  的条件下, 逐步增加自回归步长 (从 2 步增加到 12 步)。这种训练方法使 GraphCast 能够有效地处理多时间尺度的预测, 同时确保了学习过程的稳定性和高效性。

## 2.5 FengWu 模型

### 2.5.1 模型架构

FengWu 模型由 Chen 等<sup>[20]</sup>于 2023 年提出, 在涵盖 10 天范围内的 880 个预报变量中, 该模型 80% 的预报表现优于 GraphCast。

与 ClimaX 类似, FengWu 对每个变量的数据进行单独嵌入 (见图 6), 但没有将变量再次聚合以及切割成块, 而是使用基于 Transformer 的跨模态融合器 (Cross-modal Fuser) 直接进行处理。这样能够更细致地捕捉每一个变量的特征, 缺点在于进一步增加了运算量, 但是由于显卡性能的提升 (在单精度运算上 A100 显卡较 V100 显卡提升近 10 倍<sup>[21]</sup>), 上述缺点则变得不太显著。

### 2.5.2 训练数据及输入输出变量

FengWu 模型使用 ERA5 数据进行训练, 共选取了 189 个变量, 包括风速、温度、气压等关键气象参

数 (见表 1)。这些变量的分辨率为  $0.25^\circ$  ( $721 \times 1\,440$ ), 时间分辨率为 6 h。模型的训练数据覆盖 1979—2015 年, 验证集为 2016—2017 年, 2018 年的数据被用作测试集。

FengWu 借鉴了 GraphCast 模型逐步增加自回归步长的方法来提高长期预测的准确性。在训练过程中, FengWu 首先学习了从当前状态  $X(k\Delta t)$ — $X[(k+1)\Delta t]$  的映射, 其中  $\Delta t$  为固定的 6 h 间隔。为了优化长期预测性能, FengWu 逐渐增加了自回归步长。FengWu 的创新之处在于引入了回放缓冲区机制, 即通过存储初步预测结果并在后续训练中重用这些结果, 减轻了内存压力并提高了计算效率。这使得 FengWu 在训练过程中更全面地考虑了自回归误差累积, 有效支持了长期自回归预测。

## 2.6 FuXi 模型

### 2.6.1 模型架构

FuXi 模型由 Chen 等<sup>[22]</sup>于 2023 年提出, 是一个专为 15 天全球天气预测构建的级联机器学习系统。在 15 天的预报时效上, FuXi 的表现与 ECMWF 集合平均 (ECMWF ensemble prediction system, EPS) 相当。EPS 同样是 ECMWF 开发的一个先进的数值天气预报模型, 基于 IFS 的概率预报系统, 能够进行 15 天甚至更长期的预报。

与 FengWu 通过计算量来提升模型预报准确性不同, FuXi 依然对 ViT 进行改进。自 Swin Transformer 作为 ViT 的一种改进方案被提出后, 暴露了诸如训练不稳定等问题。于是, Swin

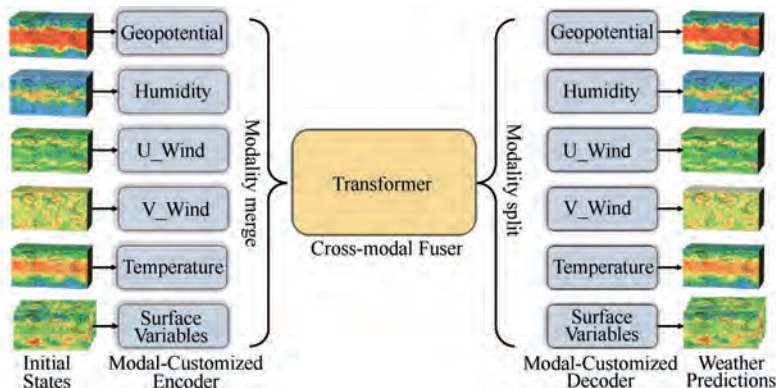


图 6 FengWu 架构示意 (引自文献[20])

Fig.6 Architecture schematic of the FengWu model (cited from reference [20])



Transformer V2应运而生<sup>[23]</sup>,它除了提高模型的训练稳定性外,还提升了模型性能,减少了对GPU内存的占用,为进一步实现更高的分辨率提供了可能。FuXi采用了由48个重复的Swin Transformer V2模块构建的U-Transformer,同时引入了U-NET<sup>[24]</sup>的采样结构,以更有效地提取数据信息。

FuXi采用了类似Pangu-weather的立方体切片方式(见图7),并更改了数据排布的方式。FuXi将两个时刻的数据同时输入,输入的数据被切割成 $2 \times c \times p \times p$ 的大小,其中的2指代不同时刻;之后嵌入数据并进行位置编码,利用U-Transformer进行处理。

## 2.6.2 训练数据及输入输出变量

FuXi模型使用ERA5数据进行训练,选取了70个变量,包括风速、温度、气压等(见表1)。这些变量大部分与Pangu-Weather相同,但在地表处增加了一个额外变量。这些变量的分辨率为 $0.25^\circ(721 \times 1440)$ ,时间分辨率与FourCastNet相同,为6 h。训练集包含1979—2015年的数据,验证集为2016—2017年的数据,2018年的数据用作测试集。

FuXi模型的独特之处在于对不同时间窗口的处理方式。模型被分为FuXi-Short(0~5天)、FuXi-Medium(5~10天)和FuXi-Long(10~15天)3个版本,每个版本都用前两个时间步的数据来预测后一

个时间步。在每个时间窗口分隔处,输入数据都源自前一个模型。FuXi首先对单步预测模型进行预训练,并逐渐增加自回归步长,这与GraphCast的训练策略类似。FuXi-Short训练完成后,其权重和预测结果被用作FuXi-Medium的初始值和部分训练数据,同样的方法也应用于FuXi-Long的训练。这种分阶段训练方法使FuXi能够有效处理不同长度的预测窗口,提高模型在短期—长期天气预报中的准确性和效率。

## 2.6.3 预报效果

从FuXi、GraphCast和HRSE纬向平均预报结果的ACC与RMSE的对比(见图8)可以很明显地看到,在第5天时,FuXi的预报精度稍有提升,其整体预报效果在后几天优于GraphCast,在前几天的效果也与GraphCast不相上下。

## 2.7 AI-GOMS模型

### 2.7.1 模型架构

AI-GOMS模型由Xiong等<sup>[25]</sup>于2023年提出,它是一个全球海洋模型,模型在30天内的预报表现较好。AI-GOMS模型在基础架构上与FourCastNet类似(见图9),即通过AFNO来解决ViT计算复杂的问题。与FourCastNet最终采用节省算力的线性模块

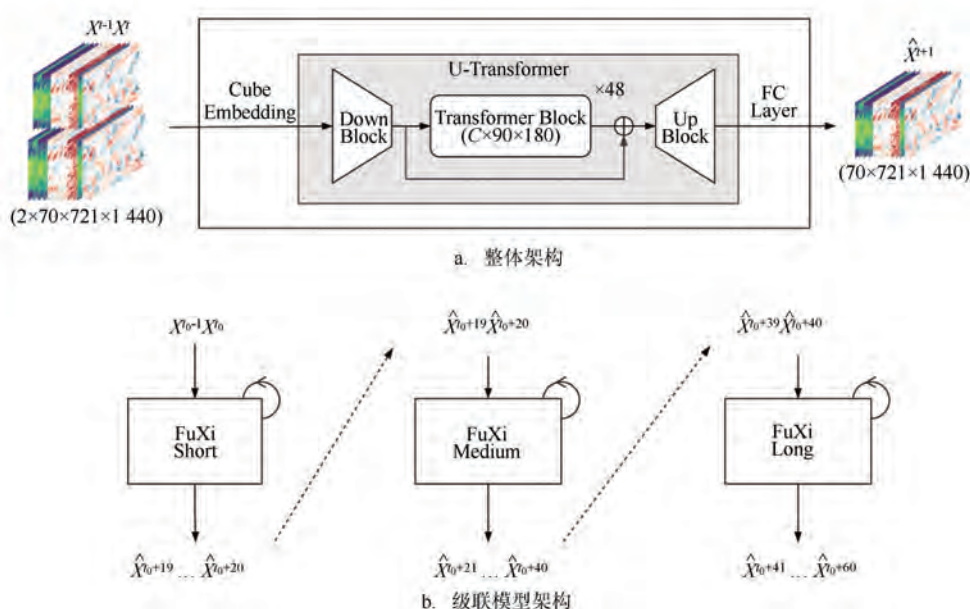


图7 FuXi架构示意(引自文献[22])

Fig.7 Architecture schematic of the FuXi mode (cited from reference [22])

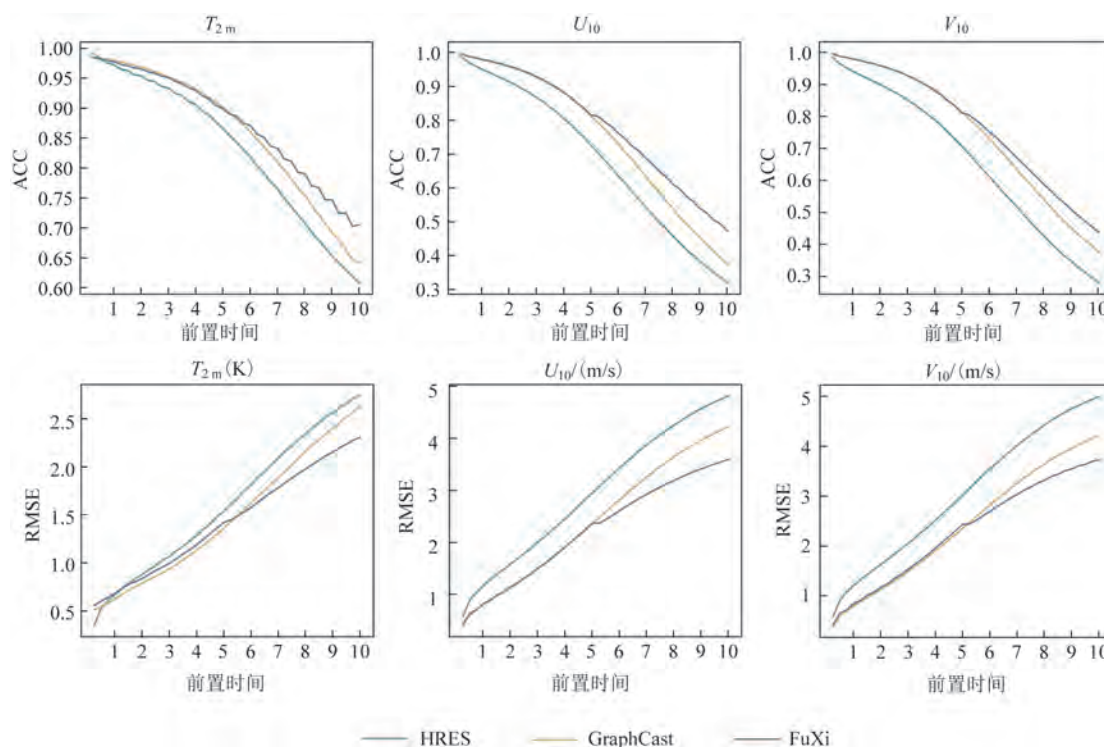


图8 HRES、GraphCast和FuXi的 $T_{2m}$ 、 $U_{10}$ 、 $V_{10}$ 预报效果评估曲线(引自文献[23])

Fig.8 Evaluation curves for the forecasting performance of HRES, GraphCast, and FuXi for  $T_{2m}$ ,  $U_{10}$  and  $V_{10}$  (cited from reference [23])

完成解码不同,受益于GPU的发展,AI-GOMS采用了浅层的AFNO模块进行解码,使得输出结果中包含更多细节。

在AI-GOMS模型中,数据被切割成 $c \times p \times p$ 大小的块,随后进行嵌入和位置编码,并通过随机掩码策略(Random Mask Strategy)进行数据增强,以提高模型性能,最后生成输入的令牌。这些令牌通过AFNO模块在空间上进行混合和处理。

### 2.7.2 训练数据及输入输出变量

AI-GOMS模型在数据处理上采用了独特的方法。它主要使用由海军耦合海洋数据同化系统(Navy Coupled Ocean Data Assimilation, NCODA) 3D-Var 算法同化的混合坐标海洋模型(Hybrid Coordinate Ocean Model, HYCOM)再分析数据集,结合了ERA5的表面数据作为边界条件以及全球水深地形高程数据集(ETOPO)的地形数据。模型共选取了67个变量,包括风速、温度、气压、海水温度、海水速度、海水盐度和海表面高度等(见表1)。这些变量的空间分辨率为 $0.25^\circ (720 \times 1440)$ ,时

间分辨率为24 h。训练数据覆盖2000—2010年,验证集为2011年的数据,测试集则为2012年的数据。

与大气预测模型不同,AI-GOMS模型的输出量包括海水温度、海水速度、盐度和海表面高度等。在其迭代预测的过程中,除了使用自身的预测结果作为输入,还需不断引入大气和地形的边界条件数据。模型训练的目标是根据当前时间步的预报数据和边界数据预测下一个时间步的状态。在预训练阶段,AI-GOMS采用了逐渐减少掩码比率的随机掩码策略,并结合了两步监督和余弦学习率计划。在此基础上,模型对区域降尺度、波浪解码和生化耦合3个下游任务进行了微调。

### 2.7.3 预报效果

由于该模型是海洋的第一个大尺度预报模型,且预报时间较长,因此与气象模型相比,该模型预报结果的ACC和RMSE并不好,但是从文献[26]中可以看到(图略),模型的预报结果也能较为准确地反映海洋特征。

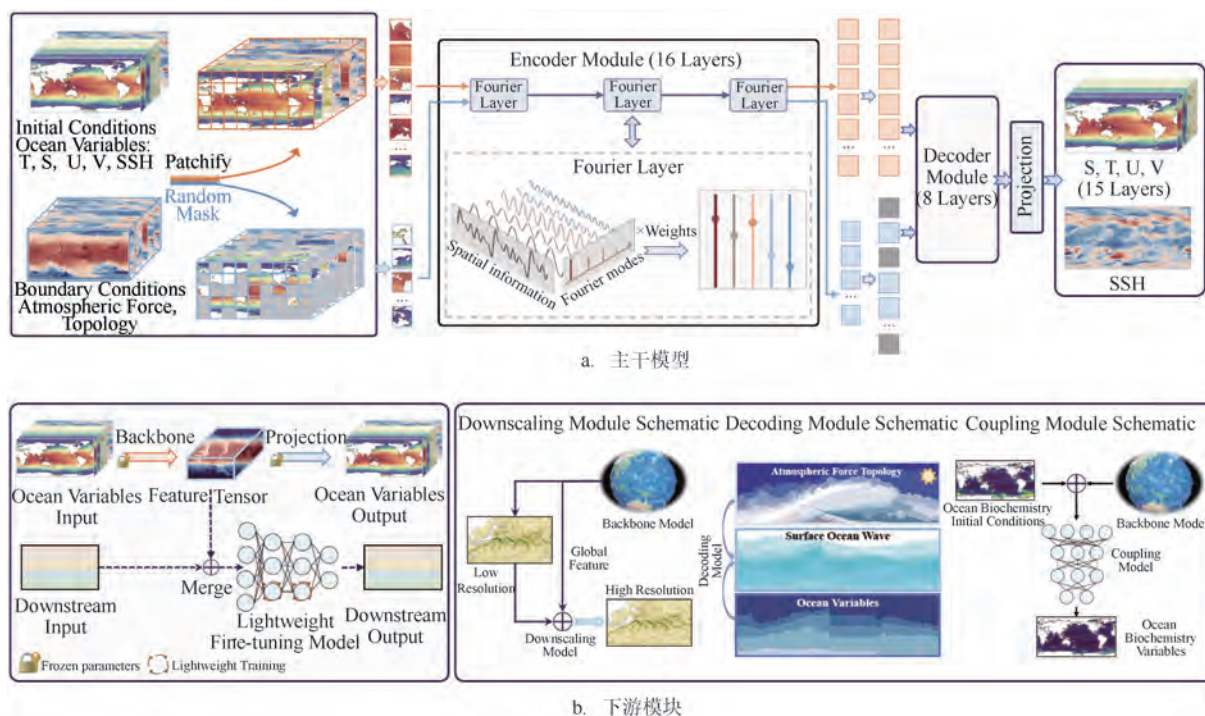


图9 AI-GOMS架构示意(引自文献[26])

Fig.9 Architecture schematic of the AI-GOMS mode (cited from reference [26])

表2 不同模型的特征

Tab.2 Characteristics of different models

模型名称	主要架构	训练代价	模型优势
FourCastNet	ViT+AFNO	64*V100*26 h	3天内0.25°空间分辨率预报的ACC和RMSE与IFS相当
Pangu-Weather	Swin Transformer	192*V100*15天	7天内0.25°空间分辨率预报的ACC和RMSE优于IFS
ClimaX	ViT	8*V100*15 h/变量	14天内1.40625°和5.625°空间分辨率预报的ACC和RMSE与IFS相当甚至略优
GraphCast	GNN	32*TPU V4*4周	10天内0.25°空间分辨率预报的ACC和RMSE优于HRES
FengWu	Transformer	32*A100*17天	10天内0.25°空间分辨率预报的ACC和RMSE优于GraphCast
FuXi	U-Transformer	8*A100*30 h	15天内0.25°空间分辨率预报的ACC和RMSE与EPS相当
AI-GOMS	ViT+AFNO	16*A100*34.2 h	30天内0.25°空间分辨率预报的ACC和RMSE较好

### 3 总结

表2为上述模型的主要结构、训练代价以及优点,其中训练代价依据显卡数量\*显卡型号\*训练耗时排列,V100和A100指NVIDIA Tesla V100 GPU和NVIDIA Tesla A100 GPU,TPU V4为谷歌专门为深度学习设计的ASIC芯片。表2展示了近期深度学习模型在海洋和大气预报领域的发展过程,其预

报精度和计算效率的显著提升已得到实证。这些模型为通用近似定理在该领域的应用提供了有力支撑,即前馈神经网络在特定条件下能够逼近任意连续函数<sup>[26-27]</sup>。值得注意的是,这些先进模型的架构中鲜少包含显式的物理机制表述,这反映了深度学习范式的核心思想:通过海量数据驱动,自适应地学习变量间的复杂非线性关系,而人为干预主要体现在对模型监督和优化策略的调整上。这种数



据驱动的方法虽显著降低了模型开发的难度,但也带来了一系列挑战。

首要挑战是计算资源的限制。在从FourCastNet到FuXi的演进过程中,尽管训练数据主要源自ERA5再分析资料,但模型参数和输入变量的规模呈现显著增长趋势。早期模型(如FourCastNet)受限于当时GPU内存容量,数据规模相对较小。强制增加模型规模将导致批量大小(batch size)降低,进而影响训练的收敛性和最终性能。为应对这一挑战,拥有充足资源的研究机构不断升级硬件设施。例如,用于训练FuXi的GPU在内存容量和计算能力上较训练FourCastNet时提升了约2倍。然而,这种策略导致了高性能GPU的供需失衡和价格攀升,这给资源受限的研究团队造成了更大压力。因此,解决这个问题的另一个重要的技术路线是通过优化模型架构,在保持或提升性能的同时降低对计算资源的需求。

另一个挑战是纯数据驱动方法可能导致对物理机制的理解不足。这种基于统计学习的方法虽然在短期预报中表现出色,但在长期预报中可能出现显著偏差。这种偏差不同于传统模型中的误差累积,而是由于模型结构缺乏物理自洽性导致的。然而,这一特性也为非计算机专业的研究者提供了利用深度学习构建模型的机会,因为它允许研究者专注于领域知识而非复杂的算法设计。海洋与大气科学长期积累的机理性知识为深度学习模型的结构优化提供了丰富的理论基础。例如,在模型设计中可以考虑动量、能量及质量守恒原理,这不仅可以指导输入/输出变量的选择,还可以通过构建物理约束的损失函数来优化模型的学习过程<sup>[28-29]</sup>。此外,物理模型与深度学习模型的融合也是一个富有前景的研究方向。最新的NeuralGCM模型<sup>[30]</sup>在GNN架构的基础上引入了物理模块,构建了物理机制与机器学习相结合的混合预报模型。该方法不仅提高了中期预报(15天内)的技巧,还在特定条件下实现了跨季节—年代际尺度的预报能力。

在海洋模型领域,海洋与大气系统间的固有差异给深度学习模型的应用带来了独特挑战。与大气的整体连通性不同,海洋被陆地分割成不同区域,导致学习效果呈现显著的区域性特征<sup>[31]</sup>。此外,深海观测数据的稀疏性进一步增加了海洋模型的

学习难度<sup>[32]</sup>。这些因素导致海洋模型的发展更倾向于聚焦数据丰富的区域,注重中尺度—区域尺度的过程,而非全球尺度的模拟。为了应对这些挑战,研究者正在探索多种创新方法,包括但不限于整合卫星遥感、Argo浮标网络等多源异构数据,开发适应性更强的区域化模型,设计能够处理多尺度问题的深度学习架构(如借鉴并应用地球科学其他领域的多尺度卷积神经网络<sup>[33-34]</sup>或层次化注意力机制<sup>[35]</sup>)。

对海洋和大气科学研究而言,深度学习模型提供了一种新颖的研究范式。在提出科学假设后,可以选取成熟的深度学习框架作为基础,结合领域特定的理论进行定制化改进,从而构建高效的应用模型。这种方法使研究者能够专注于核心科学问题,而无需过多关注底层算法的实现细节;同时,云计算平台如Azure和Google Colab的兴起使得高性能计算资源的获取更便捷,大大加速了模型的迭代验证过程。

## 4 展望

展望未来,深度学习在海洋和大气预报领域仍有广阔的发展空间。一些值得重点关注的研究方向包括:①开发更适应海洋特性的区域化和多尺度模型,如考虑海洋锋面、涡旋等中尺度过程的专用模型架构;②探索物理信息引导的深度学习方法,如物理约束的神经常微分方程,以提高模型的可解释性和长期预报能力;③利用深度学习技术改进变分和集合数据同化方法,提高预报初始场的质量;④开发能够有效处理海洋数据稀疏性和不确定性的模型结构,如贝叶斯深度学习方法等。此外,加强海洋学、气象学和计算机科学之间的跨学科合作也至关重要,这将有助于开发更符合海洋和大气动力学特性的深度学习模型。

总而言之,深度学习为海洋和大气预报带来了前所未有的机遇和挑战。通过合理配置计算资源、深度融合领域知识、开发创新的模型架构,深度学习有望在提高预报技巧、延长预报时效以及揭示新的物理机制等方面发挥关键作用。然而,我们也需要保持审慎态度,充分认识到深度学习模型的固有局限性,在实际业务应用中不断进行验证和优化,

真正发挥深度学习在海洋和大气科学中的潜力,推动学科的持续进步。

### 参考文献:

- [1] SCHIZAS C N, MICHAELIDES S, PATTICHIS C S, et al. Artificial neural networks in forecasting minimum temperature (weather)[C]//1991 Second International Conference on Artificial Neural Networks. Bournemouth: IET, 1991: 112-114.
- [2] COMRIE A C. Comparing neural networks and regression models for ozone forecasting[J]. Journal of the Air & Waste Management Association, 1997, 47(6): 653-663.
- [3] HALL T, BROOKS H E, DOSWELL III C A. Precipitation forecasting using a neural network[J]. Weather and Forecasting, 1999, 14(3): 338-345.
- [4] DEO M C, NAIDU C S. Real time wave forecasting using neural networks[J]. Ocean Engineering, 1998, 26(3): 191-203.
- [5] GAO S, HUANG J, LI Y R, et al. A forecasting model for wave heights based on a long short-term memory neural network[J]. Acta Oceanologica Sinica, 2021, 40(1): 62-69.
- [6] ZHOU S Y, BETHEL B J, SUN W J, et al. Improving significant wave height forecasts using a joint empirical mode decomposition-long short-term memory network[J]. Journal of Marine Science and Engineering, 2021, 9(7): 744.
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 6000-6010.
- [8] PATHAK J, SUBRAMANIAN S, HARRINGTON P, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators[J]. arXiv:2202.11214.
- [9] RASP S, THUEREY N. Data - driven medium - range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench[J]. Journal of Advances in Modeling Earth Systems, 2021, 13(2): e2020MS002405.
- [10] WEYN J A, DURRAN D R, CARUANA R. Improving data - driven global weather prediction using deep convolutional neural networks on a cubed sphere[J]. Journal of Advances in Modeling Earth Systems, 2020, 12(9): e2020MS002109.
- [11] LI Z Y, KOVACHKI N B, AZIZZADENESHELI K, et al. Fourier neural operator for parametric partial differential equations[C]//9th International Conference on Learning Representations. ICLR, 2021.
- [12] GUIBAS J, MARDANI M, LI Z Y, et al. Adaptive fourier neural operators: Efficient token mixers for transformers[J]. arXiv: 2111.13587.
- [13] BI K F, XIE L X, ZHANG H H, et al. Pangu-weather: A 3d high-resolution model for fast and accurate global weather forecast[J]. arXiv:2211.02556.
- [14] BI K F, XIE L X, ZHANG H H, et al. Accurate medium-range global weather forecasting with 3D neural networks[J]. Nature, 2023, 619(7970): 533-538.
- [15] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 9992-10002.
- [16] NGUYEN T, BRANDSTETTER J, KAPOOR A, et al. ClimaX: A foundation model for weather and climate[C]//Proceedings of the 40th International Conference on Machine Learning. Honolulu: ICML, 2023: 25904-25938.
- [17] RASP S, DUEBEN P D, SCHER S, et al. WeatherBench: a benchmark data set for data-driven weather forecasting[J]. Journal of Advances in Modeling Earth Systems, 2020, 12(11): e2020MS002203.
- [18] LAM R, SANCHEZ-GONZALEZ A, WILLSON M, et al. GraphCast: Learning skillful medium-range global weather forecasting [J]. arXiv: 2212.12794.
- [19] LAM R, SANCHEZ-GONZALEZ A, WILLSON M, et al. Learning skillful medium-range global weather forecasting[J]. Science, 2023, 382(6677): 1416-1421.
- [20] CHEN K, HAN T, GONG J C, et al. FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead[J]. arXiv: 2304.02948.
- [21] KRASHINSKY R, GIROUX O, JONES S, et al. Nvidia ampere architecture in-depth[EB / OL]. NVIDIA blog, 2020. <https://devblogs.nvidia.com/nvidia-ampere-architecture-in-depth>.
- [22] CHEN L, ZHONG X H, ZHANG F, et al. FuXi: A cascade machine learning forecasting system for 15-day global weather forecast[J]. npj Climate and Atmospheric Science, 2023, 6(1): 190.
- [23] LIU Z, HU H, LIN Y T, et al. Swin transformer v2: Scaling up capacity and resolution[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11999-12009.
- [24] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Proceedings of the 18th International Conference. Munich: Springer, 2015: 234-241.
- [25] XIONG W, XIANG Y F, WU H, et al. AI-GOMS: Large AI-driven global ocean modeling system[J]. arXiv:2308.03152.
- [26] CYBENKO G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals and Systems, 1989, 2(4): 303-314.
- [27] HORNIK K. Approximation capabilities of multilayer feedforward networks[J]. Neural Networks, 1991, 4(2): 251-257.
- [28] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: A deep learning framework for solving

- forward and inverse problems involving nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2019, 378: 686-707.
- [29] DAW A, KARPATNE A, WATKINS W D, et al. Physics-guided neural networks (PGNN): An application in lake temperature modeling[M]//KARPATNE A, KANNAN R, KUMAR V. *Knowledge Guided Machine Learning*. New York: Chapman and Hall/CRC, 2022: 353-372.
- [30] KOCHKOV D, YUVAL J, LANGMORE I, et al. Neural general circulation models for weather and climate[J]. *Nature*, 2024, 632 (8027): 1060-1066.
- [31] WANG X, WANG R Z, HU N Z, et al. XiHe: A data-driven model for global ocean eddy-resolving forecasting[J]. *arXiv:2402.02995*.
- [32] LEVIN L A, BETT B J, GATES A R, et al. Global observing needs in the deep ocean[J]. *Frontiers in Marine Science*, 2019, 6: 241.
- [33] VANDAL T, KODRA E, GANGULY S, et al. DeepSD: Generating high resolution climate change projections through single image super-resolution[C]//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax: Association for Computing Machinery, 2017: 1663-1672.
- [34] PAN B X, HSU K, AGHAKOUCHAK A, et al. Improving precipitation estimation using convolutional neural network[J]. *Water Resources Research*, 2019, 55(3): 2301-2321.
- [35] LIANG Y X, KE S Y, ZHANG J B, et al. GeoMAN: Multi-level attention networks for geo-sensory time series prediction[C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm: AAAI Press, 2018: 3428-3434.

## Recent developments in large model research in the ocean and atmospheric fields

ZHANG Ziliang<sup>1</sup>, YU Huaming<sup>1,2</sup>, REN Shutong<sup>3,4\*</sup>, ZHANG Chenyu<sup>5</sup>

(1. College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China; 2. Sanya Oceanographic Institution, Ocean University of China, Sanya 572025, China; 3. National Marine Environmental Forecasting Center, Beijing 100081, China; 4. Key Laboratory of Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Ministry of Natural Resources, Beijing 100081, China; 5. Qingdao Ekman Technology Co., Ltd, Qingdao 266100, China)

**Abstract:** This review comprehensively examines the major deep learning models applied to global-scale ocean and atmospheric forecasting in recent years, with a particular emphasis on autoregressive models based on the Transformer architecture. We focus on analyzing models such as FourCastNet, Pangu-Weather, ClimaX, GraphCast, FengWu, FuXi, and AI-GOMS, chronologically detailing their fundamental architectures, training data, and forecasting performance. Through a comparative analysis of these models, this paper elucidates the primary trends in model development: from simple Transformer structures to the incorporation of diverse architectures such as AFNO, Swin Transformer, and GNN, demonstrating continuous structural innovation. Concurrently, models have been optimized in terms of input variable selection, training strategies, and computational efficiency to enhance forecast accuracy and practical applicability. Looking ahead, with advancements in GPU (Graphics Processing Unit) technology and strengthened interdisciplinary collaboration, we anticipate further breakthroughs in improving forecasting skills, extending prediction lead times, and uncovering new physical mechanisms, while simultaneously focusing on enhancing computational efficiency and reducing resource requirements.

**Key words:** deep learning; numerical prediction; Transformer; large model