

马强, 颜京辉, 魏敏, 等. 北京气候中心 CMIP6 数据共享平台及应用. 应用气象学报, 2022, 33(5): 617-627.

DOI: 10.11898/1001-7313.20220509

北京气候中心 CMIP6 数据共享平台及应用

马强¹⁾²⁾ 颜京辉³⁾⁴⁾ 魏敏³⁾⁴⁾* 辛晓歌³⁾⁴⁾
张莉³⁾⁴⁾ 张芳³⁾⁴⁾ 吴统文³⁾⁴⁾

¹⁾(国家气象信息中心, 北京 100081)

²⁾(清华大学计算机科学与技术系, 北京 100084)

³⁾(中国气象局地球系统数值预报中心, 北京 100081)

⁴⁾(中国气象科学研究院灾害天气国家重点实验室, 北京 100081)

摘 要

为保障北京气候中心(Beijing Climate Center, BCC)气候模式在第 6 次耦合模式比较计划(Coupled Model Intercomparison Project Phase 6, CMIP6)中的大量试验数据产品面向国内外实现共享, 建立了试验数据共享平台。由于模式试验数据具有数据量大、要素种类繁多、元数据多样等特征, 为提供高效的数据管理, 平台采用分布式存储架构, 数据通过气候模式输出重写(climate model output rewriter, CMOR)软件进行格式规范, 并实现基于 THREDDS(thematic real-time environmental distributed data services)的数据组织与共享。在平台建设及软件设计部署等层面, 充分考虑数据安全。该平台实现 BCC 3 个模式约 190 TB 的试验数据稳定、高效共享, 为国内外气候变化领域科研工作者提供获取数据的方便快捷途径与方法, 成为推动我国气候模式国际应用的有力技术手段。

关键词: CMIP; BCC; 模式试验数据; 数据共享; 数据安全

引 言

20 世纪 90 年代, 世界气候研究计划(World Climate Research Programme, WCRP)提出并组织了耦合模式比较计划(Coupled Model Intercomparison Project, CMIP), 该计划通过设计气候模式试验, 制定试验标准, 收集、共享来自世界不同国家模式工作组的气候模式试验数据, 开展多模式相互比较和评估工作。参与该计划的试验数据被广泛应用于气候变化机理研究与未来气候变化预估等相关领域, 为政府间气候变化专门委员会(Intergovernmental Panel on Climate Change, IPCC)定期发布评估报告提供技术支撑, 也为气候模式发展提供重要平台^[1]。WCRP 先后组织了 6 次国际耦合模式比较计划, 目前正在进行的是 CMIP6^[2-3]。CMIP6 科学试验设计

包括 2 个核心试验和 23 个模式比较子计划, 试验种类较 CMIP5 明显增加。参加比较计划的气候模式由早期的海-气耦合模式发展到对大气、陆面、海洋及海冰等多圈层进行模拟的耦合模式, 以及包含碳循环过程的地球系统模式, 模式分辨率越来越精细^[4]。多种因素综合导致试验数据量激增, 从 CMIP1 约 1 GB, 到 CMIP2 超过 500 GB, CMIP3 超过 35 TB, CMIP5 超过 3.5 PB, CMIP6 产生的数据量更加巨大。如何在全球范围广泛、高效地共享这些海量试验数据, 满足气候变化及相关领域研究的需求, 成为亟需解决的关键技术问题。

2011 年美国在原有地球系统网格(Earth System Grid, ESG)项目基础上, 由美国能源部(Department of Energy, DOE)、美国航空航天局(National Aeronautics and Space Administration, NASA)、美国国家海洋和大气管理局(National Oceanic and

2022-03-31 收到, 2022-07-01 收到再改稿。

资助项目: 国家重点研发计划(2016YFA0602100)

* 通信作者, 邮箱: weim@cma.gov.cn

Atmospheric Administration, NOAA) 等多家单位共同资助,多个国家共同参与建立了地球系统网格联盟(Earth System Grid Federation, ESGF),自 CMIP5 开始负责模式试验数据的管理与共享^[5]。ESGF 是由多个分布式节点组成的全球系统,节点分为数据节点、身份认证节点、索引节点和计算节点 4 种类型,参加 CMIP 模式工作组可根据需要进行部署。数据共享平台的建设成为参与 CMIP 的重要环节。

作为国家级业务中心,北京气候中心(BCC)参

加了 CMIP5 和 CMIP6 两个阶段的比较计划^[6-7]。为了满足数据共享应用需求,BCC 建设试验数据共享平台,部署 BCC ESGF 数据节点。

1 数据特征

BCC 选用近年研发的 3 个不同版本模式参加 CMIP6,包括地球系统模式 BCC-ESM1.0、中等分辨率气候系统模式 BCC-CSM2-MR 和高分辨率气候系统模式 BCC-CSM2-HR,模式信息见表 1。

表 1 BCC 参加 CMIP6 试验的模式版本

Table 1 BCC model versions participated in CMIP6

模式	分量	模式版本	分辨率
BCC-ESM1.0 ^[8]	大气	BCC-AGCM3-Chem	T42L26(约为 280 km,水平格点数为 128×64,垂直分 26 层,模式顶为 2.19 hPa)
	陆面	BCC-AVIM2	T42(约为 280 km,水平格点数为 128×64)
	海洋	MOM4-L40v3	gx1v1(纬向分辨率为 1°,经向在 10°S~10°N 加密到(1/3)°,10°S~30°S 和 10°N~30°N 由(1/3)°逐渐过渡到 1°,30°S 以南、30°N 以北区域为 1°,水平格点数为 360×232,垂直分 40 层)
	海冰	SIS	gx1v1(水平分辨率与 MOM4-L40v3 相同)
BCC-CSM2-MR ^[9]	大气	BCC-AGCM3-MR	T106L46(约为 110 km,水平格点数为 320×160,垂直分 46 层,模式顶为 1.46 hPa)
	陆面	BCC-AVIM2	T106(约为 110 km,水平格点数为 320×160)
	海洋	MOM4-L40v3	同 BCC-ESM1.0
	海冰	SIS	同 BCC-ESM1.0
BCC-CSM2-HR ^[10]	大气	BCC-AGCM3-HR	T266L56(约为 45 km,水平格点数为 800×400,垂直分 56 层,模式顶为 0.1 hPa)
	陆面	BCC-AVIM2	T266(约为 45 km,水平格点数为 800×400)
	海洋	MOM5-L50	0.25°×0.25°(水平格点数为 1440×688,垂直分 50 层)
	海冰	SIS	0.25°×0.25°(水平格点数为 1440×688)

BCC-ESM1.0 与 BCC-CSM2-MR 完成了气候诊断、评估和描述试验(Diagnostic, Evaluation and Characterization of Klima experiments, DECK)与历史气候模拟试验(Historical)^[2]两个核心试验。在此基础上,BCC-ESM1.0 开展了气溶胶和化学模式比较计划(Aerosols and Chemistry Model Intercomparison Project, AerChemMIP)^[11], BCC-CSM2-HR 参与了高分辨率模式比较计划(High-Resolution Model Intercomparison Project, High-ResMIP)^[12],BCC-CSM2-MR 完成了包括年代际气候预测计划(Decadal Climate Prediction Project, DCP)^[13-14]在内的其他 8 个子计划。

BCC 模式开展核心试验共需完成超过 2500 年的积分,计算输出近 2000 个模式要素,完成各子计划需进行近万年积分,计算输出超过 6000 个模式要素。按照 CMIP6 要求,模式数据均采用标准 NetCDF 格式存储,水平分辨率与模式保持一致,垂直分辨率按照试验要求进行插值,时间分辨率根据试验要求不同,分为逐小时、3 h、日及月等。以应用最多

的 BCC-CSM2-MR 模式为例,该模式为大气、陆面、海洋和海冰多圈层耦合模式,仅大气模式单时次、单要素场格点就包含约 235 万个,多圈层格点之和则达到约 583 万个。历史气候模拟试验(Historical)分为 3 个集合样本,每个样本积分 165 年,分别按照 3 h,6 h,日及月不同频率输出近 200 个要素,数据量约为 7.4 TB。据测算 BCC CMIP6 试验的数据量共计 190 TB。由此可知试验具有配置复杂、模式分辨率高、要素种类多、数据输出频次多样及作业积分时间长等特点。

由于参加比较计划的模式来自不同国家模式工作组,模式数据特征各不相同,为便于进行统一管理与规范应用,需对模式原始输出数据进行加工处理,采用通用标准格式存储。数据管理与存储处理逻辑非常复杂,工作量巨大,数据共享工作极具挑战性。

2 平台架构

为满足 BCC 试验数据存储及共享服务需求,从

数据规模,服务效率、访问安全性等多方面考虑,共享平台采用多层体系架构,采用组件化设计,保障系统灵活的可扩展性和可维护性,采用数据库与文件系统结合,进行高效的数据存储管理和统一的数据

支撑服务,采用 http, Globus 等通信协议及 Web-Service 等技术,确保数据传输和共享服务的准确与高效。平台系统总体框架由系统层、数据层、服务层和用户层 4 部分组成(如图 1 所示)。



图 1 平台系统架构示意图

Fig. 1 Platform system architecture

系统层包括硬件和软件两部分,硬件主要包括高性能专用服务器和 1 套具有高速读写访问能力的可扩展分布式存储设备。存储系统采用全对称分布式架构,可横向平滑扩展文件存储,包含 7 个存储节点,提供约 900 TB 可用存储容量。平台采用的分布式数据存储系统在可扩展性、并发性能等方面克服了传统本地存储扩容步骤复杂、性能受单一控制器能力限制的缺陷。同时,为服务器配置轻量级备份虚拟机,以提升持续服务能力。软件包括支撑数据处理和展示的 Fortran, Python, NCL 等程序语言和工具软件,支持通过高级程序语言调用函数库对数据处理,支持在命令行实现对数据的集合操作及绘图等功能^[15]。

数据层存储平台所涉及的所有数据,包括模式原始计算数据、共享的数据产品及产品元数据等。服务层指运行在服务器的各种服务,包括 Web 服务、OPeNDAP 服务、Globus Gridftp 服务、OpenID 身份认证服务等。Web 服务提供互联网访问服务,OPeNDAP 服务通过 http 通信协议实现数据访问请求与响应,基于 URL 链接和 Web 服务器即可实

现对数据的分发与共享, Globus Gridftp 服务为共享数据的下载提供高性能、安全可靠的数据传输, OpenID 服务可为访问平台的用户提供便捷的单点登录功能。用户层主要为用户提供平台交互访问客户端,为用户提供多种条件组合查询界面、元数据展示、wget 和 Globus 等多种数据下载服务与数据说明文档等。

平台部署在中国气象局互联网隔离区(demilitarized zone, DMZ),既可实现互联网的公开访问,又可获得加强的网络安全防护^[16]。平台通过中国科技网(出口带宽为 1.35 Gbps)对外部用户提供服务,并实现与其他多个国家或组织的 ESGF 节点高速互联通信,气象局内网用户可通过万兆以太网访问。

3 平台实现

3.1 建设流程

平台建设包括数据收集、数据处理、数据存储发布及数据服务几个关键部分(图 2),数据安全贯穿全流程。

首先对 BCC 模式计划开展的试验进行分析,预估可能产生的数据量及后续工作任务,确定硬件性能及功能需求^[17]。硬件建设主要包括服务器及配套存储设施选型、采购及安装测试等。在此基础

上,对相应软件环境进行部署,主要包括系统软件、开发运行环境、应用软件的安装配置。建设过程中严格按照安全规范执行,在平台正式上线前通过系统安全基线检查、软件漏洞扫描等安全检测。

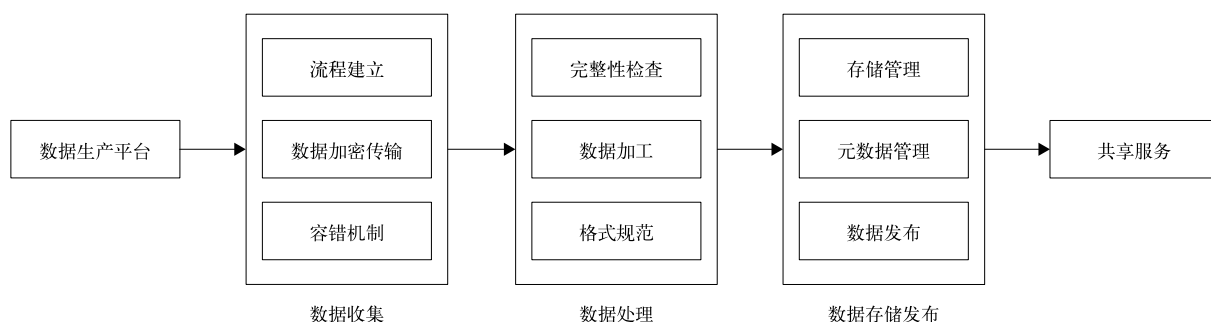


图 2 平台建设工作流程

Fig. 2 Platform construction workflow

数据收集模块实现自动化传输流程,采用加密传输策略,将在数据生产平台-高性能计算系统产生的模式试验数据收集到共享平台。数据处理模块对模式原始输出数据进行检查、处理并规范化数据格式^[18]。数据存储发布模块确定试验数据在线存储规则,建立目录结构及访问权限,完成数据发布。数据共享服务门户提供易用高效的数据检索界面,用户经过身份认证后,可通过多种方式批量下载检索结果。

3.2 数据质量控制及规范化

模式试验在高性能计算系统开展,模式程序并行规模大,运行时间长,计算过程可能受到内存争用、通信抖动或文件读写干扰等多方面因素影响,增加模式计算结果的不稳定性。确保每个试验数据的完整性和可用性是数据处理工作的前提,需要在每个试验完成后,对生成的试验数据文件数量、文件大小及是否存在计算异常等情况进行严格检查,对存在问题的试验,需要重新计算。针对各试验配置特点,结合模式运行情况,设计并实现数据检查流程,提高自动化水平,减少人为干预,提升问题筛查精度,从信息技术角度保障数据质量^[19]。

模式模拟效果合理性检查也是必不可少的重要环节。BCC CMIP6 模式均为多圈层耦合模式,每个试验完成后需对大气、海洋等主要分量模式关键要素模拟结果从全球、典型区域等空间属性,多年平均、季节平均等时间属性的物理意义合理性进行检查^[20]。如大气分量主要考查地表温度、降水、风场

等要素在全球及东亚等不同区域气候态及随时间的演变规律。海洋分量主要考虑海表及中上层海温基本气候态,全球、赤道太平洋等关键区域海表温度不同时间尺度变化等。检查方法包括空间分布合理性检查、时间序列合理性检查、气候态检查、极值统计及误差分析等。

参加 CMIP 的模式来自世界不同国家模式工作组,模式原始输出数据在数据组织、数据格式及元数据信息等方面存在较大差异,为减少对不同模式同一试验数据预处理的工作量,CMIP 给出较为严格的数据规范及明确的数据属性要求。因此,在严格的技术框架内结合 BCC CMIP6 数据特征与共享平台实际情况,实现数据规范化是十分关键的技术问题。模式输出结果采用气候模式输出重写 (climate model output rewriter, CMOR) 软件^[21]进行格式标准化,确保生成符合气候和预测元数据公约的 NetCDF 文件。经过 CMOR 软件处理,实现将同一试验相同要素的多个时间记录存储在 1 个或一系列文件中,将模式垂直层输出的要素插值到标准等压面,依据给定的经纬度排列顺序对数据进行排序,对要素量纲进行转换等。数据属性主要包含模式工作组提供的模式配置信息、CMIP6 各模式及其组织机构信息和模式要素信息等内容,以 JSON 格式存储为 MIP 表文件^[22],与模式原始输出数据文件共同作为 CMOR 软件的输入和输出数据的元数据。CMOR 输出的数据采用 CMIP 的统一规则命名。

数据处理流程如图 3 所示。

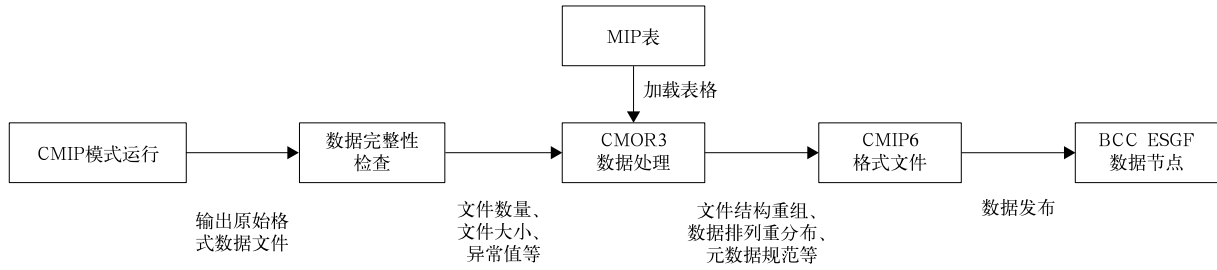


图 3 数据处理流程

Fig. 3 Data processing workflow

3.3 数据发布与服务

CMIP6 试验数据具有数据量大、数据文件离散但命名规则统一及元数据多维等特征^[23]。平台采用专题实时环境数据分发系统(thematic real-time environmental distributed data services, THREDDS)提供数据服务(THREDDS data server, TDS),支持 http, Girdftp, OPeNDAP 和 OGC WMS/WCS 等多种数据访问接口。THREDDS 是一个面向服务体系结构(service oriented architecture, SOA)的数据服务系统,实现试验数据组织、元数据管理和数据发布的功能。THREDDS 提取数据文件的元数据信息,用于支撑数据访问接口以 Web 服务形式提供用户访问。其中 http, Girdftp 提供完整文件下载功能,OPeNDAP 支持数据空间、时间、要素等范围子集的选取功能,

OGC WMS/WCS 提供 GIS 访问接口。

目录结构设计对于 THREDDS 非常重要,CMIP6 试验数据存储目录采用具有自描述信息的层级管理结构,实现不同试验不同要素数据分级分类存储。在试验数据发布过程中,THREDDS 遍历扫描所选数据集,自动分析提取要素变量、时空分辨率等多种元数据信息,保持试验数据及元数据的一致性和准确性^[24]。将元数据存储到本地 PostgreSQL 数据库相应数据表中,生成对应的数据集元数据描述文件 catalog.xml,并纳入 TDS 提供服务,用户即可通过 URL 方式对多层次的目录进行访问,将数据下载至本地或者通过客户端直接读取访问所需要的数据。同时,元数据信息发布到 ESGF 索引节点,实现 CMIP6 试验数据统一检索,如图 4 所示。

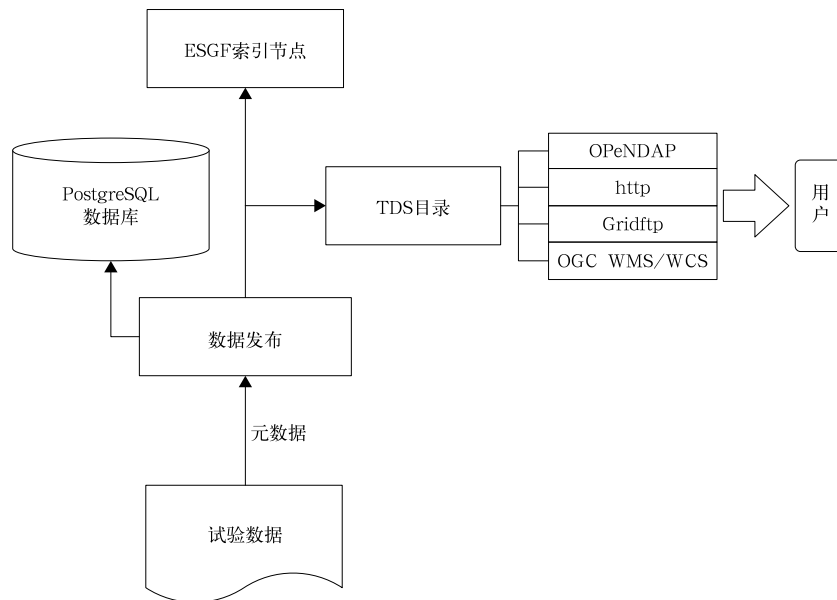


图 4 BCC CMIP6 数据发布与服务

Fig. 4 BCC CMIP6 data publishing and service

试验数据通过 ESGF 索引节点以 Web 门户方式提供共享服务,主要包含 CMIP6 模式及其试验数据的详细描述、数据检索及获取等服务。数据检索

功能支持试验类型、模式信息、组织机构、数据要素、集合样本等多种条件组合检索,根据需求可从结果中进一步检索所需时间序列的数据子集,为用户使

用提供极大的便利。平台收到数据节点响应索引节点请求,为用户提供数据浏览以及基于 http, Globus, OPeNDAP 等协议的数据获取服务。平台依据数据检索结果为用户自动生成 Bash 或 Python 格式脚本,用户登录认证通过后即可实现批量数据文件下载。

3.4 数据安全

平台面向世界各国 ESGF 用户,通过互联网对外提供数据共享服务,保障数据安全极其重要。

物理层面上,通过增加物理介质的副本存储等方式保护存储在分布式存储系统上的试验数据的物理安全,采用轻量级备份虚拟机保障服务的连续性,确保数据不丢失、服务不中断。同时,BCC CMIP6 数据同步到澳大利亚、美国和英国等其他国家的 ESGF 数据节点,实现数据异地备份,提升共享服务能力,用户可选择最优站点进行数据下载。

系统层面上,系统及应用软件参照 ESGF 软件安全规范部署,并采用密码增强、系统加固、访问控制、日志审计等多种技术方法对平台进行基础的系统安全防护。

网络层面上,平台部署在互联网 DMZ,纳入中国气象局国家级的整体网络安全防护体系,通过防火墙、抗 DDOS、入侵检测、入侵防御、Web 应用防护、安全威胁和态势感知等一系列安全防护手段保障网络安全。

对于模式试验数据自身的安全,从数据处理、收集、发布、共享服务等各个环节采取安全管控措施,避免数据的非授权访问,保护数据不被篡改和伪造。在数据处理过程中,利用高性能计算机安全管理体系^[25],针对数据加工处理过程进行严格授权,仅允许特定用户进行相关操作,控制输入、输出和中间结果数据文件的访问权限,确保数据来源可信。在数据收集过程中,采用 SFTP 加密协议防止用户名、口令等敏感信息泄露,确保数据不被非法访问,并通

过校验机制确保数据的完整性。

数据发布前,BCC 需向 ESGF 内部 CA(certification authority)证书颁发机构申请认证审查,获取并在 BCC 数据节点安装其签署的主机证书,获得数据发布授权。主机证书符合 RFC3280 定义的 X.509 规范,使用 4096 位密钥,用于数据节点管理程序、数据发布程序等客户端应用的安全认证。发布过程中对外提供检索服务的元数据信息与实际的模式数据文件分别进行存储管理,并结合系统层和应用层的访问控制设置访问权限。发布的元数据存储在数据库中,并同步发布到索引节点,模式数据文件存储在本地安全可靠的分布式存储系统上^[26]。

平台针对发布后的数据共享服务进行授权管理,仅允许通过认证的用户进行批量数据访问和下载操作。用户通过浏览器检索访问 BCC CMIP6 数据时,需要通过 OpenID(open identification)进行身份认证。利用去中心化的 OpenID 用户身份认证协议,用户只需在 ESGF 的某一个身份认证节点注册获取 OpenID 帐号,就可以凭此帐号登录并获取数据,在加强数据应用安全的同时兼顾用户便捷访问的需求。

4 平台应用

4.1 模式产品

BCC CMIP6 数据共享平台自 2017 年开始建设,2018 年正式开始发布 BCC CMIP6 数据,截至 2021 年共计发布数据 190 TB,实现 BCC CMIP6 全部数据共享。模式试验数据包括 DECK 和 Historical 2 个核心试验,以及 ScenarioMIP, C4MIP 等 10 个子计划,共计 50 多个试验,输出模拟结果包括约 8000 个要素,涵盖了对全球开展的中等分辨率、高分辨率多圈层模拟及大气化学模拟数据,各试验集合样本量和模拟结果要素数量有所差异,如表 2 所示。

表 2 BCC CMIP6 数据

Table 2 BCC CMIP6 data

核心试验/子计划	试验名称	模式	样本量	要素数量	数据量/TB
气候诊断、评估和描述试验 (DECK) ^[2]	1pctCO2	BCC-CSM2-MR	1	142	29.00
		BCC-ESM1.0	1	179	
	abrupt-4xCO2	BCC-CSM2-MR	1	142	
		BCC-ESM1.0	1	169	
	amip	BCC-CSM2-MR	3	97	
		BCC-ESM1.0	3	134	
	esm-hist	BCC-CSM2-MR	3	155	
	esm-piControl	BCC-CSM2-MR	1	143	

续表 2

核心试验/子计划	试验名称	模式	样本量	要素数量	数据量/TB
气候诊断、评估和描述试验(DECK) ^[2]	piControl	BCC-CSM2-MR	1	142	32.00
		BCC-ESM1.0	1	168	
历史气候模拟试验(Historical) ^[2]	Historical	BCC-CSM2-MR	3	154	6.50
		BCC-ESM1.0	3	189	
检测归因模式比较计划(DAMIP) ^[27]	hist-GHG	BCC-CSM2-MR	3	149	26.00
	hist-aer		3	144	
	hist-nat		3	145	
情景模式比较计划(ScenarioMIP) ^[28]	SSP1-2.6	BCC-CSM2-MR	1	154	7.40
	SSP2-4.5		1	153	
	SSP3-7.0		1	158	
	SSP5-8.5		1	154	
耦合气候碳循环比较计划(C4MIP) ^[29]	1pctCO2-bgc	BCC-CSM2-MR	1	144	2.50
	1pctCO2-rad		1	144	
	esm-ssp585		1	155	
全球季风模式比较计划(GMMIP) ^[30]	amip-hist	BCC-CSM2-MR	1	74	29.00
	hist-resAMO		1	130	
云反馈模式比较计划(CFMIP) ^[31]	amip	BCC-CSM2-MR	1	113	0.01
	amip-4xCO2		1	115	
	amip-future4K		1	114	
	amip-m4K		1	114	
	amip-p4k		1	118	
陆面、雪和土壤湿度模式比较计划(LS3MIP) ^[32]	Land-Hist-princeton	BCC-CSM2-MR	1	40	2.00
土地利用模式比较计划(LUMIP) ^[33]	deforest-globe	BCC-CSM2-MR	1	147	4.10
	esm-ssp585-sspl26Lu		1	147	
	hist-nolu		1	146	
	land-hist		1	40	
	land-nolu		1	40	
	sspl26-ssp370Lu		1	147	
	ssp370-sspl26Lu		1	147	
气溶胶和化学模式比较计划(AerChemMIP) ^[11]	hist-piAer	BCC-ESM1.0	3	186	9.07
	hist-piNTCF		3	179	
	histSST		1	121	
	histSST-piCH4		1	120	
	histSST-piNTCF		1	120	
	piClim-BC		1	123	
	piClim-CH4		1	119	
	piClim-NOx		1	123	
	piClim-NTCF		1	120	
	piClim-O3		1	123	
	piClim-SO2		1	123	
	piClim-VOC		1	123	
	piClim-aer		1	123	
	piClim-control		1	119	
	ssp370		3	187	
	ssp370-lowNTCF		3	182	
ssp370SST	1	121			
ssp370SST-lowNTCF	1	121			
年代际气候预测计划(DCPP) ^[13]	deppA-hindcast	BCC-CSM2-MR	8	74	70.00
高分辨率模式比较计划(HighResMIP) ^[12]	control-1950	BCC-CSM2-HR	1	113	70.00
	highresSST-present		1	97	
	hist-1950		1	130	

模式试验大部分利用耦合模式完成,试验输出要素分别由大气、陆面、海洋及海冰多个圈层的分量模式生成,试验数据集内容丰富,要素种类多,时间序列长,包含地球表面及大气层多种科学数据,对地球系统科学研究具有重要价值。数据均采用国际通用标准 NetCDF 格式存储,每个试验数据集设置国

际通用 DOI(数字对象唯一标识符),有利于数据长期保存和有效管理,同时也体现数据权威性,对数据提供者和制作者知识产权进行有效保护。为便于推广使用,对每个试验数据集提供说明文档,以 BCC-CSM2-MR 模式完成的历史气候模拟试验数据为例,具体信息如表 3 所示。

表 3 BCC-CSM2-MR 模式历史气候模拟试验数据

Table 3 Historical experiment data of BCC-CSM2-MR

模式	分量模式	要素	格式	时间段	时效	分辨率
BCC-CSM2-MR	大气	近地表气温、地表气压、降水等	NetCDF	1850—2014 年	月、日、3 h 等	320×160, L46, L19 等
	陆面	土壤总含水量、地表径流等	NetCDF	1850—2014 年	月、日、3 h 等	320×160
	海洋	海表面温度、海水质量、海表面压力等	NetCDF	1850—2014 年	月、日	360×232, L40
	海冰	海冰厚度、海冰面积、海冰表面温度等	NetCDF	1850—2014 年	月、日	360×232

4.2 应用成效

BCC CMIP6 数据共享平台投入运行以来,两次存储容量扩充均未中断服务,同步提升整体吞吐和并发能力,有力支撑平台的存储发布和服务,已为来自亚洲、欧洲及美洲等多个国家和地区的用户提供数据服务,取得显著成效。以 2022 年上半年为例,中国、韩国、新加坡等亚洲国家数据下载量达到 32.37 TB,位居各大洲之首,美国、加拿大等北美洲国家数据下载量约为 9.70 TB,位居第二,英国、挪威、西班牙等欧洲国家数据下载量约为 7.63 TB, BCC CMIP6 数据共享平台下载数据量共计 50.97 TB,各大洲访问共享平台情况如图 5 所示。BCC CMIP6 数据共享平台的稳定运行为全球科研工作者提供获取 BCC 试验数据的方便快捷途径与方法,降低获取试验数据的难度与成本,成为推动我国气候模式国际应用的有力技术手段。

BCC CMIP6 数据共享平台为国内外气候变化及相关领域的研究提供数据支撑,不同国家科研人员利用平台获得的试验数据开展科学研究工作,在气候变化模拟与预估^[34]、全球变暖与人类活动^[35]及模式评估改进^[36]等领域取得成果。2020 年以来,国际公开发表的气候变化相关领域研究引用 BCC CMIP6 数据 60 余次(不完全统计),在参与 CMIP6 的近 200 个模式中位居前列^[37]。2021 年 8 月 IPCC 正式发布 IPCC 第 6 次评估报告第 1 工作组报告,该报告的多个章节引用 BCC 模式多个试验数据集^[38],主要涉及人类活动对气候系统的影响,全球碳和其他生物地球化学循环与反馈及地球的能量收

支、气候反馈和气候敏感性等方面科学研究。

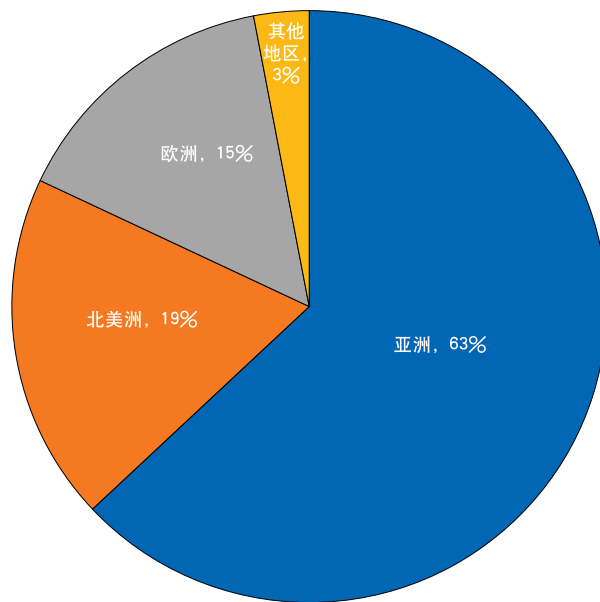


图 5 2022 年上半年 BCC CMIP6 数据访问统计

Fig. 5 BCC CMIP6 data access statistics in the first half of 2022

我国科学家也利用平台共享数据开展气候变化研究工作,如基于 BCC-CSM2-MR 模式历史气候模拟试验与情景模式比较计划等数据开展西南暴雨洪涝灾害风险预估^[39]、青藏高原夏季水汽长期变化趋势分析^[40]与黄河流域生态径流影响预估^[41]等研究工作,有助于揭示我国主要区域及典型气候事件的变化规律,推动气候变化与水资源、生态安全等领域关系研究,促进我国地球系统模式的评估与改进。

5 小结

分析表明:

1) 基于 BCC CMIP6 数据特征分析,实现完整性检查与标准化处理。采用分布式存储架构,设计并实现分层级的数据共享平台,基于专题实时环境数据分发系统提供数据服务,实现存储管理和共享发布,并在系统、应用及网络等方面采取措施保障数据安全。BCC CMIP6 数据共享平台为我国国家级业务中心参与国际 CMIP6 比较计划提供了基础技术支持。

2) BCC CMIP6 数据共享平台为国内外用户提供持续稳定的数据共享服务,BCC CMIP6 数据应用于气候变化模拟与预估、全球变暖与人类活动、地球系统模式评估与改进等研究领域,支撑国内外多项科研成果,多个数据集被 IPCC 第 6 次评估报告引用,有效提升我国气候模式的国际影响力,推动气候变化研究工作的开展。

未来气候模式将向更高分辨率、更高输出频次、更多集成员等方向发展,CMIP 试验设计将更为复杂、灵活,试验数据量将迅速增长,这些变化对共享平台软硬件的高需求也将更突显。今后将对 CMIP 提供持续数据服务,并从基础设施及技术集约化考虑,将扩展平台功能,为其他相关模式比较计划提供统一支撑。

参考文献

- [1] 周天军,邹立维,吴波,等.中国地球气候系统模式研究进展:CMIP 计划实施近 20 年回顾.气象学报,2014,72(5):892-907.
Zhou T J, Zou L W, Wu B, et al. Development of earth/climate system models in China: A review from the Coupled Model Intercomparison Project perspective. *Acta Meteor Sinica*, 2014, 72(5): 892-907.
- [2] Eyring V, Bony S, Meehl G A, et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev*, 2016, 9: 1937-1958.
- [3] 赵宗慈,罗勇,黄建斌. CMIP6 的设计. 气候变化研究进展, 2016, 12(3): 258-260.
Zhao Z C, Luo Y, Huang J B. Design of CMIP6. *Climate Change Research*, 2016, 12(3): 258-260.
- [4] 赵宗慈,罗勇,黄建斌.从检验 CMIP5 气候模式看 CMIP6 地球系统模式的发展. 气候变化研究进展, 2018, 14(6): 643-648.
Zhao Z C, Luo Y, Huang J B. The detection of the CMIP5 climate model to see the development of CMIP6 Earth system models. *Climate Change Research*, 2018, 14(6): 634-648.
- [5] ESGF Home Page. [2019-02-16]. <https://esgf-node.llnl.gov/>.
- [6] 辛晓歌,吴统文,张洁. BCC 气候系统模式开展的 CMIP5 试验介绍. 气候变化研究进展, 2012, 8(5): 378-382.
Xin X G, Wu T W, Zhang J. Introduction of CMIP5 Experiments carried out by BCC Climate System Model. *Climate Change Research*, 2012, 8(5): 378-382.
- [7] 辛晓歌,吴统文,张洁,等. BCC 模式及其开展的 CMIP6 试验介绍. 气候变化研究进展, 2019, 15(5): 533-539.
Xin X G, Wu T W, Zhang J, et al. Introduction of BCC models and its participation in CMIP6. *Climate Change Research*, 2019, 15(5): 533-539.
- [8] Wu T, Zhang F, Zhang J, et al. Beijing Climate Center Earth System Model version 1 (BCC-ESM1): Model description and evaluation of aerosol simulations. *Geosci Model Dev*, 2020, 13: 977-1005.
- [9] Wu T W, Lu Y X, Fang Y J, et al. The Beijing Climate Center Climate System Model (BCC-CSM): Main progress from CMIP5 to CMIP6. *Geosci Model Dev*, 2019, 12: 1573-1600.
- [10] Wu T, Yu R, Lu Y, et al. BCC-CSM2-HR: A high-resolution version of the Beijing Climate Center Climate System Model. *Geosci Model Dev*, 2021, 14: 2977-3006.
- [11] Collins W J, Lamarque J-F, Schulz M, et al. AerChemMIP: Quantifying the effects of chemistry and aerosols in CMIP6. *Geosci Model Dev*, 2017, 10: 585-607.
- [12] Haarsma R J, Roberts M J, Vidale P L, et al. High Resolution Model Intercomparison Project (HithResMIP V1.0) for CMIP6. *Geosci Model Dev*, 2016, 9: 4185-4208.
- [13] Boer G J, Smith D M, Cassou C, et al. The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geosci Model Dev*, 2016, 9: 3751-3777.
- [14] Wei M, Li Q Q, Xin X G, et al. Improved decadal climate prediction in the North Atlantic using EnOI-assimilated initial condition. *Sci Bull*, 2017, 62(16): 1142-1147.
- [15] 米前川,高西宁,李玥,等.深度学习应用在干旱预测中的应用.应用气象学报,2022,33(1):104-114.
Mi Q C, Gao X N, Li Y, et al. Application of deep learning method to drought prediction. *J Appl Meteor Sci*, 2022, 33(1): 104-114.
- [16] 邱彬,谭彬.安全域划分研究与应用.计算机安全,2012(6):39-42.
Qiu L, Tan B. Research and application of security domain. *Comput Secur*, 2012(6): 39-42.
- [17] 张正秋,祝从文,苏京志,等.气候动力诊断和分析系统设计与应用.应用气象学报,2021,32(5):542-552.
Zhang Z Q, Zhu C W, Su J Z, et al. Designing and implementation of Climate Dynamic Diagnosis and Analysis System. *J Appl Meteor Sci*, 2021, 32(5): 542-552.
- [18] 刘娜,熊安元,张强,等.强对流天气人工智能应用训练基础数据集构建.应用气象学报,2021,32(5):530-541.
Liu N, Xiong A Y, Zhang Q, et al. Development of basic dataset of severe convective weather for artificial intelligence

- training. *J Appl Meteor Sci*, 2021, 32(5): 530-541.
- [19] 郭启云, 杨荣康, 程凯琪, 等. 基于探空观测的多源掩星折射率质量控制及对比. *应用气象学报*, 2020, 31(1): 13-26.
Guo Q Y, Yang R K, Cheng K Q, et al. Refractive index quality control and comparative analysis of multi-source occultation based on sounding observation. *J Appl Meteor Sci*, 2020, 31(1): 13-26.
- [20] 董晓云, 余锦华, 梁信忠, 等. CWRf 模式极端降水模拟误差订正. *应用气象学报*, 2020, 31(4): 504-512.
Dong X Y, Yu J H, Liang X Z, et al. Bias correction of summer extreme precipitation simulated by CWRf model. *J Appl Meteor Sci*, 2020, 31(4): 504-512.
- [21] CMOR. [2021-01-08]. <https://cmor.llnl.gov/pdf>.
- [22] CMIP6 Table Excerpt. [2022-02-01]. https://cmor.llnl.gov/mydoc_cmip6_table_excerpt.
- [23] 丁一汇, 李霄, 李巧萍. 气候变暖背景下中国地面风速变化研究进展. *应用气象学报*, 2020, 31(1): 1-12.
Ding Y H, Li X, Li Q P. Advances of surface wind speed changes over China under global warming. *J Appl Meteor Sci*, 2020, 31(1): 1-12.
- [24] 成菲, 李巧萍, 沈新勇, 等. BCC-CSM1. 1m 对欧亚积雪覆盖的预测评估. *应用气象学报*, 2021, 32(5): 553-566.
Cheng F, Li Q P, Shen X Y, et al. Evaluation of Eurasian snow cover fraction prediction based on BCC-CSM1. 1m. *J Appl Meteor Sci*, 2021, 32(5): 553-566.
- [25] 魏敏. 气象高性能计算应用服务环境适应性研究. *气象*, 2015, 41(1): 92-97.
Wei M. Study on adaptability of meteorological high performance computing application services environment. *Meteor Mon*, 2015, 41(1): 92-97.
- [26] 王娟怀, 李清泉, 汪方, 等. 基于 DERF2. 0 的华南前汛期降水订正. *应用气象学报*, 2021, 32(1): 115-128.
Wang J H, Li Q Q, Wang F, et al. Correction of precipitation forecast predicted by DERF2. 0 during the pre-flood season in South China. *J Appl Meteor Sci*, 2021, 32(1): 115-128.
- [27] Gillett N P, Shiogama H, Funke B, et al. The Detection and Attribution Model Intercomparison Project Project (DAMIP V1. 0) contribution to CMIP6. *Geosci Model Dev*, 2016, 9: 3685-3697.
- [28] O'Neill B C, Tebaldi C, van Vuuren D P, et al. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci Model Dev*, 2016, 9: 3461-3482.
- [29] Jones C D, Arora V, Friedlingstein P, et al. C4MIP: The Coupled Climate-Carbon Cycle Model Intercomparison Project: Experimental protocol for CMIP6. *Geosci Model Dev*, 2016, 9: 2853-2880.
- [30] Zhou T J, Turner A G, Kinter J L, et al. GMMIP (V1. 0) contribution to CMIP6: Global Monsoons Model Inter-comparison Project. *Geosci Model Dev*, 2016, 9: 3589-3604.
- [31] Webb M J, Andrews T, Bodas-Salcedo A, et al. The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6. *Geosci Model Dev*, 2017, 10: 359-384.
- [32] van den Hurk B, Kim H, Krinner G, et al. LS3MIP (V1. 0) contribution to CMIP6: The Land Surface, Snow and Soil moisture Model Intercomparison Project: Aims, setup and expected outcome. *Geosci Model Dev*, 2016, 9: 2809-2832.
- [33] Lawrence D M, Hurtt G C, Arneth A, et al. The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6: Rationale and experimental design. *Geosci Model Dev*, 2016, 9: 2973-2998.
- [34] Xin X, Wu T, Jie W, et al. Impact of higher resolution on precipitation over China in CMIP6 HighResMIP Models. *Atmosphere*, 2021, 12(6): 762.
- [35] Zhang J, Furtado K, Turnock S T, et al. The role of anthropogenic aerosols in the anomalous cooling from 1960 to 1990 in the CMIP6 Earth system models. *Atmos Chem Phys*, 2021, 21: 18609-18627.
- [36] Ahmed F, Neelin J D. A Process-oriented diagnostic to assess precipitation-thermodynamic relations and application to CMIP6 models. *Geophys Res Lett*, 2021, 48, e2021GL094108.
- [37] Publication Hub. [2022]. <https://cmip-publications.llnl.gov/view/CMIP6/?type=model>.
- [38] IPCC. Annex II: Models//Climate Change 2021: The 19 Physical Science Basis. Cambridge: Cambridge University Press, 2021.
- [39] 黄晓远, 李谢辉. 基于 CMIP6 的西南暴雨洪涝灾害风险未来预估. *应用气象学报*, 2022, 33(2): 231-243.
Huang X Y, Li X H. Future projection of rainstorm and flood disaster risk in Southwest China based on CMIP6 models. *J Appl Meteor Sci*, 2022, 33(2): 231-243.
- [40] 俞静雯, 李清泉, 丁一汇, 等. 气候变暖背景下青藏高原夏季水汽的长期变化趋势分析. *中国科学(地球科学)*, 2022, 52(5): 942-954.
Yu J W, Li Q Q, Ding Y H, et al. Long-term trend of water vapor over the Tibetan Plateau in boreal summer under global warming. *Sci China (Earth Sci)*, 2022, 52(5): 942-954.
- [41] 刘绿柳, 魏麟骁, 徐影, 等. 气候变化对黄河流域生态径流影响预估. *水科学进展*, 2021, 32(6): 824-833.
Liu L L, Wei L X, Xu Y, et al. Projection of climate change impacts on ecological flow in the Yellow River Basin. *Adv Water Sci*, 2021, 32(6): 824-833.

Implementation and Application of BCC CMIP6 Experimental Data Sharing Platform

Ma Qiang¹⁾²⁾ Yan Jinghui³⁾⁴⁾ Wei Min³⁾⁴⁾ Xin Xiaoge³⁾⁴⁾
Zhang Li³⁾⁴⁾ Zhang Fang³⁾⁴⁾ Wu Tongwen³⁾⁴⁾

¹⁾ (National Meteorological Information Center, Beijing 100081)

²⁾ (Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

³⁾ (CMA Earth System Modeling and Prediction Centre, Beijing 100081)

⁴⁾ (State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081)

Abstract

The experimental data of ongoing CMIP6 (Coupled Model Intercomparison Project Phase 6) are widely used to study the mechanism of climate change and provide technical support for the assessment report of the Intergovernmental Panel on Climate Change (IPCC). With more types of model experiments and more complex climate model, the amount of CMIP experimental data are also increasing rapidly. Therefore, Beijing Climate Center (BCC) has established Earth System Grid Federation (ESGF) data node to share experimental data of BCC CMIP6.

BCC has three latest versions of models to participate in the project through model development in recent years. The hardware of the platform adopts a distributed storage architecture and is deployed in the demilitarized zone (DMZ) of China Meteorological Administration, which provides a strong guarantee for its network access rate and security. The data processing module mainly checks the integrity, processes the original model output and adopts the climate model output rewriter (CMOR) software to standardize the format. Thematic real-time environmental distributed data services data server is used for local storage management and data sharing, publishing metadata to ESGF index node for unified data retrieval. The data storage directory adopts hierarchical management structure with self-describing information to realize hierarchical and classified storage of different elements in different experiments. To ensure the security of data sharing, the platform is optimized based on ESGF security framework in addition to physically adding replica storage, and the needs of easy access are also considered.

Totally, 190 TB experimental data of BCC CMIP6 have been released and shared since the establishment of the platform. The platform has provided important technical support for BCC to participate in the CMIP6, and it has also supported scientific research in the fields of climate change simulation and prediction, weather and climate extremes, global warming and human activities.

Subsequent work will provide continuous data services to the CMIP and can be extended to other related model comparison programs. It is also important to further improve the capabilities of customized data sharing services.

Key words: CMIP; BCC; model experimental data; data sharing; data security