

南京师范大学

毕业设计（论文）

（2019 届）



题 目：____ 陆地生态系统碳循环模型
 ____ 的开放式对比系统构建方法研究

学 院：____ 地理科学学院

专 业：____ 地图学与地理信息系统

姓 名：____ 沈超然

学 号：____ 161302088

指导教师：____ 陈旻

南京师范大学教务处 制

目 录

第 1 章 绪论	1
1.1 研究背景及意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 国内外研究现状综述	3
1.2.1 陆地生态系统碳循环模型	3
1.2.2 陆地生态系统碳循环模型的验证与对比	4
1.2.3 开放式服务化系统框架	5
1.2.4 研究现状分析与总结	5
1.3 研究目标与研究内容	6
1.3.1 研究目标	6
1.3.2 研究内容	6
1.4 研究方法与技术路线	7
1.4.1 研究方法	7
1.4.2 技术路线	8
1.5 论文组织结构	8
第 2 章 陆地生态系统碳循环及其模型和数据	10
2.1 陆地生态系统碳循环	10
2.2 陆地生态系统碳循环模型	11
2.2.1 陆地生态系统碳循环模型及其分类	11
2.2.2 IBIS, Biome-BGC, LPJ 模型简介	13
2.3 陆地生态系统碳循环相关数据	15
2.3.1 气象数据集	15
2.3.2 土壤数据集	15
2.3.3 植被功能类型数据集	15
2.3.4 通量观测数据集	15
2.3.5 MODIS GPP/NPP 数据集	17
2.3.6 其他数据集	17

第3章 陆地生态系统碳循环模型开放式对比框架	18
3.1 陆地生态系统碳循环模型对比情景和业务分析归纳	19
3.1.1 对比情景分析和总结	19
3.1.2 对比业务抽象和归纳	20
3.1.3 对比系统功能模块设计	22
3.2 基于服务的开放式地理资源库	23
3.2.1 模型服务资源库	23
3.2.2 数据服务资源库	23
3.2.3 单位量纲资源库	24
3.2.4 数据重构服务资源库	24
3.2.5 可视化服务资源库	24
3.2.6 对比服务资源库	24
3.3 基于微服务的分布式网络架构设计	25
3.3.1 资源服务器层	26
3.3.2 计算服务器层	26
3.3.3 对比服务器层	26
3.3.4 服务器通信	28
3.4 开放式对比科学工作流引擎	29
3.4.1 对比流程分析和归纳	29
3.4.2 对比自动化执行引擎	29
3.5 本章小结	31
第4章 开放式对比资源接入方法	32
4.1 开放式碳循环相关数据资源接入方法	32
4.1.1 碳循环相关数据资源结构特征分析	32
4.1.2 碳循环相关数据资源描述方法	35
4.1.3 碳循环相关数据资源的服务化封装	37
4.2 开放式碳循环模型资源接入方法	39
4.2.1 碳循环模型资源特征分析	39
4.2.2 碳循环模型资源描述方法	39
4.2.3 碳循环模型资源的封装、部署、发布	43
4.3 开放式碳循环模型对比方法	46
4.3.1 碳循环模型对比方法	46
4.3.2 开放式碳循环模型对比方法接入方法	49
4.4 本章小结	49

第 5 章 原型系统与案例验证.....	50
5.1 原型系统构建	50
5.1.1 系统网站组织	50
5.1.2 系统开发运行环境.....	50
5.1.3 网络架构和系统功能模块设计.....	50
5.1.4 陆地生态系统碳循环对比资源库	50
5.1.5 开放式对比任务执行引擎	50
5.2 实验案例	50
5.2.1 模型资源和数据资源	50
5.2.2 对比方案.....	50
5.2.3 对比结果.....	50
第 6 章 结论与展望.....	51
6.1 研究结论	51
6.1.1 开放式对比系统	51
6.1.2 IBIS、Biome-BGC、LPJ 模型的验证和评价	51
6.1.3 IBIS、Biome-BGC、LPJ 模型对气候变化的响应	51
6.1.4 IBIS、Biome-BGC、LPJ 模型的不确定性分析	51
6.1.5 GPP、NPP 的时空分布格局.....	51
6.2 创新点	51
6.2.1 开放式、可共享、可重用的对比系统架构	51
6.3 不足与展望	51
6.3.1 不足	51
6.3.2 展望	51
参考文献	52

第 1 章 绪论

1.1 研究背景及意义

1.1.1 研究背景

地理模型是对地理过程和机理的抽象表达,是通过模拟、预测或重现这些地理现象来解决地理问题的重要手段。通过地理模型的模拟可以有效解决复杂的地理问题,促进地理学研究的发展。地理模型的对比是检验和验证模型模拟效果的必要方式,可以有效促进模型的发展改进。

以陆地生态系统碳循环领域为例,为了模拟生态系统中大气、海洋、陆表和化石燃料四个碳库之间的碳循环过程,预测不同尺度的碳收支格局和变化情况(Cao M K et al., 2005; Cao, et al., 2003), 大气学家研发了众多的碳循环模型,大致可以分为统计模型、遥感参数模型、生态过程模型和遥感、过程耦合模型四类。代表性的有 Miami 模型^[1]、CASA^{[2][3]}、GLO-PEM^{[4][5]}、BIOME-BGC^{[6][7][8]}、IBIS (Foley et al., 1996)、LPJ DGVM^{[9][10]} 等。各个模型由于其模拟机理不同,有其各自适用的尺度和范围。

为了全面客观地检测气候变化,评估人为因素对气候变化的影响,以及更好的模拟气候变化,国际上成立了联合国政府间气候变化专门委员会(IPCC, Intergovernmental Panel on Climate Change)。IPCC 每 5 年开展并出版一次评估报告,提供有关气候变化的现象、成因、影响以及对策等信息^[11]。除此之外,世界气候研究计划(WCAR, World Climate Research Programme)成立了耦合模型对比计划(CMIP, Coupled Model Intercomparison Project),针对每届 IPCC 评估报告中出现的科学问题制定标准实验方案,对比各个模型的模拟能力^[12];其中,第六阶段耦合模型对比计划(CMIP6)背书的模型对比项目有 23 余个,包括碳循环模型对比项目、云反馈对比项目、十年际气候预测项目、海洋模型对比项目、古气候模型对比项目、辐射强迫模型对比项目、协调区域降尺度实验等。^[13]可见,模型对比评价是当前气候模型研究领域的一个热点。

据统计,在第五次耦合模型比较计划(CMIP5)中,超过 20 个建模组使用 50 多个碳循环模型参与模拟^[14],越来越多的模型参与极大地促进了对气候变化的监测和评估。但是,模型对比仍然面临着诸多困难。第一,由于地球系统模式的复杂性,碳循环模型通常涉及到生物(光合作用、呼吸作用)、物理(植被冠层能量交换)、化学(碳、水循环)、遥感(数据同化)等模块,每个模型由于其机理不同,

涉及到的模块组成往往不同，而且各个模块的实现方法也不尽相同；第二，由于地理现象的时空分异特性，地理模型在模拟地理现象时往往受到研究区域的局限，对于模型全面的评估应该是从多尺度、多视角进行；第三，由于模型参数的复杂性，在应用模型模拟时，在调参方面需要花费巨大的精力，调参过程也需要耗费大量的计算资源；第四，碳循环模型由不同的科研人员编写，在实现层次上，面临着操作系统、编程语言、软硬件环境依赖、运行方式、输入输出数据格式等多方面的异构性，难以部署、编译和应用，阻碍了模型的推广。这些原因共同限制了模型对比工作的发展，模型对比也是当前气候模型研究领域的一个**难点**。

随着计算机技术的发展，**Web Service** 和面向服务的软件架构（**SOA, Service-Oriented Architecture**）技术不断成熟，为分布式网络环境下的地理模型的共享与重用提供了技术上的有效支持，地理模型和数据逐渐朝着网络服务化的方向进行封装与共享^[15]。服务化的地理模型将复杂的地理模型封装为一个黑箱，屏蔽了模型在使用平台、编程语言、软硬件环境依赖以及数据格式等方面的复杂性和异构性，使用者不必关心其内部实现的细节，简单了解其调用接口就能够开展地理模拟；而地理数据服务促进了数据的共享重用，降低了模型应用时数据搜集的困难^[16]。以地理模型服务和数据服务为基础，将网络服务技术手段应用到地理模型对比上，不仅可以将对对比过程公开化透明化，而且可以将对比方案共享出来，达到可重用的目的，为开放式模型对比提供了一种新的视角和技术手段。**本文基于 Web Service 技术规范**，设计了一套开放式模型对比框架，可以支持开放式网络环境下异构地理模型的参与式和可共享式对比。并以 **IBIS**、**LPJ**、**Biome-BGC** 三个开源碳循环模型为例，模拟了三个模型在全球 150 个站点上的碳收支排放，并与 **Fluxdata** 和 **MODIS** 观测数据进行对比，评估了三个模型的模拟能力，验证了开放式模型对比框架的可用性，为地理模型对比提供了一种新的思路。

1.1.2 研究意义

采用传统的单机式地模型部署和对比需要将所有模型部署在同一台计算机上，面临着部署过程复杂、对比结果难以再现、对比方法不公开等一系列问题。本文通过基于 **Web Service** 的对比框架设计，对地理模型来说，屏蔽了其在操作系统、编程语言、软硬件环境依赖以及数据格式依赖等方面的复杂性和异构性，降低了模型的使用门槛；对地理数据来说，降低了数据收集、存储、管理、处理的繁琐流程；对于对比方案来说，一方面可以将传统的线下对比过程公开化、透明化，使对比过程具有可追溯性、可验证性和易重复性，另一方面对比方案也可以共享出来，达到可共享、可复用的目的。能够提供一个更加公平公正公开的“擂台”。

对于陆地生态系统碳循环领域，本平台提供了一个开放式的模型对比“擂台”，通过碳循环模型之间的对比，一方面验证和诊断了模型的模拟效果，促进了模型的应用和改进，另一方面促进了耦合模型的协同研究和实验，从而在人类活动对自然环境的影响方面获取更可信地评判，指导人类面对气候问题及时做出正确的决策。

1.2 国内外研究现状综述

开放式地理模型对比系统的构建涉及多个方面，包括模型和数据资源的深入了解、模型对比评价方法的总结归纳、以及开放式服务化系统架构的搭建。认识地理模型的运行行为特征和数据特征是驱动其模拟计算的前提，剖析前人对比领域模型的统计学方法是丰富对比形式的手段，熟悉开放式服务化系统的框架是构建系统的基础。因此，本文从陆地生态系统碳循环模型的发展历程，前人对碳循环模型的对比评价方法以及开放式服务化系统的框架三个方面进行回顾，分析陆地生态系统碳循环领域地理模型对比的研究现状、存在问题和发展趋势。

1.2.1 陆地生态系统碳循环模型

陆地、大气、海洋以及化石燃料碳库是地球系统碳循环的四个主要组成部分，其中陆地生态系统又是其中最为活跃的碳库，也是人类活动聚集的场所。陆地生态系统碳循环模型通过模拟植物光合器官碳库、植物支持器官碳库、凋落物碳库和土壤有机碳库之间的碳源汇交换，来计算植被的初级生产力^[17]。

陆地生态系统碳循环模型从机理上可分为统计模型、生态过程模型和遥感、过程耦合模型。统计模型可分为气候统计模型和遥感统计模型。其中气候统计模型主要通过气候因子与植被净初级生产力（Net Primary Productivity, NPP）的实测数据之间建立回归方程；遥感统计模型通过遥感光谱指数（如 NDVI）与 NPP、生物量等数据间的相关关系进行统计回归。统计模型简单直观，具有较强的区域适用性，但其完全依赖于地面观测数据，对于不同的区域，模型不具备普适性和推广性。同时统计模型没有考虑陆地生态系统碳循环过程的内部机理，无法揭示生态系统与环境间的相互影响关系，不能用于对未来的预测研究^{[18][19]}。生态过程模型按照是否考虑实际环境对植被类型、组成和结构的影响分为基于动态植被类型的模型和基于静态植被类型的模型^{[20][21][17]}，按照涉及到的机理类型分为地球化学过程模型、陆面物理过程模型和生物过程模型^[19]。过程模型由于其综合考虑了碳循环过程的动力学特征，结合了气候、土壤和植被生理生态参数，以及陆地生态系统与大气、海洋之间的相互作用，模拟结果相对来说更加准确，逐渐占据了主

导地位。遥感、过程耦合模型通过将遥感观测数据（如叶面积指数 LAI）同化到模型之中，来提高模型模拟的精度^[22]，他融合了遥感统计模型和生态过程模型的优点，可以反映区域和全球尺度的 NPP 空间分布和变化^[23]。

综上所述，生态过程模型在全球范围内不仅模拟效果好，而且与遥感、过程耦合模型相比更加简单、容易实现。因此，本文选取了三个适合全球尺度的生态过程模型参与对比：IBIS、BIOME-BGC、LPJ DGVM，并总结了各自的特征，如表 1 所示。从中可知，这些模型的时空适用范围、输入数据、输出要素都很相似，均具备全球范围内碳循环的模拟能力，彼此之间可以进行对比。

1.2.2 陆地生态系统碳循环模型的验证与对比

对于陆地生态系统碳循环模型模拟结果的验证和对比，前人大多从多个视角进行综合对比，包括对模拟结果的可视化观察对比，使用统计学方法定量地对比等。没有单独的评价手段被认为是优越的，相反的，多种对比技术的结合使用能够为模型模拟能力提供全面评价^[14]。

在对比对象方面，目前主要分为三类：基于通量站点数据的对比、基于卫星遥感数据的对比和基于多模型模拟结果的交叉对比。杨延征^[24]、王磊（2010）、刘曦^[25]、张海燕（2006）等人分别在中国全国尺度、中国东部、东北东部森林和东北帽儿山应用 IBIS 模型估算碳排放情况，将模型模拟的结果与通量站点观测数据对比，验证其在中国的模拟能力。胡瀚予（2011）、NIU Ben^[26] 等人分别以台湾陆域生态区和青藏高原为研究区域，将生物测定法、Rahman's 方程式以及 VPM、PCM、AVM 的模拟结果与 MOD17 数据集相对比，评估其模拟能力。P.Friedlingstein（2006）在 CMIP4 中通过 11 个碳循环模型之间模拟结果的相互比较来评价其模拟能力。

在对比方法方面，通过绘制模拟值等值线图、观测值等值线图、差异等值线图可以非常直观地展示出各个模型模拟结果的差异；通过泰勒图展示模拟值与观测值之间的标准差之比、均方根误差以及相关系数；通过纵向图和时间序列折线图可以清晰地展现模型在各个子区域中的适用情况^{[27][28][29][30]}（Kim J, 2012; Kim J, 2014; Kim J, 2018; 郭彦, 2013; 刘敏, 2009）。Stoner A M K^[31] 使用统计降尺度的方法建立当前观察值、当前模拟值、未来模拟值与未来观察值之间的关系，预测未来的气候因子发展状态。

在对比框架方面，国际上比较知名的是 WCAR 开展的 CMIP。通过制定一系列证明模型模拟能力的标准实验，参与比较的模型都遵循实验协议进行测试。使用一致的预测因子（如数据源、空间分辨率、时间分辨率等），针对特定的预测指

标（如温度、湿度、降水量、NPP等），在一致的时间范围内模拟，最后将模拟结果提交给专家评审，通过审核的对比结果被公开发布在网站上^[14]。耦合模型对比计划、协调区域降尺度实验以及区域气候模型评估系统都遵循着这种对比框架开展对比^{[32][33][34]}。

1.2.3 开放式服务化系统框架

随着计算机技术的发展，越来越多的数据被发布为服务，其中具有代表性的是开放地理信息联盟（OGC, Open Geospatial Consortium）的 WMS、WFS、WCS。在地理模型和数据处理方法方面，OGC WPS（Web Processing Service）也为地理模型的共享提供了解决办法。

袁爽^[35]基于 OGC WFS、WMS 标准，实现了空间数据的服务发布；吴楠^[36]、毛曦^[37]、宋东泽^[38]基于 OGC WPS 标准，分别设计并实现了 VPM 模型和日均蒸散量模型的服务发布。Yue S^{[39][16]}基于 Web Service 提出了一套模型服务的元数据描述、封装、打包和发布的技术体系。

1.2.4 研究现状分析与总结

纵观陆地生态系统碳循环模型的研究现状，陆地生态系统碳循环模型的理论体系越来越完善，应用越来越广泛。随着 IPCC 等国际组织的牵头推动，模型的对比和评价也逐渐开始开展，极大地促进了模型的改进和发展。分析当前的研究现状总结为以下几点：

(1) 陆地生态系统碳循环模型：

随着技术手段的不断成熟，越来越多的模型被开发出来。但由于地球系统的复杂性，模型在类别、机理、模块结构、参数、数据等方面形态各异，这种多样性阻碍了模型在不同区域以及全球范围内的应用推广，开展模型对比是解决该问题的有效技术手段；

(2) 碳循环模型的对比：

碳循环模型的对比是当前的研究热点和难点，在对比对象方面模型可以与通量观测站点、遥感影响影像数据以及其他模型的模拟结果相对比；在对比方法方面大多数采用可视化方法和统计学手段；在对比框架方面世界气候研究计划制定了一套模型对比的实验方案，参与者通过实验内容的实施、提交、评审，最终可以将对比结果公开出来。目前的对比相对完善，但是在互联网技术的高速发展下，有待进行一些新的技术尝试；

(3) 开放式服务号系统框架：服务号包括模型资源的服务化和数据资源的服务

化。目前基于 OGC 的 WFS、WMS 和 WPS 标准形成了一套成熟的技术体系。模型服务和数据服务地发布，可以降低模型的使用难度，促进模型的应用和完善，但是这种技术手段在气候模型以及模型的对比方面缺乏广泛地应用。

1.3 研究目标与研究内容

1.3.1 研究目标

针对目前地理模型和数据资源分散、模型安装部署困难、模型模拟效果难以评估、评估成果难以再现的问题，本文以陆地生态系统碳循环模型为例，基于 Web Services 技术，旨在设计并实现一个提供模型对比的“擂台”，并用 IBIS、BIOME-BGC 和 LPJ 三个模型验证平台的实用性。具体研究目标包括：(1) 设计出一套新的基于网络服务的开放式地理模型对比系统框架；(2) 以陆地生态系统碳循环模型为例，实现地理模型资源 and 数据资源的开放式接入；(3) 以陆地生态系统碳循环模型为例，实现对比方案的构建和实施。

1.3.2 研究内容

为实现上述研究目标，设计研究内容如下：

(1) 开放式地理模型对比系统框架的构建

开放式对比系统框架是开展详细对比工作的基础支撑，它描述了模型资源、数据资源、对比方法资源的标准接入方式，所有这些资源在框架的规范下以科学工作流的方式，有条不紊地运行、耦合、集成。

考虑到网络资源的传输成本、模型运算和模型对比的业务逻辑，本文将整体服务器框架分为三部分：1) 资源服务器：负责数据存储、数据库管理、数据服务发布和数据缓存；2) 模型计算服务器：负责模型服务发布、模型调用、数据处理和分布式计算；3) 模型对比服务器：负责资源汇总和展示、对比方案的构建、对比任务的分发和比较结果的可视化。三个部分之间通过 Http 和 Ftp 协议通信。这种情况下，地理模型、地理数据、数据处理方法、数据可视化方法和数据对比方法都以组件的形式分别插入到各个模块中，实现了系统架构的解耦。

(2) 开放式地理模型资源的接入

地理模型资源是参与模型对比的主体，通过模型资源的汇聚形成模型资源库，而模型资源库是模型对比的重要组成部分。为了实现模型资源的开放式接入，首先需要定义一套模型的元数据描述接口，其中包括模型的属性描述、模型的调用接口描述和模型的软硬件依赖环境描述，结合 Web Service 技术，通过

应用服务器向外暴露 Restful API 可以将地理模型发布为服务，模型服务向数据库中注册可以实现模型服务的管理。在本文中发布的模型条目包括 IBIS、BIOME-BGC、LPJ DGVM。

(3) 开放式地理数据资源的接入

地理数据资源的接入是驱动模型运行的基础。对于地理数据资源的接入方法和模型资源类似。对于本文涉及到的碳循环模型的运行，需要发布的数据服务包括陆面掩膜数据、气象数据、土壤数据、植被数据和其他配置数据。其中气象数据的时间分辨率是 1 天，气象参数包括最高温、最低温、日均温、降水量、相对湿度、平均风速、大气压和云覆盖率等；土壤数据包括沙粒含量和黏土含量等；植被数据包括植被类型、凋落物储量和植被功能型数据等。

(4) 地理模型对比方案的构建和实施

由于单一的对比方法不能够完备地展示出各个模型之间的差异，本文的对比从构建对比方案入手。对比方案是从多个视角综合对比几个模型的模拟能力，是多个对比方法的组合。因此，在构建对比方案之前，首先要构建一个对比方法库，包括现阶段流行的偏差等值线图、泰勒图、时间序列折线图、纵向图、区域降尺度等统计学方法都作为对比方法库的成员，每种对比方法针对特定的结果数据集作出评价。最后，对比方案通过配置具体的输入数据开始模拟，模拟结束会生成一个评价报告文档。在这个过程中，对比方案也作为一种开放的资源共享和重用起来。

1.4 研究方法与技术路线

1.4.1 研究方法

(1) 文献分析法

搜集并整理国内外相关领域研究者关于陆地生态系统碳循环模型、模型对比方法、模型和数据共享与服务化方法的研究成果，总结前人研究方法的特点和使用范围，针对这些成果的优劣势进行分析和借鉴，为本文的研究方法提供思路。

(2) 归纳演绎法

分析地理模型在运行行为上的特征，对比并归纳其调用方式，在此基础上设计模型的元数据描述接口，为模型的服务化方法做准备。

(3) 数理统计法

本文对模型模拟结果的对比评价分析是通过在模型模拟结果与实测值之间进行统计分析，根据统计特征值来评价模拟结果的优劣。

(4) 案例验证法

本文以 IBIS、BIOME-BGC、LPJ 模型为例,对各个模型进行封装,以满足模型在站点和区域上的模拟计算,并对模拟结果进行数理统计,分析其优缺点。

1.4.2 技术路线

根据本文的研究目标和研究内容,本文设置的技术路线如图 1 所示,包括 5 个阶段:1) 问题分析与资料整理阶段。收集并整理文献,分析现有陆地生态系统碳循环模型的特征、碳循环模型的对比方法和开放式服务化系统架构,确定本文的研究思路;2) 模型对比系统框架的构建:搭建系统的整体网络架构,设计各个服务器模块的功能职责和服务器之间的通信方式;3) 模型和数据资源的开放式接入:根据模型资源 and 数据资源的特点,分别设计其开放式资源接入方法;4) 模型对比方案的构建和实施:首先丰富模型对比方法库,然后根据模型对比方法的组合和综合评价标准构建模型对比方案,配置数据后调用模型生成评价对比报告;5) 实验验证:根据选取的三个模型分别在全球陆地范围内进行模拟,对比其模拟水平。

1.5 论文组织结构

第一章:绪论。介绍开放式对比的意义,梳理当前国内外的研究现状,得出模型对比是碳循环领域内的一个热点和难点,并提出了当前时代下开放式服务化技术体系可以为模型对比提供了新的思路,在此基础上确定了本文的研究目标、研究内容、研究方法,并设计了具体的技术路线。

第二章:碳循环、碳循环模型及数据。本章概述地球系统碳循环的基本结构,并确定陆地生态系统碳循环在其中的位置和作用。总结归纳了陆地生态系统碳循环模型的概念、结构、分类体系和优缺点,指出当前的陆地生态系统碳循环模型可分为统计模型、生态过程模型和遥感、过程耦合模型,并详细介绍了参与本文对比的三个生态过程模型模型:IBIS、Biome-BGC、LPJ。最后介绍这三个模型运行所需的数据、通量网的观测数据和遥感通量数据产品。

第三章:陆地生态系统碳循环开放式对比框架。本章从四个角度详细叙述了开放式的对比框架。首先,对国际上的开展的对比计划中所涉及到的对比情景进行分析总结,得出对比情景包括历史情景、未来预测情景、敏感性分析情景和参数校准情景等,在对这些情景的业务执行流程进行抽象归纳后,将对比流程抽象为创建对比话题、对比方案和对比任务的三步,并对系统的功能进行了整体设计;其次,从组件视角上分析了框架下面包含的地理资源组件模块,这些模块包括模型服务资源库、数据服务资源库、单位量纲资源库、数据重构服务资源库、可视化

服务资源库和对比服务资源库，并详细介绍了每种组件服务的标准开放接口；再次，从网络架构的视角上选择了微服务的分布式架构，将计算节点分布在 n 台计算节点、1 台资源节点、1 台对比节点上，使系统整体功能有效分割同时又不显得过于臃肿；最后，从科学 workflows 执行流程的视角上，分析总结了对比的流程，并设计自动化执行引擎以驱动模型对比工作。

第四章：开放式对比资源接入方法。本章指出在开放式对比框架下，对比资源的开放式接入方法是系统的关键，并将对比资源分为数据资源、模型资源 and 对比方法资源三类。因此，本章分三个小节详细介绍三种资源的开放式接入方法。对于数据资源，首先从格式、尺度和编排四个角度分析本文所用数据的特点，对领域通用标准数据设计结构化描述文档，对非标准数据采用示例数据 URL 的形式进行描述。然后遵循 OGC WMS/WFS/WCS 标准，使用 GeoServer 发布数据服务，并提供数据下载服务和数据处理服务；对于模型资源，首先针对其特点进行分析，总结模型在参数策略、运行流程上的相同点和在开发层面的异构点。针对这些特征分三个视角进行结构化描述：面向人类理解的基本信息描述；面向机器运行的运行信息描述；面向模型部署的软硬件依赖环境描述。最后设计支持有源代码、无源代码和简单模型、集成模型的封装策略，在此基础上进行模型服务部署和发布。对于对比资源...

第五章：原型系统和案例验证。

第六章：结论和展望。

第2章 陆地生态系统碳循环及其模型和数据

2.1 陆地生态系统碳循环

地球系统碳循环是指碳在岩石圈、水圈、大气圈和生物圈之间，以各种形式相互转换和迁移的过程。碳元素主要分布于陆地生态系统、海洋、大气和化石燃料四个碳库，其中陆地生态系统中的碳库是含量最小的碳库，但由于其是人类活动聚集的场所，所以又是最为活跃的碳库。它主要由植物光和器官碳库、植物支持器官碳库、凋落物碳库和土壤有机物碳库四部分组成，如图 2.1 所示，植物通过光合作用将大气中的 CO_2 转换为碳水化合物，再通过自养呼吸、分配光合作用产物和死亡凋落过程把碳输送到大气、植物支持器官碳库和凋落物碳库；植物支持器官通过自养呼吸和死亡凋落把有机碳分别输送到大气和凋落物碳库；凋落物碳库通过异养呼吸和分解作用分别把碳输送到大气和土壤有机物碳库中；土壤有机物碳库最后通过异养呼吸把碳输送回大气碳库中。因此，碳在陆地生态系统中的转移路径为：大气——植物光合器官——植物支持器官——凋落物——土壤——大气，形成了一个首尾相连的闭环，因此，可以把陆地生态系统中碳的转移看成一个循环过程。

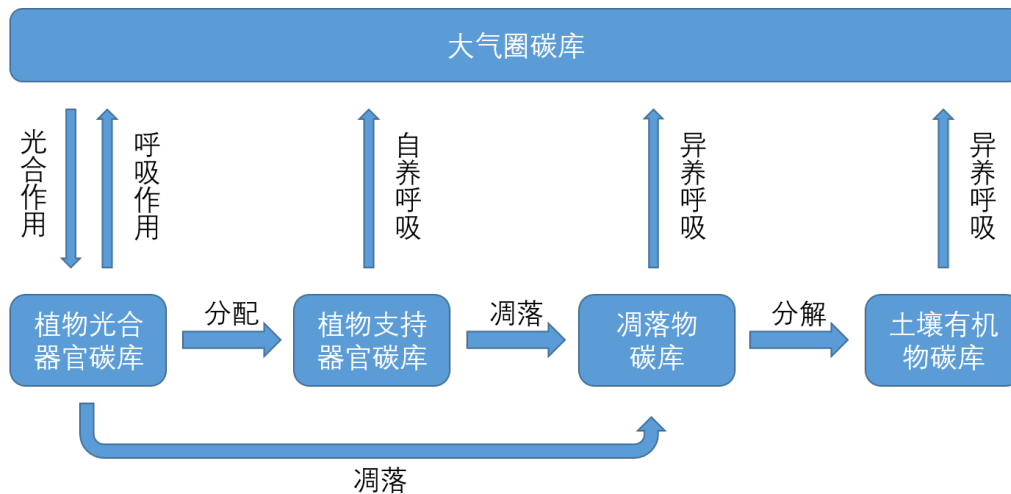


图 2.1 陆地生态系统碳循环基本过程

2.2 陆地生态系统碳循环模型

2.2.1 陆地生态系统碳循环模型及其分类

陆地生态系统碳循环中碳主要在植被——大气——土壤碳库之间转移，模型一般包括植物的光合作用、自养呼吸和异氧呼吸以及土壤的呼吸作用这四个过程，影响这些过程的因素有大气 CO_2 浓度、辐射、气温、土壤湿度、叶片中的叶绿素含量，所以许多模型耦合了水循环过程和氮循环过程。陆地生态系统碳循环模型通过生物生产力来描述碳通量的变化，这些指标主要有总初级生产力（Gross Ecosystem Production, GPP）、净第一性生产力（Net Primary production, NPP）、净生态系统生产力（Net Ecosystem Production, NEP），他们之间的关系如公式 2-4 和 2-2 所示，式中， R_A 表示自养呼吸， R_H 表示异氧呼吸。

$$NPP = GPP - R_A \quad (2-1)$$

$$NEP = (GPP - R_A) - R_H = NPP - R_H \quad (2-2)$$

除了这些指标外，模型计算过程中一般还包括有生物量（Biomass）、蒸散发（Evapotranspiration, ET）、径流（Runoff）、叶面积指数（LAI）等指标。

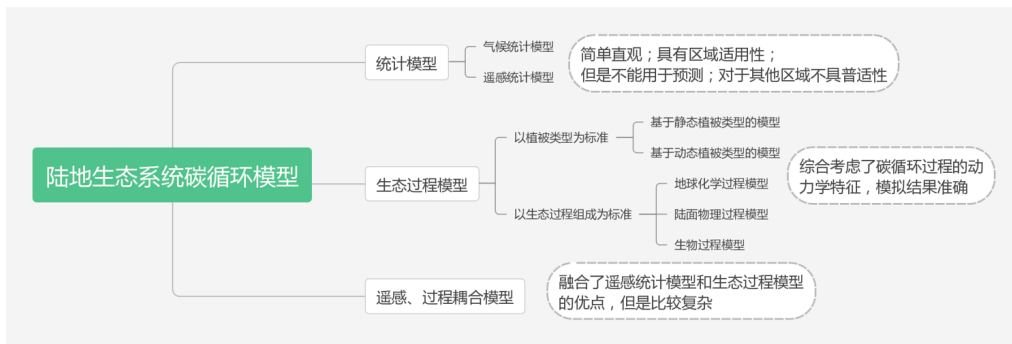


图 2.2 碳循环模型分类

陆地生态系统碳循环模型从机理上可分为统计模型、生态过程模型和遥感、过程耦合模型，如图 2.2 所示。统计模型可分为气候统计模型和遥感统计模型。其中气候统计模型主要通过温度、降水等气候因子与植被净初级生产力（Net Primary Productivity, NPP）的实测数据之间建立回归方程，代表性的有 Miami 模型（曼阿密模型）^[40]、Chikugo 模型（筑后模型）^[41] 和 Thornth-waite Memorial 模型（桑斯维特纪念模型）^[42]；遥感统计模型是基于遥感光谱指数建立的模型。由于总初级生产力（GPP）与截留光合有效辐射量和辐射的光合利用效率（LUE）成正比，而

截留光合有效辐射与总光合有效辐射（PAR）和光合有效辐射截留率（FPAR）成正比。其中 PAR 可以通过气候学方法计算得到，FPAR 可以通过遥感方法得到，因此，GPP 可以表达为公式 2-3，NPP 可以表达为公式 2-4

$$GPP(FPAR, PAR) = FPAR \cdot PAR \cdot LUE \cdot f(T, W, CO_2) \quad (2-3)$$

$$NPP = GPP - R_A \quad (2-4)$$

式中， $f(T, W, CO_2)$ 表示温度 T、土壤湿度 W 和大气中 CO_2 浓度对辐射的诗集光合利用效率的影响； R_A 表示自养呼吸；代表性的模型有 CASA、GLO-PEM、TURC、SIB2 等统计模型简单直观，具有较强的区域适用性，但其完全依赖于地面观测数据，对于不同的区域，模型不具备普适性和推广性。同时统计模型没有考虑陆地生态系统碳循环过程的内部机理，无法揭示生态系统与环境间的相互影响关系，不能用于对未来的预测研究^{[18][19]}。

生态过程模型按照是否考虑实际环境对植被功能类型、组成和结构的影响分为基于动态植被类型的模型和基于静态植被类型的模型^{[20][43][17]}，按照涉及到的机理类型分为地球化学过程模型、陆面物理过程模型和生物过程模型^[19]。过程模型由于其综合考虑了碳循环过程的动力学特征，结合了气候、土壤和植被生理生态参数，以及陆地生态系统与大气、海洋之间的相互作用，模拟结果相对来说更加准确，逐渐占据了主导地位。图 2.3 是生态过程模型的典型结构，其中冠层物理主要包括辐射传输、能量平衡、水平衡和气溶胶动力等过程，冠层生理主要用于计算光合作用和气孔导度，土壤物理包括能量平衡、水平衡和土壤温度等过程，自养呼吸主要指植被的维持呼吸和生长呼吸，异氧呼吸则指土壤呼吸和枯枝层分解，植被生理主要包括同化物分配和组织流转等过程，营养循环主要指碳、氮循环，植被竞争包括光竞争、水竞争和空间竞争，干扰过程主要涉及火灾、采伐（或收割）、林隙、热胁迫、干旱胁迫、背景胁迫（如接近植被生命年限，死亡不可避免）等，萌生过程主要指植被幼苗或种子在自然环境下的成活或萌发过程。生态过程模型的输入数据一般包括气象数据、 CO_2 浓度数据、土壤质地数据、高程数据和植被功能类型数据等。

遥感、过程耦合模型通过将遥感观测数据（如叶面积指数 LAI）同化到模型之中，来提高模型模拟的精度（张延龙，2015），他融合了遥感统计模型和生态过程模型的优点，可以反映区域和全球尺度的 NPP 空间分布和变化^[23]。但由于遥感、过程耦合模型难度过大，在应用时不太方便。

综合分析三类模型的优缺点后，本文选取了 IBIS、Biome-BGC、LPJ 三个具有

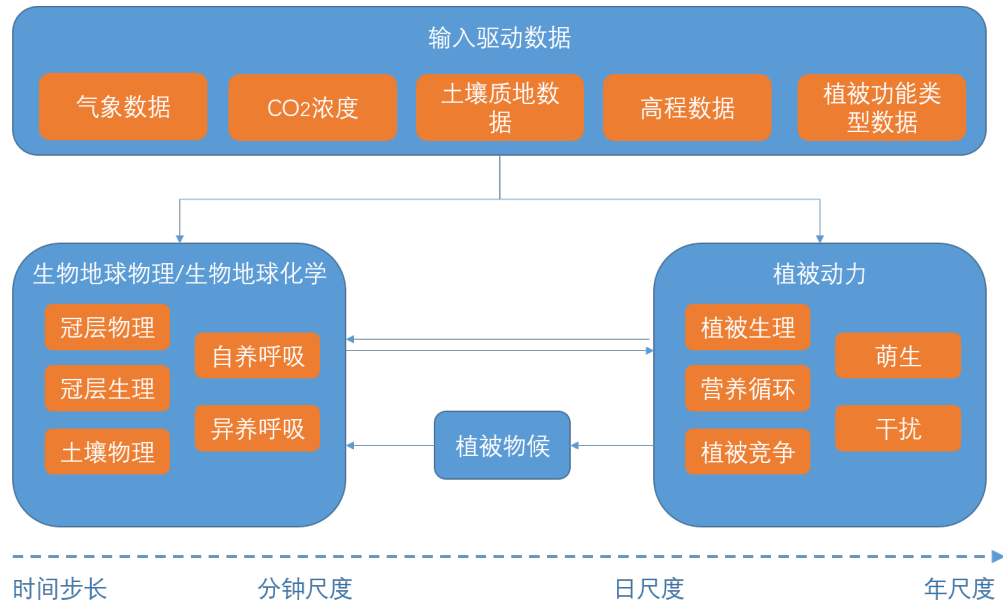


图 2.3 生态过程模型的典型结构

代表性的生态过程模型，在全球尺度上进行模拟和对比分析。

2.2.2 IBIS, Biome-BGC, LPJ 模型简介

2.2.2.1 IBIS 模型

IBIS 全称为集成生物圈模拟器（Integrated Biosphere Simulator），由美国威斯康星——麦迪逊大学 Foley 等^[44]提出，随后，Kucharik 等^[45]于 2000 年对其进行了改进，现在最新版本是 IBIS 2.6。模型包括地表模块、土壤地球生物化学模块、植被物候模块和植被动态模块 5 个模块，属于动态植被类型模型。IBIS 模型综合考虑了地球化学过程、地球物理过程和植被动态过程，能够耦合大气环流模式，在时间上适用于跨度从秒到数百年的碳循环过程，在空间上适用于区域尺度和全球尺度。

IBIS 模型的输入数据包括日分辨率的气象数据（平均温度、最大温度、最小温度、相对湿度、风速和云量）、土壤质地数据（沙粒含量、黏粒含量和粉粒含量）、植被功能类型（PFT）、网格点的经纬度。IBIS 模型将植被功能类型分为 12 种，其中林地分为热带常绿阔叶林、热带干旱落叶阔叶林、温带常绿阔叶林、温带常绿针叶林、温带阔叶林、北方常绿针叶林、北方干旱落叶阔叶林、北方冷落叶针叶林，灌木和草地分为常绿灌木、冷落叶灌木、C3 草本植物、C4 草本植物，其他详细参数见 Foley 等^[44]，Kucharik 等^[45]

2.2.2.2 Biome-BGC 模型

Biome-BGC^{[46][47][48]} 模型由美国蒙拿大大学森林学院陆地动态数值模拟团队 (Numerical Terradynamic Simulation Group, NTSG) 基于 FOREST-BGC 模型改进而来, 广泛应用于模拟植被、凋落物、土壤中的碳、氮、水等物质的循环过程^{??}。模型使用静态植被类型, 属于生物地球化学循环模型, 包括的生理生态过程有光合作用、蒸散发、生态系统呼吸、分解、光合产物的分配以及植物的死亡, 基本原理是物质和能量守恒定律, 即进入生态系统中的物质和能量, 等于离开生态系统的与生态系统中累积的之和。模型的植被功能类型被分为 7 类: 落叶阔叶林、常绿阔叶林、常绿针叶林、灌木林、落叶针叶林、C3 草本植物和 C4 草本植物, 每种植被功能类型对应一套生理生态参数。该模型在空间上适用于区域尺度和全球尺度。

Biome-BGC 模型的运行分为两阶段: spin-up 阶段是获取模拟的初始状态, 第二步是使用第一阶段的结果参数正式运行模拟程序。初始状态的获取有两种方法, 一是直接输入长期定点观测的参数值, 二是运行模型自带的 spin-up 程序。在 spin-up 过程中, 将模拟起点时的碳氮存量设为极低的值 (叶片的碳存量为 $0.001\text{kgC}/\text{m}^2$, 其他库的碳氮存量均为 0), 反复模拟数千年, 直到连续两年土壤碳差值小于 $0.005\text{kgC}/\text{m}^2$, 或达到设定的最大迭代年数为止, 这时认为系统达到稳定状态^{[49][50]}。

2.2.2.3 LPJ 模型

LPJ (Lund-potsdam-Jena Model, LPJ) 模型建立在 Biome3 模型的基础上, 包括的植被生理过程包括光合作用、呼吸作用、植被冠层能量交换、土壤水平衡等。LPJ 模型适用于大尺度、中尺度以及全球尺度范围内的碳循环模拟。LPJ 模型的时间步长是 1 月, Venevsky^{??}对运行时间步长进行了修改, 时间步长变成了 1 天。为了和 IBIS 模型与 Biome-BGC 模型的时间分辨率一致, 本文应用的是 Venevsky 的 LPJ-Daily。如无特殊说明, 本文后面所说的 LPJ 均指 LPJ-Daily。LPJ 模型的植被功能类型分为 10 种, 包括热带常绿阔叶林、热带雨绿阔叶林、温带常绿针叶林、温带常绿阔叶林、温带夏绿针叶林、北方常绿针叶林、北方夏绿针叶林、北方夏绿阔叶林、C3 草本植物和 C4 草本植物。LPJ 模型的运行也分为两个阶段: spin-up 阶段假设没有植被 (全为裸土地) 和生物量, 进行 1000 年积分直到植被覆盖和土壤碳达到平衡态; 常规运行阶段使用 spin-up 阶段运行出的平衡态作为初始值进行模拟。与 Biome-BGC 不同的是, LPJ 模型在内部将两个阶段直接串联在一起, 不用分两步运行。

2.3 陆地生态系统碳循环相关数据

本文所选的三个模型均以网格点为基本模拟单位,针对全球范围内的区域,本文使用 $0.5^{\circ} \times 0.5^{\circ}$ 的经纬网划分格点,并筛选去除海洋和裸土地的格点,最终在全球范围内共有 40595 个格点。

如前文所分析,生态过程模型一般都需要气象数据集、土壤数据集和植被功能类型数据集,而在进行对比时,要确保使用相同的通用数据集。而由于不同模型对这些数据在时空分辨率、植被类型等方面的要求不同,因此,本文采用相同的数据域,针对每个模型的特殊要求,对原始数据进行数据处理。

2.3.1 气象数据集

本文采用 MERRA 2 (Modern Era Retrospective-analysis for Research and Applications 2) 气象数据集, MERRA 是 NASA 为卫星提供的在大气再分析资料,致力于天气和气候时间尺度与水循环相关研究,具有较高的空间分辨率和较完整的时间序列范围,气象要素包括平均温、最高温、最低温、相对湿度、降水量、风速和云量。本文的数据时间范围从 1982 年到 2013 年共 32 年,以日为时间分辨率。数据原始空间分辨率为 $0.625^{\circ} \times 0.5^{\circ}$, 对其进行重采样转换为 $0.5^{\circ} \times 0.5^{\circ}$, 以符合全球离散网格点的空间分布。

2.3.2 土壤数据集

主要包括土壤的颗粒组成成分,即沙粒含量、黏粒含量、粉粒含量、土壤深度,由中山大学土壤—大气相互作用研究组开发 ??。数据的空间分辨率为 $0.0085^{\circ} \times 0.0107^{\circ}$, 对其进行重采样转换为 $0.5^{\circ} \times 0.5^{\circ}$ 。数据详情参考 <http://globalchange.bnu.edu.cn/research/soilw>

2.3.3 植被功能类型数据集

生态过程模型普遍采用植被功能类型作为植被基本处理单元,本文所选的三个模型也是如此。由于这三个模型内部对植被功能类型的划分不同,因此对原始数据进行类型映射,如表 ??, 2.2, 2.3 所示,分别为三个模型的植被功能类型映射表。

2.3.4 通量观测数据集

全球通量观测网络 (FLUXNET) 是一个以全球广泛分布的通量塔为基础的全球通量观测网络。通量观测站点遍布世界主要国家,在北美、欧洲、亚洲、非洲都

表 2.1 IBIS 模型植被功能类型映射表

植被功能类型	生物量转换时间			碳分配系数		比叶面积
	叶	根	干	叶	根	

数据来源于 Foley 等^[44]

表 2.2 Biome-BGC 模型植被功能类型映射表

植被功能类型	生物量转换时间			碳分配系数		比叶面积
	叶	根	干	叶	根	

数据来源于 Foley 等^[44]

表 2.3 LPJ 模型植被功能类型映射表

植被功能类型	生物量转换时间			碳分配系数		比叶面积
	叶	根	干	叶	根	

数据来源于 Foley 等^[44]

有它的区域网络存在（如亚洲通量网 AsiaFlux、美洲通量网 AmeriFlux、中国通量网 ChinaFlux）。到如今为止，累计有 800 多个活跃和历史的通量测定站点，分布在全球大部分气候空间和代表性生物群系中。覆盖的植被功能类型包括热带森林、温带针叶和阔叶森林、寒带森林、农田、草原、灌木丛、苔原和湿地。它在全球范围内连续测量大气状态变量，如温度、湿度、风速、降雨量、大气二氧化碳、植被生产力、生物量等。本文使用的是 La Thuile 数据集，包括站点每天的温度、降水、光合有效辐射比例（FPAR）、GPP、NEP，共计有 231 个通量观测站点，累计 900 余个站点·年次数据，作为观测数据集与模型模拟的结果进行站点尺度上的对比。站点分布如图 2.4 所示。

图 2.4 La Thuile 观测站点分布

表 2.4 站点植被功能类型统计表

2.3.5 MODIS GPP/NPP 数据集

MODIS MOD17 A3 提供了时空连续的遥感 GPP/NPP 产品，MODIS/GPP 产品以其他 MODIS 产品为基础计算而得，时间范围从 2000 年到 2015 年，时间分辨率为 8 天，空间分辨率为 $0.0083^{\circ} \times 0.0083^{\circ}$ ，对其进行重采样转换为 $0.5^{\circ} \times 0.5^{\circ}$ 。本文使用 MODIS GPP/NPP 数据集作为全球尺度上的对比验证数据集，与三个模型模拟的结果进行对比分析。

2.3.6 其他数据集

模型还需要网格点的物理常量数据，包括有格点高程数据、短波反射率等。其中高程数据使用的是 DEM 数据

第 3 章 陆地生态系统碳循环模型开放式对比框架

根据第2章的分析，陆地生态系统碳循环模型具有种类多样、参数复杂、难以应用的特点，模拟过程参与的要害众多（如 GPP、NPP、NEP、Biomass、LAI 等），模拟站点丰富，数据需求量大、难以搜集。这些因素都共同阻碍了碳循环模型的应用和发展，开展模型对比工作往往十分艰巨，像 CMIP 项目那样，各个模式开发组共同参与进来才能更顺利地完。但是传统的对比框架也面临着一些问题，它将模型计算和对比过程都放在本地计算机运行，模型难以共享和重用，对比过程不够公开，换一套数据或研究区域就难以重现或重复，这些都是由于基于本地计算的对比框架所导致的。针对这一问题，本章旨在设计一个开放式的对比框架，支持将各种对比资源以微服务的形式开放地接入进来，实现可共享、可重用、公开化的陆地生态系统碳循环模型的对比。

本章首先分析了陆地生态系统碳循环模型的对比情景，并从中归纳总结出“对比话题——对比方案——对比任务”的三部流程；其次详细探讨了对比过程中的需要接入的地理资源组件库，包括有模型服务资源库、数据服务资源库、单位量纲资源库、数据处理服务资源库、对比服务资源库，并设计了对应接口来满足相应资源的接入，实现了对比资源的标准一致性和可扩展性；然后设计了分布式的网络系统架构，将整体复杂的系统功能分割为三部分，实现了功能的解耦和系统的稳定和；最后通过科学工作流引擎支持网络环境下对比流程地自动化执行。本系统的整体架构如图3.1所示。

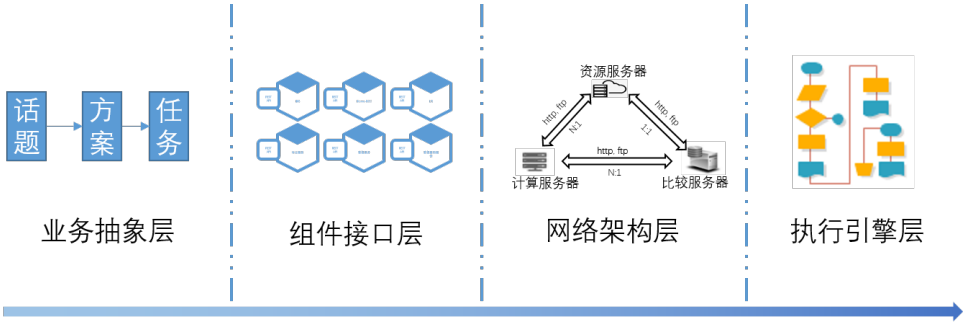


图 3.1 陆地生态系统碳循环模型开放式对比框架

3.1 陆地生态系统碳循环模型对比情景和业务分析归纳

3.1.1 对比情景分析和总结

陆地生态系统碳循环模型的应用广泛，未来的发展方向也主要停留在以下几个关键问题上：

(1) 陆地生态系统碳循环模型对历史植被生长的模拟

在 CMIP5 和 CMIP6 的设计中，都设定了历史情景模拟，即根据历史时期的太阳常数、温室气体浓度、臭氧浓度、气溶胶浓度等观测资料来模拟气候因子，是气候系统模式参与 CMIP 项目的必要实验。而陆地生态系统碳循环模型作为气候系统模式机理中的一个子过程，也需要对历史时期进行模拟，来评估模型的模拟能力。历史情景分析是最常见的模拟情景，本文将在实验案例中详细介绍模拟结果。

(2) 陆地生态系统碳循环模型的未来预测情景

在 CMIP5 和 CMIP6 中也设定了未来气候情景分析，即预测未来 CO₂ 浓度和气候因子（温度、降水）等指标的变化情况。在陆地生态系统碳循环模型对植被生产力的模拟方面，未来情景设定为气温、降水、CO₂ 浓度的变化对植被生产力的影响，即研究陆地生态系统碳循环模型对气候变化（如温度、降水等）和 CO₂ 浓度的响应。CMIP5 对未来 100 年全球气候变化的研究结果表明，全球平均气温从 1990 年到 2100 年上升 1.4 ~ 5.8°C^{[51][52]}；全球平均降水将会增加，但是降水格局也发生了很大的变化，即有些地区降水增加，有些地区降水减少；未来大气 CO₂ 浓度最大升高到当前的 2 倍^[53]。根据这些预测的未来气候情景，对温度、降水和 CO₂ 浓度的变化做出变量组合，表 3.1 可以表示一种参考的组合。

(3) 陆地生态系统碳循环模型的敏感性分析情景

在模型评价中，不确定分析研究的是模型参数、驱动变量等不确定性因素发生变化时，所引起的模型模拟结果的变化和变化程度。敏感性分析是模型不确定分析的一种常用方法，它用来研究和预测不确定性因素发生变化时，对模型结果影响程度的分析方法，又称为灵敏度分析。敏感性分析通常用来评估模型参数的重要性，分为总体敏感性分析和局部敏感性分析。局部敏感性分析相对来说更加简单，他分析单个模型参数对模拟结果的影响，但是他忽略了模型之间多个参数的相互作用对模拟结果的影响；全局敏感性分析不仅分析单个参数对模拟结果的影响，还分析参数之间的相互作用对结果的总影响，但计算起来比较复杂，耗费时间。在本文中，将敏感性分析也理解为一种对比情景，即在不同的模型参数下模拟结果的对比，可以与正常模拟与观

表 3.1 陆地生态系统碳循环模型的未来预测情景参数设定

因素	值	标记
气温	不变	$Temp_0$
	+2°C	$Temp_1$
	+4°C	$Temp_2$
	+6°C	$Temp_3$
降水	不变	$Prec_0$
	全年 +10%	$Prec_{+10}$
	全年 -10%	$Prec_{-10}$
	全年 +20%	$Prec_{+20}$
	全年 -20%	$Prec_{-20}$
	全年 +30%	$Prec_{+30}$
	全年 -30%	$Prec_{-30}$
CO ₂ 浓度	不变	$CO2_0$
	两倍	$CO2_{*2}$
多因素	温度增加、降水增加	$Temp_3 + Prec_{+10}$
	温度增加、降水减少	$Temp_3 + Prec_{-10}$
	温度增加、降水增加、CO ₂ 浓度翻倍	$Temp_3 + Prec_{+10} + CO2_{*2}$
	温度增加、降水减少、CO ₂ 浓度翻倍	$Temp_3 + Prec_{-10} + CO2_{*2}$

测数据的对比采用相同的对比方法。敏感性分析的情景设定需要从每个模型中选出一系列关键参数，根据参数的取值设定具体的情景，由于本文的研究重点不在于此，不做详细设定。

(4) 陆地生态系统碳循环模型的参数校准情景

碳循环模型的模拟精确度很大成分上依赖于模型参数的设定，目前在参数校准方面有人工校准和自动校准两种。人工校准通常有三种方法：对比实验、实验观测和文献查找；自动校准通过神经网络等各种模型进行校准。模型参数的校准可以理解为在不同模型参数下模型模拟结果的对比，因此可以作为对比的一种情景。

3.1.2 对比业务抽象和归纳

在进行对比业务抽象和归纳之前，先对本文中的一些名字做出解释：本文将 GPP、NPP、NEP、Biomass、LAI 等模拟指标称之为**对比要素**；将模型对比过程中

所需要的所有数据称之为**对比数据集**，一套对比数据集包括有输入数据集、观测数据集。其中输入数据集由气象数据集、土壤数据集、植被功能类型数据集、DEM数据集等组成，每个具体的数据集可以选择可替换的模块，如气象数据集既可以采用 MERRA 2 再分析资料，又可以使用 CRU 再分析资料。在第2.3节中具体介绍的数据本文称之为**标准对比数据集**，作为系统分析的基础原始数据，而在其之上参照表 3.1对数据进行的一系列修改，比如对温度 $\pm 2^{\circ}\text{C}$ ，对降水 $\pm 10\%$ 等，本文称之为**衍生对比数据集**。每一个模拟情景都有一套相对应的对比数据集。本文将参与对比的模型称之为**对比参与者**。

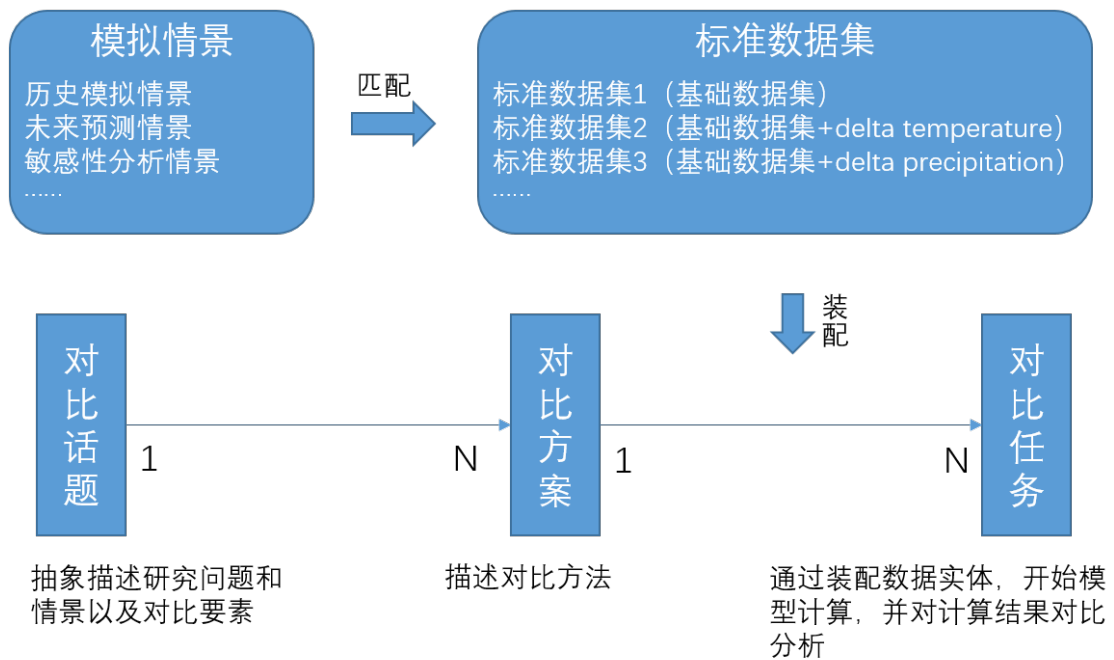


图 3.2 陆地生态系统碳循环模型对比业务抽象和归纳

陆地生态系统碳循环模型在各种情景中可以归纳出三个主要关注点：研究问题、对比方法和模型运行数据的设定。这三点可以抽象为对比话题、对比方案 and 对比任务三个层次，如图 3.2所示，对比业务的开展对应为对比话题、对比方案和对比任务创建和执行上。其中，**对比话题**中指明了研究问题、对比情景和对比要素；**对比方案**配置了对比参与者和对比方法，对比方案不包括具体的输入数据、观测数据和模拟结果，他表示的仅仅是对比的配置项，从而实现对比过程的可共享性、可重用性和可重现性；**对比任务**描述的是对比方案配置输入数据和观测数据实体的结果。模型在对比任务中开始具体的运算，并将运算结果与观测数据进行统计学对比和可视化对比。1 个对比话题可以对应多个对比方案，一个对比方案通过配置不同的标准数据集可以对应多个对比任务。图 3.3展示了对比话题、对比方案和

对比任务的详细接口设计。

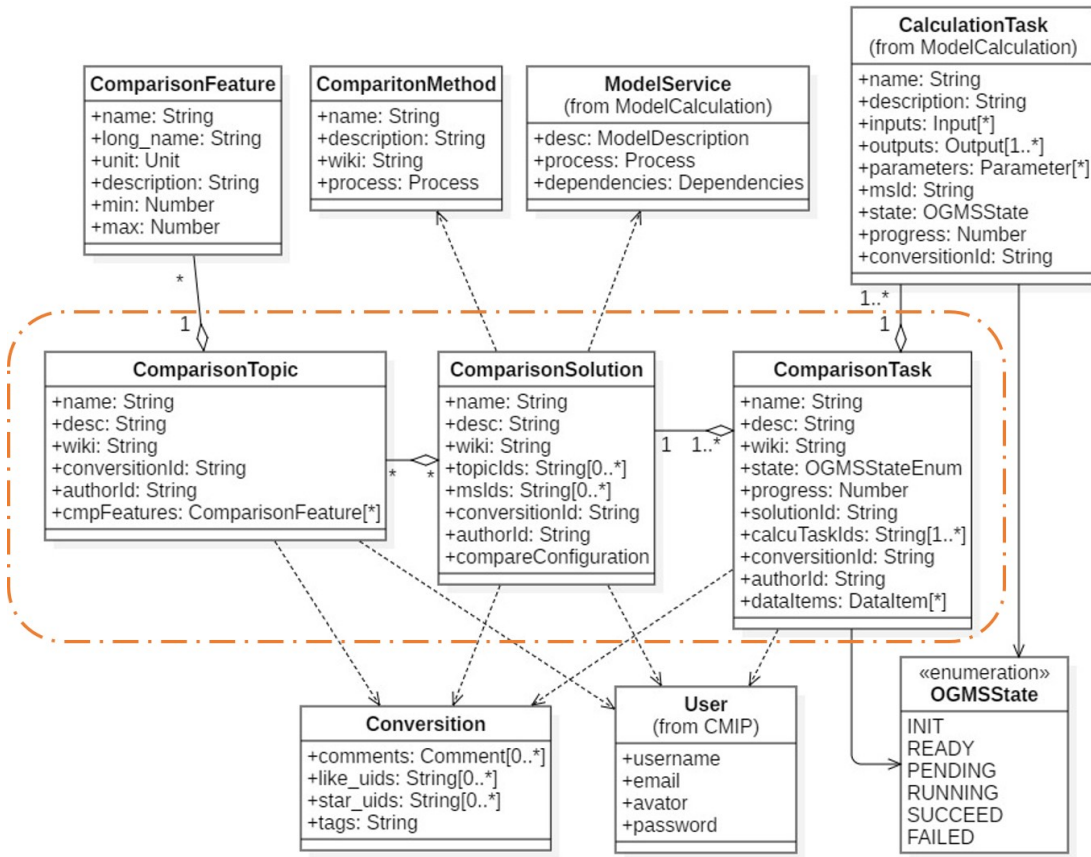


图 3.3 对比话题、对比方案和对比任务的接口设计 UML 图

以植被生产力的对比为例，如表3.2所示，针对全球历史时期的植被生产力创建一个对比话题，对比要素包括有 GPP、NPP、NEP 等，这一个话题关联两个对比方案，其中对比参与者相同，都是 IBIS、Biome-BGC 和 LPJ 三个模型，第一个方案研究的是站点尺度，它使用的对比方法包括统计学和可视化两大类，通过配置不同的站点数据可以对应 231 个对比任务；第二个方案研究的是全球范围，它使用的对比方法只有可视化对比方法，只能创建一个相应的对比任务。这样的三层抽象使对比方案可共享与重用，简化了模型对比的难度，并将最后的对比结果公开发布为服务，使对比过程公开透明，更加具有可信度。

3.1.3 对比系统功能模块设计

本文将系统的主要功能模块设计为如图3.4，功能模块分为资源模块、对比业务模块、结果展示模块和用户模块，其中资源模块包括模型资源和数据资源，对比

表 3.2 针对全球植被生产力评估的对比话题、对比方案 and 对比任务

对比话题		对比方案		对比任务
模拟情景	对比要素	对比参与者	对比方法	数据区域
历史情景	GPP	IBIS	统计学对比方法	CN-Cha（长白山）站点
	NPP	Biome-BGC	可视化对比方法	其他站点，共 231 个
	NEP	LPJ	可视化对比方法	全球

业务模块则对应于对比话题、对比方案 and 对比任务，结果展示模块基于所有对比任务的对比结果，以网站的形式发布出去。

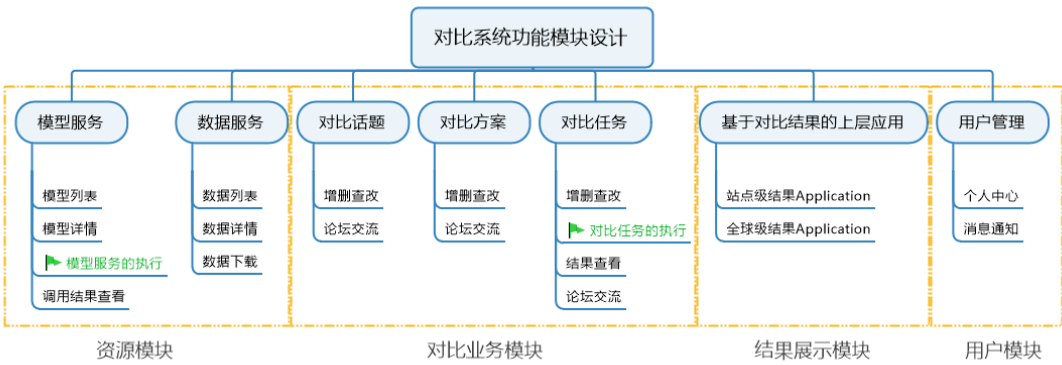


图 3.4 对比系统功能模块设计

3.2 基于服务的开放式地理资源库

3.2.1 模型服务资源库

地理模型服务资源库包括有一系列陆地生态系统碳循环模型发布出的服务，

3.2.2 数据服务资源库

数据服务资源库包括第标准输入数据集、衍生输入数据集和结果数据集。

表 3.3 模型服务 API

表 3.4 数据服务 API

表 3.5 对比要素及其单位量纲

(a) 陆地生态系统碳水循环要素表				(b) 单位量纲表	
名称	单位	最小值	最大值	单位	量纲
GPP	gCm^2d^{-1}	0	100		质量 (M)
NPP	gCm^2d^{-1}	0	100	gCm^2d^{-1}	长度 (L)
NEP	gCm^2d^{-1}	-100	100		时间 (T)
NEE	gCm^2d^{-1}	-100	100		长度 (L)
$Biomass$	gCm^2d^{-1}	0	-	mmd^{-1}	时间 (T)
R_A	gCm^2d^{-1}	0	100		
R_H	gCm^2d^{-1}	0	100		
ET	mmd^{-1}	0	100		
$Runoff$	mmd^{-1}	-	-		
LAI	1				

3.2.3 单位量纲资源库

3.2.4 数据重构服务资源库

3.2.5 可视化服务资源库

3.2.6 对比服务资源库

3.3 基于微服务的分布式网络架构设计

分布式将不同物理区域的计算资源组织整合起来，与传统的集中式相对比，分布式能够有效利用多个分布式节点上的计算能力和数据共享能力，从而提高服务端的性能。微服务是一种分布式的系统架构风格，由 Martin Fowler 提出^[54]，微服务是颗粒比较小的服务，一个大型的复杂软件可以由多个微服务组成。微服务采用 UNIX 的设计哲学，每种服务只做一件事，是一种松耦合的能够被独立开发和部署的无状态化服务。微服务通过服务来实现应用的组件化，微服务中将组件定义为可被独立替换和升级的软件单元，在应用架构设计中通过将整体应用切分成可独立部署的微服务方式进行组件化设计。这种设计方式使得应用程序效率高、灵活性强、可用性高、可扩展性强。

本文基于微服务的分布式网络架构，设计了三种服务器：计算服务器、对比服务器和资源服务器，如图3.5所示，计算服务器负责模型服务的发布、注册、管理和调用；对比服务器作为门户网站的后台服务器，负责资源的汇总与展示、对比话题和方案的创建、计算任务的分发和对比结果的展示；资源服务器负责数据相关操作，包括数据存储、数据库管理、数据服务的发布和结果数据的缓存。其中计算服务器对外提供模型微服务，资源服务器对外提供数据微服务，三者通过资源交换和消息通信完成协同完成对比任务。

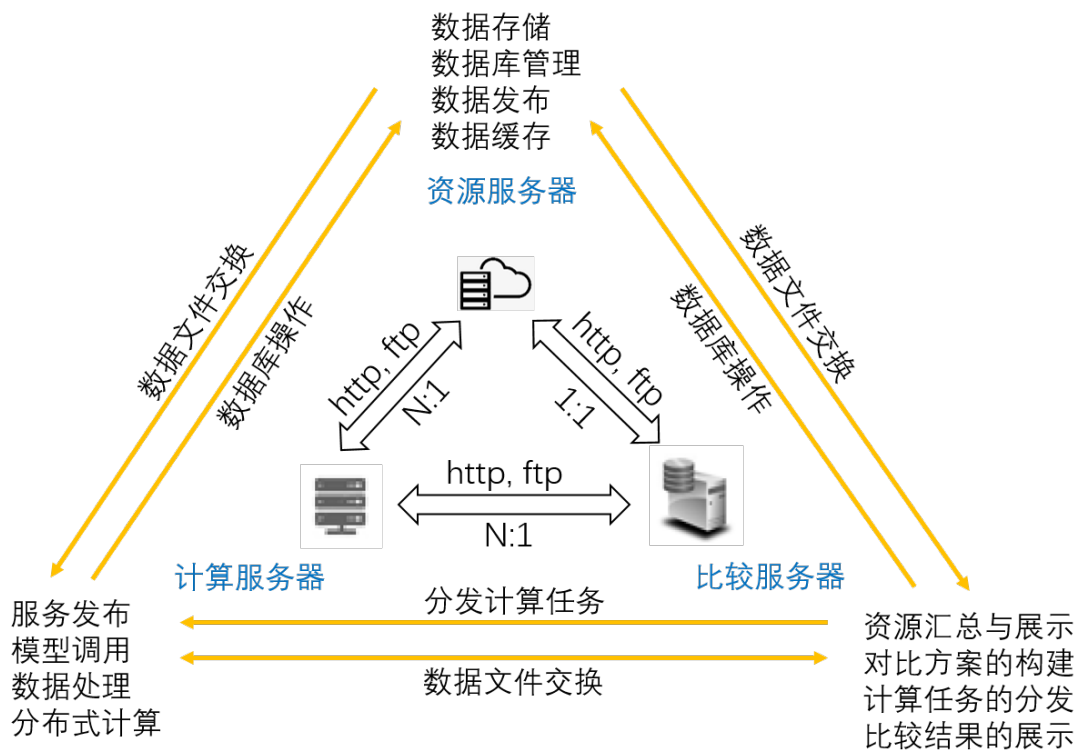


图 3.5 分布式网络架构

3.3.1 资源服务器层

资源服务器主要有以下三点功能：

(1) 数据文件存储和缓存

数据文件包括各种输入数据、对比参照数据和输出数据。模型在运行时从资源服务器请求输入数据，运行结束后将输出数据存储到资源服务器。对比任务执行时从资源服务器请求模型输出数据和对比参考数据，并将对比结果的图表存储到资源服务器。

(2) 数据库存储管理

整个对比系统的数据库存放在资源服务器上，数据库选择的是 MongoDB，通过 MongoDB Wire Protocol 协议对计算服务器和对比服务器上的客户端开放。

(3) 数据服务的发布

资源服务器使用 OGC WMS/WFS/WCS 对外发布地图服务、矢量和栅格数据服务，从而实现空间数据的查询、预览。

资源服务器作为微服务的服务提供者，将这些功能通过微服务的方式对外发布，如图 3.7 所示，计算服务器、对比服务器以及客户端浏览器以对应的协议请求微服务。针对这些功能，资源服务器的硬件配置要求是硬盘大，网络带宽高，以方便数据存储和交换。

3.3.2 计算服务器层

计算服务器的主要功能是模型服务的发布、注册、管理和调用，针对这些功能，需要在服务器上部署计算服务容器，如图 4.11 所示，计算服务容器由 Node.js 和 MongoDB 开发。Node.js 对外暴露模型调用的接口，当客户端经过 HTTP 协议发送调用请求时，在计算服务容器内部调用模型应用程序。当模型运行结束后，计算服务容器获取子程序的句柄，将运行结果的状态保存到 MongoDB 中。模型微服务与数据微服务有一些不同之处：每个独立的模型服务可以根据软硬件需求存放在不同的服务器上，即模型微服务有多计算节点的特性。因此，模型微服务发布后，还需要将服务注册到模型服务资源库中，从而能够被服务消费者发现服务。

3.3.3 对比服务器层

对比服务器的主要功能是分发计算任务，并把模型计算出来的结果文件和对比参考数据进行对比。他是模型微服务和数据微服务的消费者、对比微服务的生产者。如图 ?? 所示，模型对比的开展主要包括以下几步：

(1) 用户在浏览器客户端通过 HTTP 协议发送对比任务的调用请求；



图 3.6 资源服务器功能

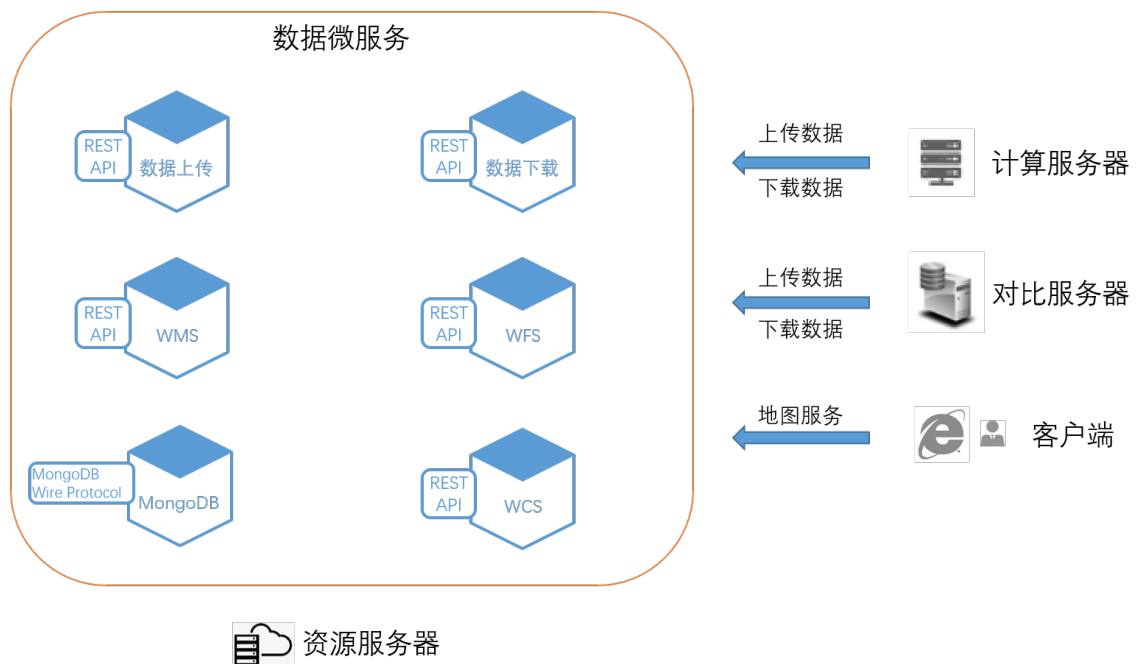


图 3.7 资源服务器微服务

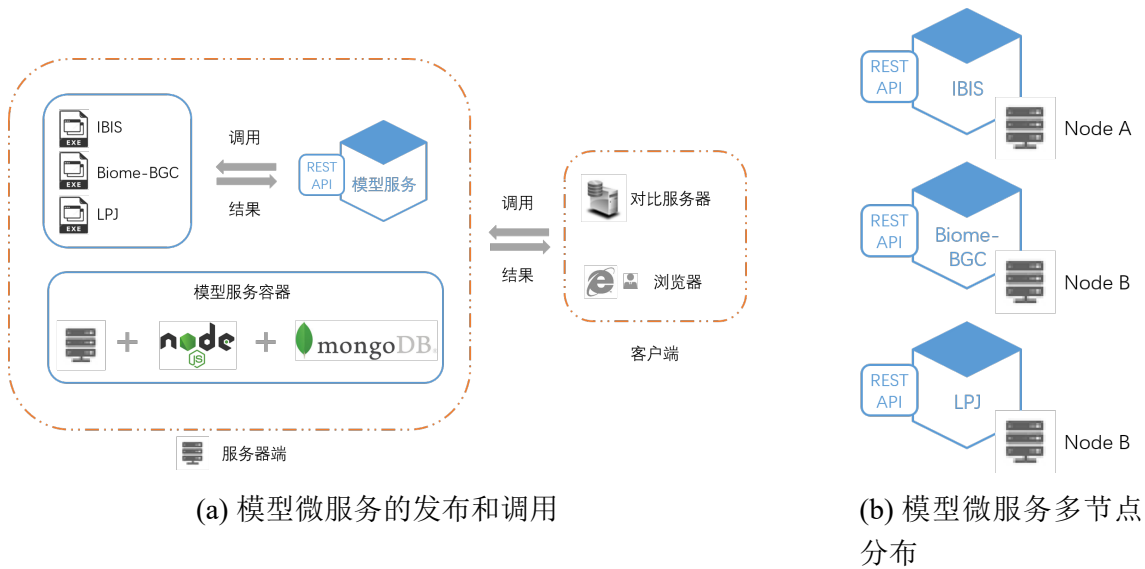


图 3.8 计算服务器微服务

- (2) Node.js 的路由器收到请求后，将对对比任务中包含的计算任务拆分出来，并通过调用计算服务器上的模型微服务启动计算任务，并不断轮询监控模型的运行进度；
- (3) 当监测到所有模型都运行成功后，从资源服务器上发布的模型运行结果微服务和对比参考数据微服务下载数据，然后调用本地的对比方法脚本程序，这些脚本可由 JavaScript、Python 或 Matlab 等编写而成；
- (4) 最后将对对比脚本的结果文件保存到数据库中，并返回给客户端。

另外他作为门户网站的后台服务器，还负责浏览器前台的一系列操作，如资源汇总与展示、对比流程的创建、对比结果的展示等。

3.3.4 服务器通信

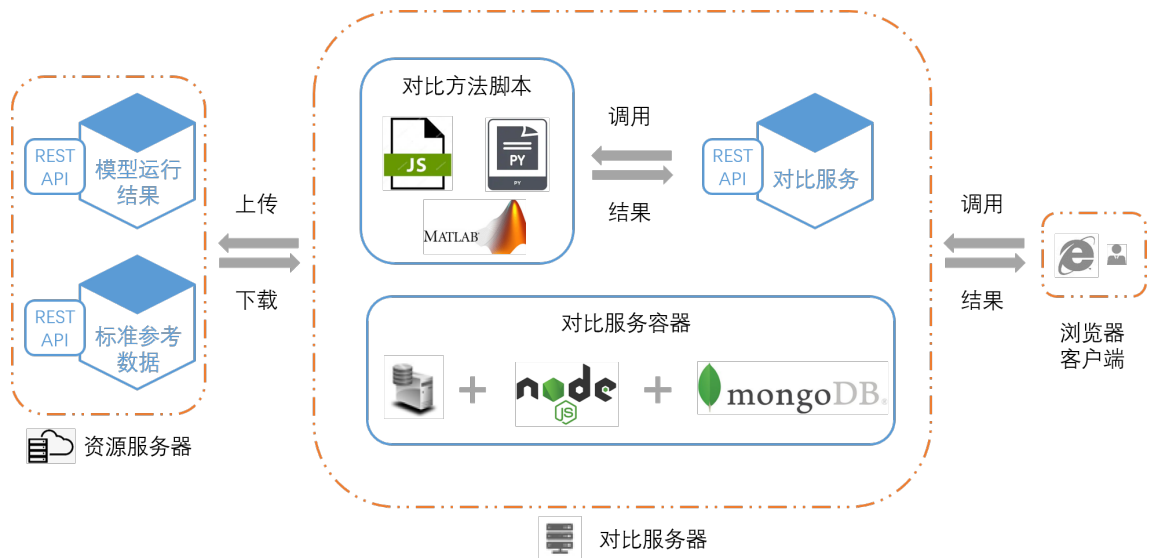


图 3.9 对比服务器微服务

3.4 开放式对比科学工作流引擎

3.4.1 对比流程分析和归纳

如图 3.10所示，

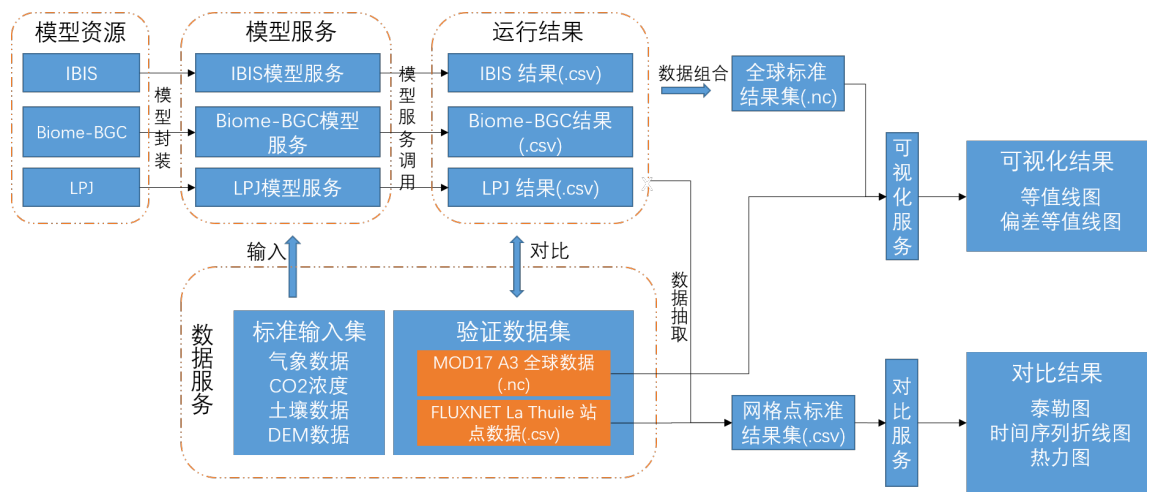


图 3.10 以 IBIS、Biome-BGC、LPJ 三个模型为例的对比流程

3.4.2 对比自动化执行引擎

工作流是一类能够完全或者部分自动执行的经营过程，根据一系列过程规则，文档、信息、或任务能够在不同的执行者之间传递、执行。科学工作流是工作流的一种，它主要面向科学实验过程，以数据驱动，用来描述和控制科学实验和过程

的执行^{[55][56]}。科学工作流的特点是面向数据、数据规模大、具有动态适应性以支持工作流执行过程中的动态资源绑定。本文的对比工作流执行流程如图 3.11所示，其中数据服务库中的数据量巨大、存储节点分散，模型服务库中的模型可以动态添加和删除，并通过标准化的接口暴露出来。工作流首先调用模型服务，然后将模型运行出的结果和观测数据交给数据重构服务进行数据重构，将两者的单位量纲统一，并抽取出需要的那一部分，最后通过对比服务对数据进行统计学对比，使用可视化服务对数据进行统计图表和地图形式的展示。

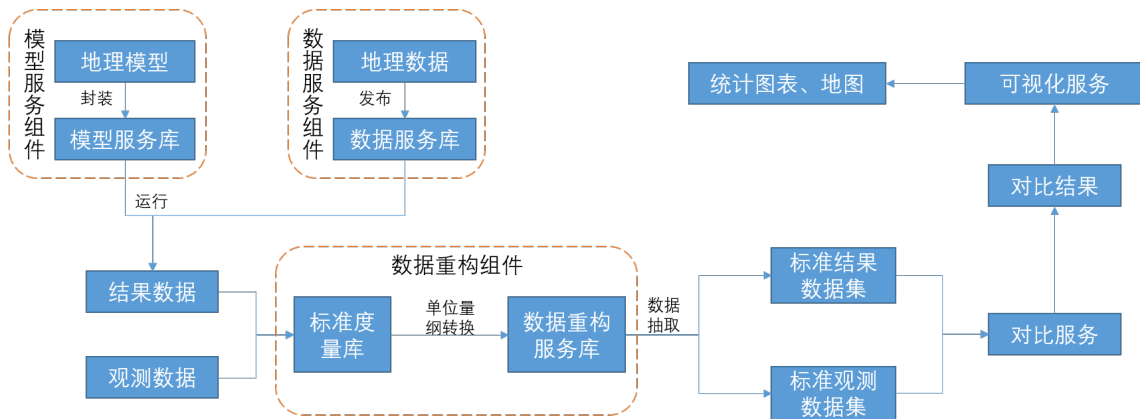


图 3.11 开放式对比科学工作流流程图

从微服务的视角来看，如图 3.12所示，对比科学工作流的各个数据处理过程都是一个微服务，并分散在分布式的网络节点上，对比工作流在执行时不用关心数据处理程序的具体网络节点位置，通过微服务暴露的 IP 地址和端口来寻找计算节点，这样就将分布式节点有效管理起来。

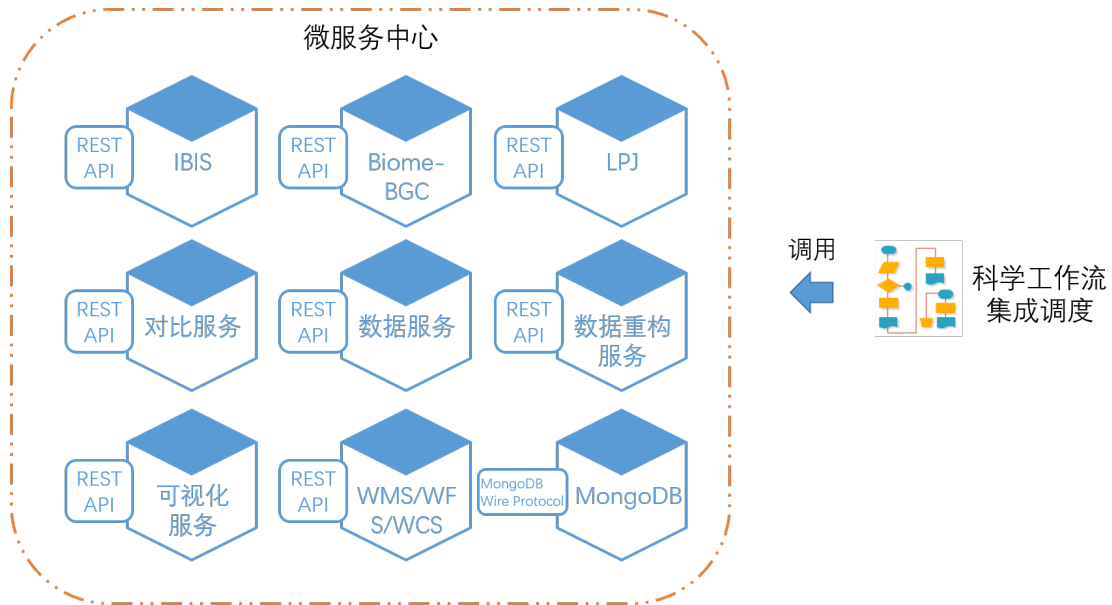


图 3.12 对比科学工作流引擎和微服务

3.5 本章小结

陆地生态系统碳循环模型的开放式对比系统有两个关键问题：一是如何设计一个开放式的对比框架，二是这些对比资源如何接入框架，对于问题一，本章从对比业务、资源组件、网络架构和执行引擎四个角度进行了详细论述，设计的系统框架具有以下特点的：

- (a) 模型计算和对比可共享与重用；
- (b) 模型计算和对比过程公开化；
- (c) 模型和数据资源可扩展性地动态接入；
- (d) 大规模计算场景下稳定可用；
- (e) 以自动化实现其易用性。

对于问题二将在下一章详细论述。

第 4 章 开放式对比资源接入方法

本文将数据、模型和对比方法称之为对比资源，这三者的开放性决定了对比结果的准确性、公开性和公平性。其中模型作为对比的对象，是本系统中最重要要素之一。传统的模型对比方案将模型部署在本地环境运行，针对第 3 节分析的多对比情景需求，会造成难以共享和重现等问题，本章描述的模型服务化封装、部署和发布对此提出了对策；而数据资源作为支撑模型成功运行的必备要素，传统的基于数据拷贝的方式在这种情景下也不能满足要求，因此本章对数据资源的开放式接入也提出了相应的方案；在这两者的基础之上，将统计学对比方法和可视化对比方法也一并接入到框架之下，实现了模型的开放式对比。

4.1 开放式碳循环相关数据资源接入方法

网络空间中的数据按照其组织结构可以分为结构化数据、半结构化数据和非结构化数据，地理模型在运行时强烈依赖于输入数据的结构，所以半结构化和非结构化数据的半结构化描述是数据应用时重要部分。本节首先从数据格式、尺度、编排三个方面详细地分析了碳循环相关数据资源的特征，对领域通用数据设计了结构化的描述接口，以帮助数据资源介绍、接入、匹配和展示。最后遵循 OGC WMS、WFS、WCS 标准，使用 GeoServer 发布碳循环数据服务，实现了空间数据的在线预览和查询检索，对于对比过程中需要进行的数据抽取和重组，设计了数据重构服务以支持他的网络化应用。

4.1.1 碳循环相关数据资源结构特征分析

4.1.1.1 数据格式分析

碳循环相关数据主要有几种格式：NetCDF、CSV、Shapefile 和 TXT，如表 4.1 所示，这些数据都是非结构化的。

表 4.1 数据格式和特征

数据格式	数据类型	维数	数据名称
Shapefile	矢量数据	2 维	网格点
			观测站点
NetCDF	场数据	n 维	DEM
			MERRA 2
			Soil
			PFT
			全球计算结果
CSV	二维表	2 维	网格点输入气象数据
			网格点计算结果
TXT	文本数据	1 维	CO ₂ 浓度
			PFT
			Biome-BGC 的运行配置

(1) NetCDF

NetCDF (network Common Data Form) 网络通用数据格式是由美国大学大气研究协会的 Unidata 项目科学家针对科学数据的特点开发的, 是一种面向数组型并适于网络共享的数据的描述和编码标准, 已被作为 OGC 的一项标准使用。NetCDF 数据通常用来表示多维场数据, 从数学上来看, 它存储的数据就是一个多自变量的单值函数: $f(x, y, z, \dots) = value$ 。其中 x, y, z 等在 NetCDF 中被称为维 (Dimension), 函数值 $value$ 被称为变量 (Variables), 维和变量在物理学上的一些性质, 被称为属性 (Attributes)。当维度只有经纬度时, 可以用来表示栅格数据, 如本文中的植被功能类型数据、DEM 数据、土壤质地数据。维度除了经纬度以外通常还有时间维, 如本文中的 MERRA 2 气象数据和全球范围内的模型计算结果。具体的维度和变量信息如表 4.2

表 4.2 NetCDF 数据的维度和变量列表

数据名称	维度	变量
DEM	经度、纬度	高程
MERRA 2	经度、纬度、时间	最低温、最高温、平均温、相对湿度、降水、风速、云量
Soil	经度、纬度	沙粒含量、黏粒含量、粉粒含量
PFT	经度、纬度	植被类型编号
全球计算结果	经度、纬度、时间	GPP、NPP、NEP、Biomass、ET

(2) CSV

CSV (Comma-Separated Values) 逗号分隔符文件是以文本形式存储的表型数据。CSV 文件由任意数量的记录组成，每行称之为一条记录，每条记录由多个字段组成，字段之间由分隔符分割，通常是逗号或制表符。CSV 文件有一系列的规则，比如可以选择性的包含列名，开头不能留空行，每条记录不能跨行等。本文中的 CSV 文件有网格点气象数据、网格点结果数据和站点观测数据，他们除了具备常规 CSV 文件的特点以外，还都用一列来表示时间。

(3) Shapefile

Shapefile 是 ESRI 开发的一种矢量空间数据格式，也属于 OGC 的一项标准。Shapefile 文件用点线面描述空间对象的几何信息，用属性表描述对象的属性，另外还包括有数据的空间参考坐标、图形索引等。本文中用到的 Shapefile 数据是由 $0.5^{\circ} \times 0.5^{\circ}$ 的经纬网划分而来的网格点。

(4) TXT

TXT 文件是最常用的文件格式，没有固定的结构。

4.1.1.2 数据尺度分析

尺度是地理学数据的重要特征，指数据集表达的时空范围和精度，不同尺度数据表达的信息密度有很大差异。

(1) 时间尺度

主要指数据的时间范围和精度，碳循环的数据和模拟有很强的多样性，精度上可细分为日尺度、月尺度、季节尺度、年尺度、百年尺度等，范围上包括 1 年、多年甚至几十年。本文的模拟全部在日尺度上进行，模拟时间范围为 1982 年到 2014 年，观测数据的时间范围视具体站点而定，一般为 1-10 年左右。在对比时，由于日尺度的时间序列长度比较大，且数据噪点比较多，因

此,也需要对其时间分辨率进行调整,以月尺度、季节尺度和年尺度对其进行平滑,可以提高数据的精度。

(2) 空间尺度

主要体现在数据的空间范围和分辨率上。如表 4.3所示,各种输入数据的分辨率和范围不尽相同,本文统一重采样为 $0.5^{\circ} \times 0.5^{\circ}$, 并以其划分网格在全球范围内模拟。

表 4.3 碳循环模拟的空间多尺度特征

数据名称	经度分辨率	纬度分辨率	北纬边界	南纬边界
模拟输入和输出	0.5°	0.5°	82.25°	-54.75°
MERRA 2	0.625°	0.5°	90°	-90°
DEM	0.5°	0.5°	89.75°	-89.75°
土壤	0.0085°	0.0107°	90°	-90°
植被功能类型	0.66°	0.5°	90°	-90°
MODIS 17 A3 GPP/NPP	0.0083°	0.0083°	80°	-60°

4.1.1.3 数据编排分析

由于三个模型对于数据的要求不完全相同,所以合理的数据编排对数据检索、下载、模型运行都是至关重要的。如图 ??所示,每套数据集存放在固定的路径下,可由数据库解析到。在运行一个格点时,首先从网格点 shapefile 的属性表中获取他的编号,再对应到具体的数据集和模型下面,可以获取详细的输入输出文件列表。

4.1.2 碳循环相关数据资源描述方法

元数据是关于数据的数据,它主要描述数据的内容、质量、表示方式、空间参照系、管理方式、数据的所有者、数据的提供方式和数据集的其他特征等。地理信息元数据可以用来管理和维护空间数据,方便查询和检索空间数据,为用户提供数据内容,质量状态和相关特性的数据编目,为地理信息共享提供了基础。对于数据的描述,国内外都有很多对应的研究,OGC 采用 XML Schema (如 GOC Simple Feature 规范)来描述 GML,来支持 WPS 的运行^[57]。OpenMI (Open Modelling Interface)设计了一套标准接口用于多模型集成耦合时对数据的定义、描述和传递^[58]。乐松山设计的通用数据表达——交换模型 (Universal Data Description eXchange Model, UDX),结合数据映射服务和数据重构服务,面向模型集成应用场景可以对数据结

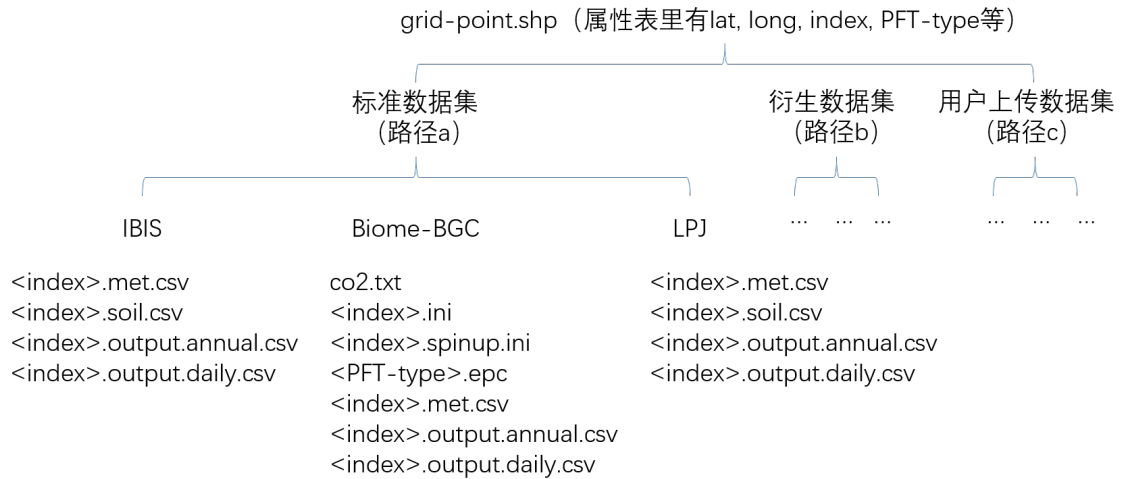


图 4.1 陆地生态系统碳循环模型输入数据编排

构进行结构化、语义化描述^[59]。

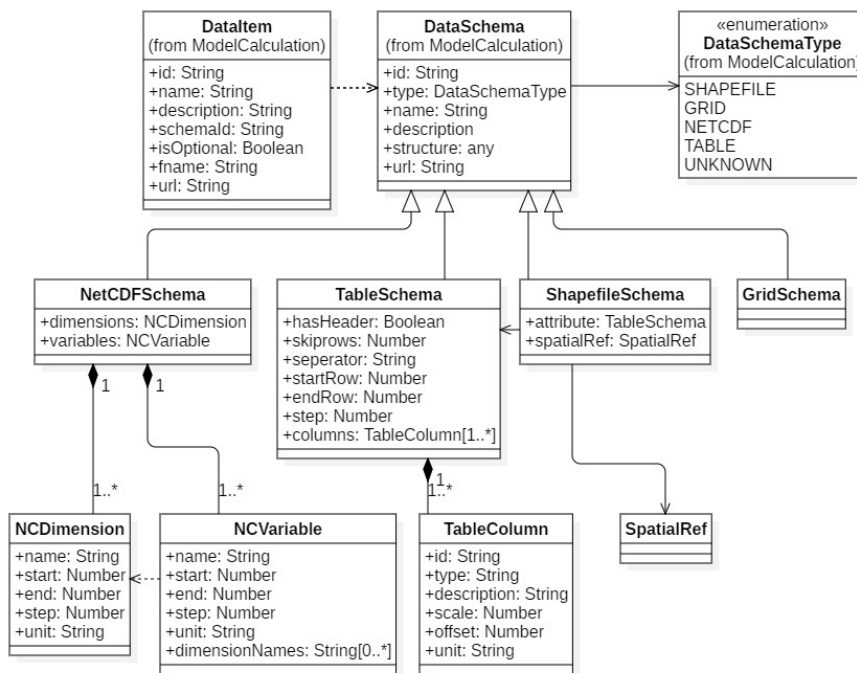


图 4.2 数据描述接口设计

为适应本系统的需求，本文并没有采用这些数据描述标准，而是采用自定义结构进行结构化描述。对于数据的描述分为两部分：领域通用格式数据的描述和自定义结构数据的描述。对于前者，本文设计了如图 ?? 所示的描述文档，针对每种格式数据各自的特点，以可扩展的形式对其进行描述。对于自定义结构的数据，本文采用示例数据的形式描述。

如图 4.3, 对于 Table 数据的描述, 主要包括解析数据的表头和列信息组成, 表头信息包括有无表头、跳过行数、分隔符、起始列号、终止列号、步长信息。列信息包括数据缩放因子、偏移量、描述和单位信息组成。对于 NetCDF 数据的描述, 包括维度和变量信息。他们都包括名称、缺省值、缩放因子、偏移量等信息。

Table数据描述示例	{ 9 fields }
type	table
header	null
skiprows	0.0
separator	\\s+
start	0
end	11680
step	365
timeUnit	days since 1982-01-01
columns	[4 elements]
0	{ 7 fields }
id	GPP
metricName	GPP
scale	1000.0
offset	0.0
type	number[]
description	
unit	gC m-2 d-1
1	{ 7 fields }
2	{ 7 fields }
3	{ 7 fields }

NetCDF数据描述示例	{ 5 fields }
type	NETCDF4
dimensions	[3 elements]
0	{ 5 fields }
name	lat
start	-54.75
end	82.25
step	0.5
unit	degrees_north
1	{ 1 fields }
2	{ 1 fields }
variables	[7 elements]
0	{ 6 fields }
1	{ 6 fields }
2	{ 6 fields }
3	{ 8 fields }
name	GPP
dimensions	[3 elements]
missing_value	null
metricName	GPP
scale	1
offset	0.0
unit	gC m-2 d-1
layerid	Carbon_Cycle:Biome-BGC-annual-GPP

图 4.3 Table 和 NetCDF 数据描述文档示例

4.1.3 碳循环相关数据资源的服务化封装

4.1.3.1 WMS、WFS、WCS 三种服务的发布

为实现一套与厂商无关的空间数据互操作规范, 开放地理空间信息联盟 (OGC) 制定了 WMS、WFS、WCS 服务标准, 具体接口如表 4.4 所示。其中 WMS (Web Map Service) 利用具有地理空间信息的数据制作地图, 在国际规范中, 地图是地理数据的可视化表现, WMS 返回的地图并非是地图数据, 而是地图图像, 格式上可以是 PNG、GIF、JPEG、SVG 等^[60]。WFS (Web Feature Service) 通过 GML (Geography Markup Language) 传递地理空间数据, 它支持在基于 HTTP 协议的分布式计算平台上对地理要素进行增删查改等操作, 并在这些操作过程中保证了地理数据变化的一致性^[61]。WCS (Web Coverage Service) 在网络上以覆盖 (Coverage) 的形式共享地理空间数据, 能够返回栅格时空范围中任意指定点的值, 实现了栅格影像数据集的共享^[62]。本文将网格点的 Shapefile 发布为 WFS, 通过界面上的交互查询到选择的格点要素, 解析其属性后可以获取输入数据路径。将全球范围下的计算结果和对比结果 NetCDF 数据发布为 WMS 和 WCS, 可以在线查看和下载。

表 4.4 OGC WMS、WFS、WCS 接口

类型	接口	描述
WMS	GetCapabilities	返回服务级元数据，包括对服务信息内容和可接受请求参数的描述
	GetMap	返回一幅地图影像
	GetFeatureInfo	返回地图上要素的空间实体信息
WFS	GetCapabilities	返回服务级元数据
	DescribeFeatureType	返回表示要素结构的 XML 文档
	GetFeature	返回 GML 形式的要素实例
	Transaction	增删查改事务请求
	LockFeature	在事务期间对一个或多个要素实例上锁
WCS	GetCapabilities	返回描述服务和数据集的 XML 文档
	DescribeCoverage	返回对覆盖的详细描述
	GetCoverage	使用覆盖格式（图片）返回地理位置的值或属性

详细接口参考 OGC^{[60][61][62]}。

4.1.3.2 数据下载服务

数据下载服务主要是针对图 4.1 设计的数据编排方式，在数据下载时，对数据路径进行反编码，查询到具体的数据条目，并通过 HTTP 协议返回给用户。

4.1.3.3 数据处理服务

如图 4.4 所示，本文用到的数据处理服务有三个：

(1) 单位量纲转换服务

IBIS、Biome-BGC、LPJ 三个模型的输出项单位不完全相同，比如 GPP 的单位有 $gC \cdot m^2/d$ 、 $kgC \cdot m^2/d$ 、 $kgC \cdot m^2/a$ 等，单位量纲转换服务以单位量纲资源库为参考，对三个模型输出数据项进行统一单位。

(2) 数据抽取服务

模型输出的数据步长是 1 天，而观测数据的时间范围往往有 1-10 年，时间序列过长，而且日间隔的数据噪点过多，通过数据抽取服务可以将其抽取为步长为 1 月的数据。

(3) 数据重组服务

模型输出数据是网格点的 CSV 数据，对于全球范围的数据可视化处理起来不够方便，因此使用数据重组服务将全球 40595 个格点的数据重组到一起，

得到一个 NetCDF 数据。

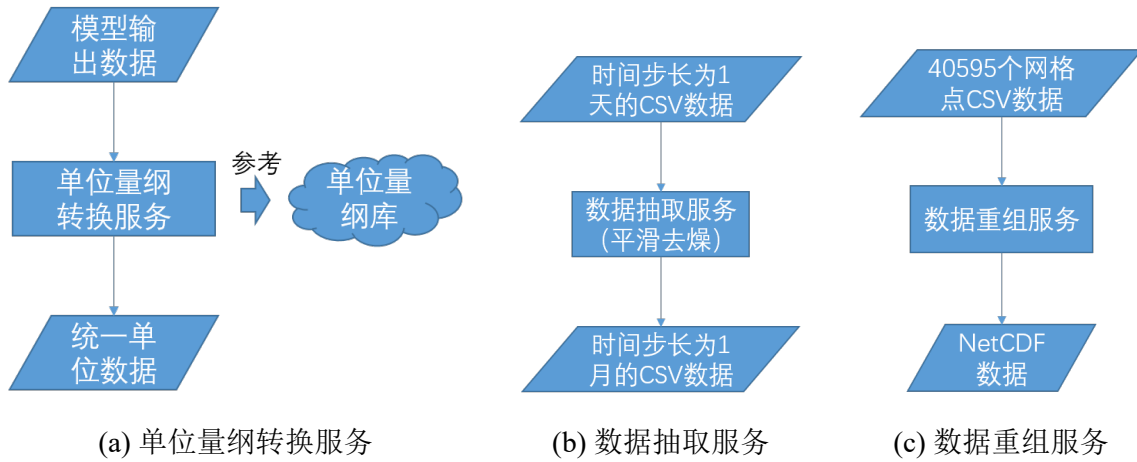


图 4.4 数据处理服务

4.2 开放式碳循环模型资源接入方法

4.2.1 碳循环模型资源特征分析

(1) 模型参数策略特征

生态过程模型普遍采用植被功能类型（PFT）作为植被基本处理单元，通常将 PFT 划分为林地（覆盖热带、温带、寒带，常绿、落叶、针叶和阔叶等）、灌木（常绿和落叶等）、草地（ C_3 、 C_4 等）、农田（水稻、玉米、温带谷物、大豆、热带根系作物、太阳花、花生、油菜等）等类型。每种植被功能类型对应有许多植被生理生态参数。

(2) 模型异构性分析

地理模型从开发视角上有很多不同的分类体系，如表 4.5 所示，这些类别中，本文暂不考虑有用户交互的模型。

4.2.2 碳循环模型资源描述方法

模型服务的描述方法是模型服务化的基础。模型服务描述文档应该是面向人的，通过模型服务描述文档提供的信息，模型服务使用者就能够方便地了解和使用模型，并能够理解模型的返回结果；同时模型服务描述文档又是面向机器的，通过模型服务描述文档提供的模型运行输入输出选项，能够判断出模型运行数据的完备性，并在用户发送请求时将模型调用起来。因此，本文将模型服务的描述文档设计为如图 4.5。其中 AttributeSet 表示基本信息描述接口，面向人的理解；Behavior

表 4.5 开发视角上的模型分类

分类标准	一级分类	二级分类
语言	编译型	C
		C++
		C#
		Fortran
		Java
	解释型	Python
		MATLAB
		JavaScript
平台	Windows	
	Linux	
	Unit	
交互	无交互	
	图形用户界面交互	
	控制台用户界面交互	
表现形式	源代码	
	可执行程序	

表示模型的运行行为接口，面向机器理解模型的运行接口；**Runtime** 表示模型的软硬件环境依赖，也是面向机器，表示模型所依赖的软硬件环境信息，从而支持模型能够成功调用起来。模型服务描述的难点在于对于模型依赖数据的结构的有效表达和对于运行输入输出的合理描述。对于前者，本文采用第 4.1.2 的方法，即采用领域通用标准数据的描述加上示例数据具体展示的形式，对于后者，将在第 (2) 节介绍。

模型描述文档由 JSON 形式表达，与本文选择的具体技术体系有关：一方面它与 MongoDB 数据库结构同构，可以直接存储于数据库中，有助于增删查改，另一方面它与后台服务器 Node.js 天生兼容，解析起来方便。

(1) 基本信息描述接口设计

基本信息的描述如图 4.6 所示，包括模型名称、作者、关键字、所属分类体系、DOI、wiki 和计算节点信息。其中计算节点信息描述了模型服务的 IP、端口和路由前缀等信息，从而在模型请求时能够解析出对应的 URL。而 wiki 的设定允许模型服务使用者也能够维护模型服务的介绍，在经过模型服务拥有

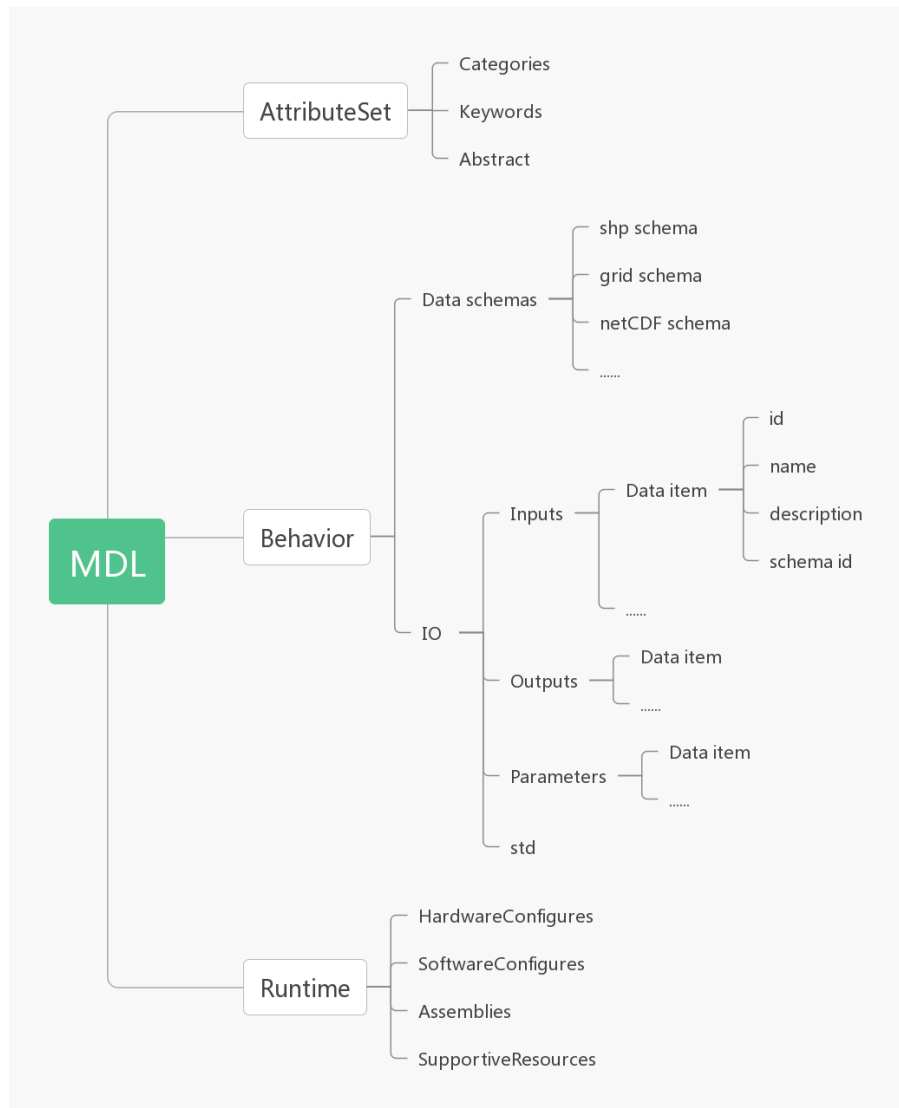


图 4.5 模型描述文档设计

者的审核后，会补充到模型服务的描述文档中，这样就可以弥补描述接口的不足之处。

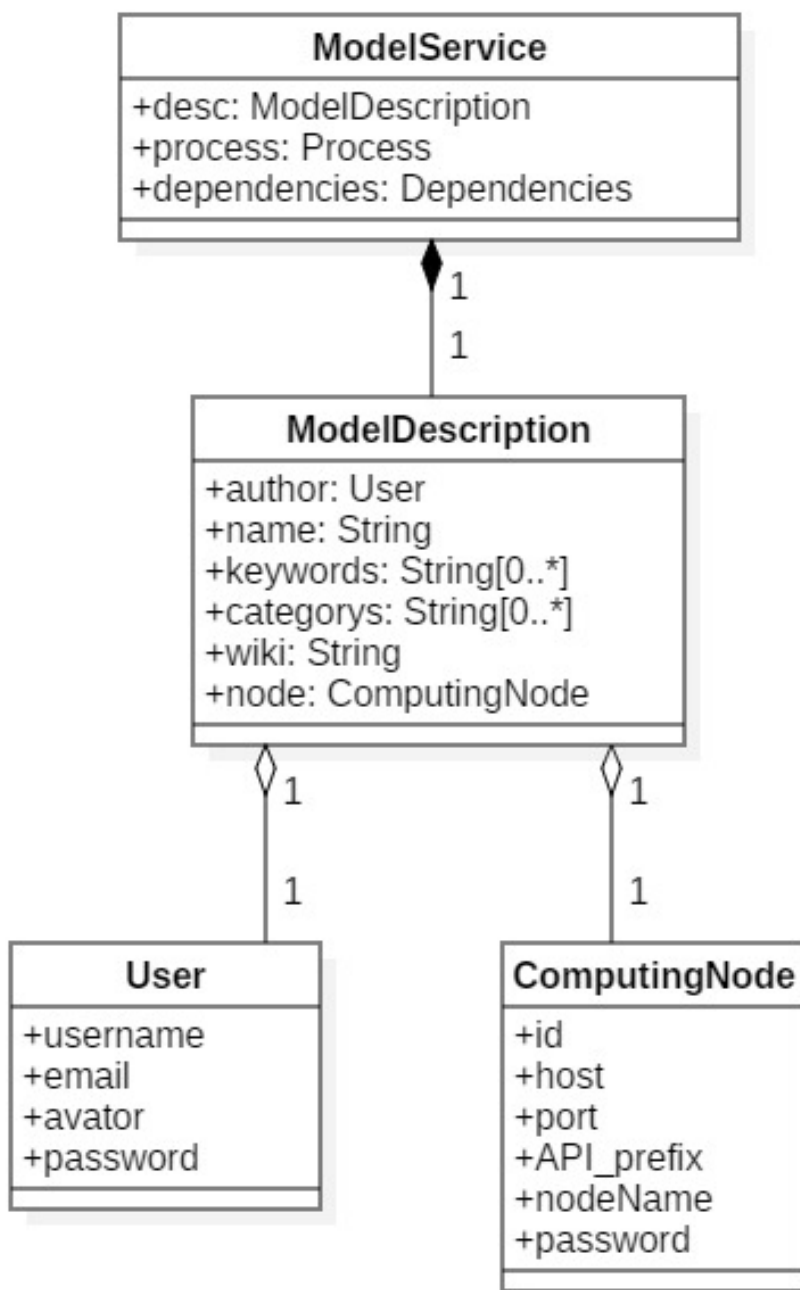


图 4.6 基本信息描述接口设计

(2) 运行信息描述接口设计

运行信息的描述接口如图 4.7 所示，包括模型的可执行程序文件名、程序类型、解释器类型、输入数据、输出数据和参数以及各种数据的结构化元数据

描述。本文将模型从开发语言上分为编译型和解释型两种，对于解释型的语言，需要记录模型的解释器，在运行时通过解释器调用才能运行。通过运行信息的描述，可以在用户调用时执行 `execute()` 函数，并返回子进程的 `pid`，在需要时可以通过 `kill(pid)` 杀死进程。

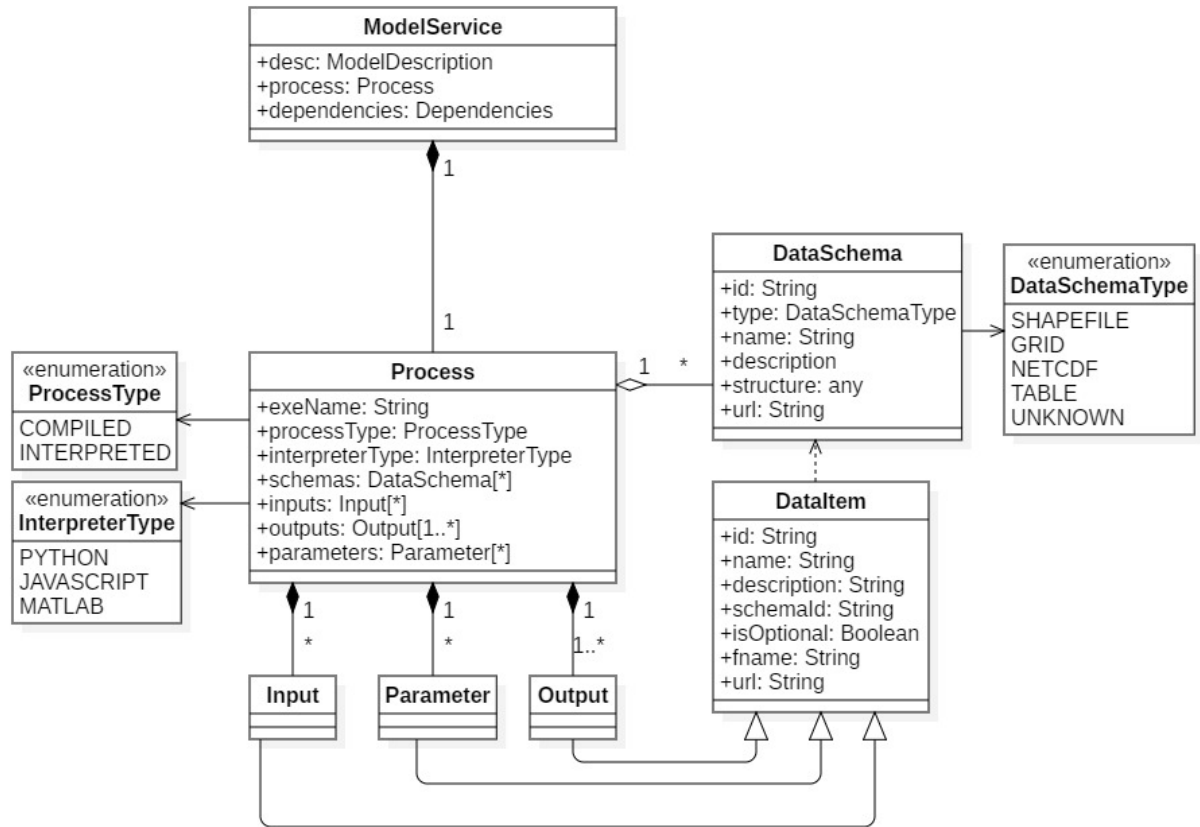


图 4.7 运行信息描述接口设计

(3) 部署信息描述接口设计

部署信息的描述接口如图 4.8 所示，包括模型运行所需的软件和硬件依赖环境。通过与计算节点的软硬件环境数据库相匹配可以得到部署环境是否匹配，当环境相匹配时才能调用成功。

4.2.3 碳循环模型资源的封装、部署、发布

4.2.3.1 服务封装

在计算机领域，封装是对实体进行包装或组合形成具有稳定接口的新实体的方法和技术。^[15] 通过封装，可以屏蔽原有程序的实现细节，提高模型的可复用性和可维护性，降低用户的使用难度。封装可以体现在程序设计的各个层次上，如

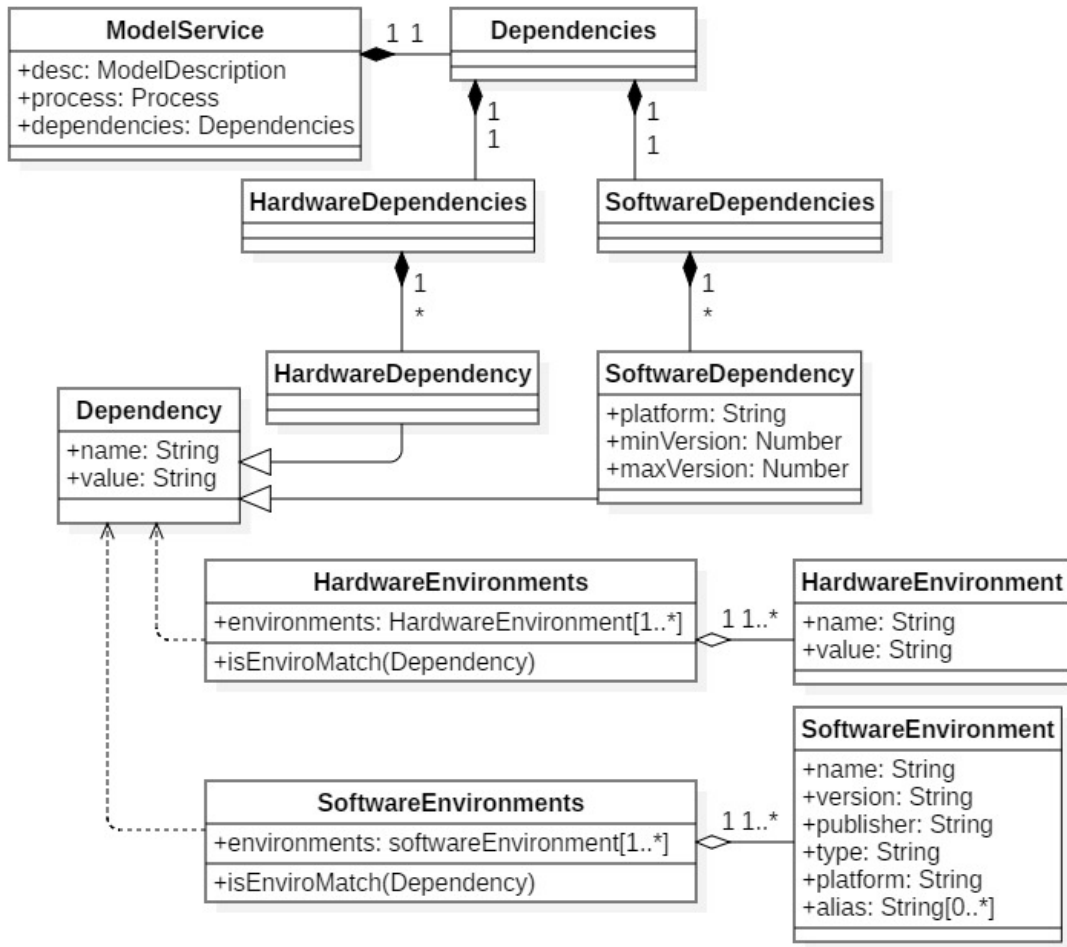


图 4.8 部署信息描述接口设计

函数、类、模块、组件、系统、可执行程序 and Web 服务这些粒度上都可以进行封装，本文的封装将模型包装成具有标准接口的控制台应用程序。

模型封装最关键的是运行接口的设计，而模型运行接口的设计本质上是设计一个命令行参数解析标准，将输入输出选项映射给程序内存变量，业界有很多成熟的标准，如 libc 的 `getopt/getopt_long` 接口、google 的 `gflags` 接口等。本文采用的 `getopt/getopt_long` 接口，它在很多其他语言上都有对应的移植版本，方便对不同语言模型封装的实现。`getopt/getopt_long` 接口支持单字符选项、多字符选项和可选参数选项，可以方便地生成帮助信息。通过 `getopt/getopt_long` 的标准化，模型的调用命令可表示为 “`exeName -a -i=“filepath1” -p=“10” -o=“filepath3”`” 或 “`exeName -a -inputFile1=“filepath2” -parameter1=“10” -outputFile1=“filepath3”`”。其中选项列表来自于运行信息描述文档，以 “-” 开头表示单字符选项，以 “-” 开头表示多字符选项，不带等号的如 “-a” 为参数选项。对于 IBIS 的封

装，标准化后调用方式可表示为 “IBIS -meteorology=”meteorology_file_path” -soil=”soil_file_path” -output=”output_file_save_path””。对于解释型的语言，需要在命令前面加上解释器，如 Biome-BGC 模型的调用方式可表示为 “node Biome-BGC.js -a -ini=”ini_config_path” -met=”meteorology_file_path” -co2=”co2_file_path” -epc=”PFT_parameters_file_path” -output=”output_file_path””。

模型封装根据有无源码可以分为源码封装和模型代理封装，如图 4.9所示，源码封装直接对模型输入输出接口做出修改使其标准化，模型代理封装通过写一个模型代理壳将 IO 标准化，然后通过子进程启动模型完成封装。按照模型是否包含集成分为简单模型封装和包含集成流程的模型的封装。Biome-BGC 模型的执行过程包括 spin-up 模式和正常模式，整个运行过程中有一个简单的串联集成过程，如图 4.10所示，正常模式的运行依赖于 spin-up 模式，通过 Node.js 脚本可以非常方便的串联集成。

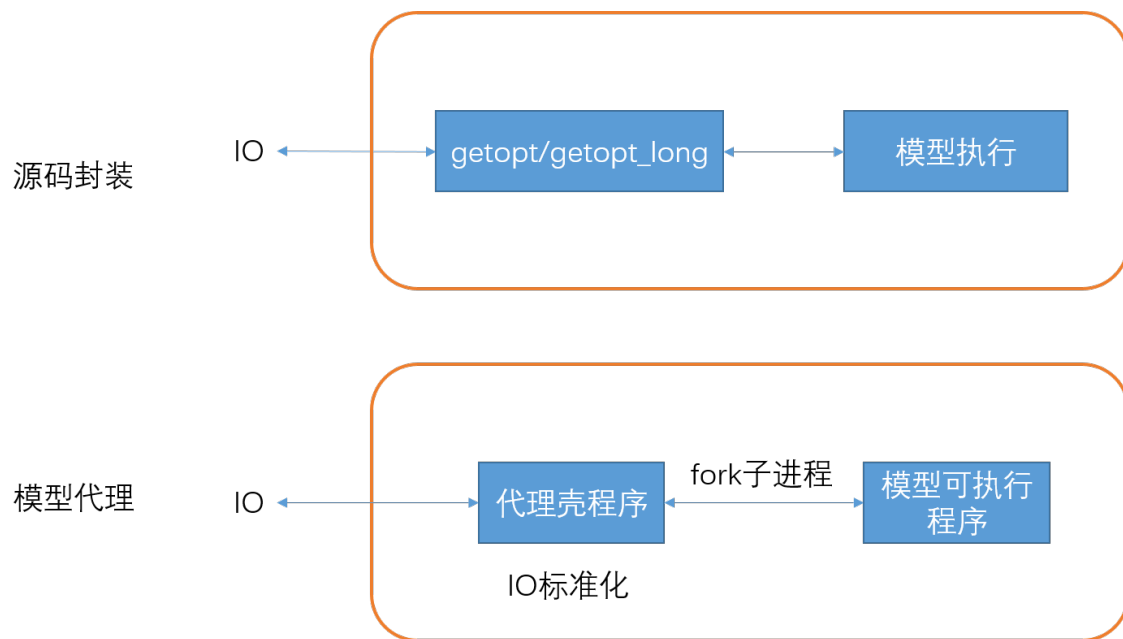


图 4.9 源码封装和模型代理封装

4.2.3.2 服务部署和发布

服务部署和发布是将封装好的模型程序迁移到目标计算节点上的过程，主要包括模型服务描述文档数据库条目的录入和模型运行软硬件环境的匹配。对于模型服务描述文档，本文将数据库设计与第 4.2.2 节的描述接口同构，可以直接录入。对于软件环境。。。对于硬件环境，本文所用的三个模型对硬件要求都不高，可以直接部署在普通 PC 机上。

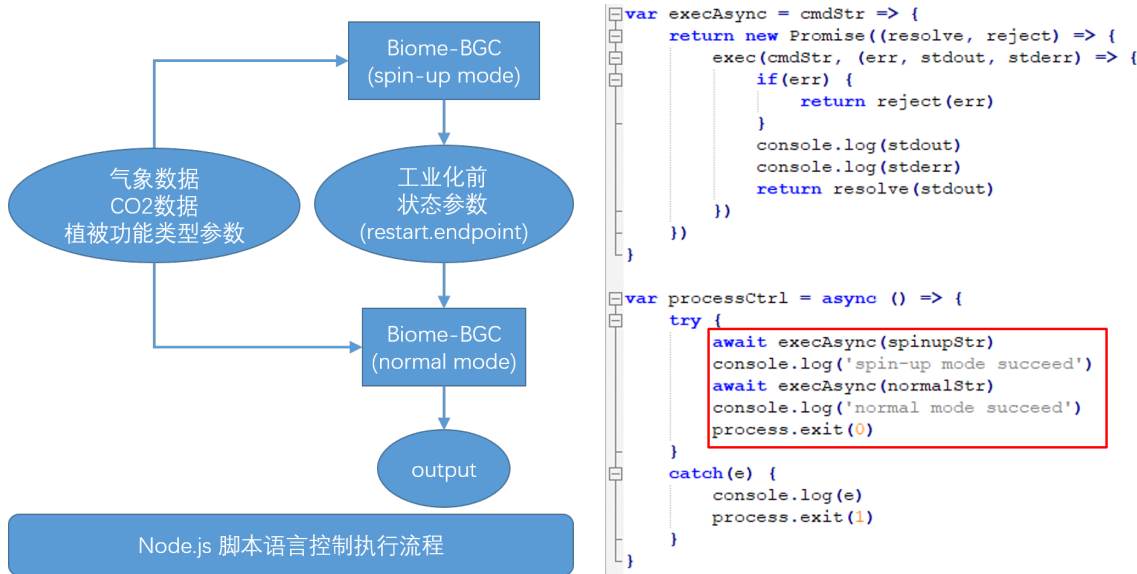


图 4.10 模型的串联集成封装，以 Biome-BGC 为例

4.3 开放式碳循环模型对比方法

4.3.1 碳循环模型对比方法

4.3.1.1 统计学对比方法

$$SD = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}} \quad (4-1)$$

$$MSE = \frac{\sum_{i=1}^N (y_i - Y_i)^2}{N} \quad (4-2)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - Y_i)^2}{N}} \quad (4-3)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (y_i - Y_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (4-4)$$

本文主要用以下几个统计变量评价 IBIS、Biome-BGC、LPJ 三个模型在网格点的模拟准确度：

(1) 标准差 (Standard Deviation, 记为 SD)

标准差又称均方差，用来表示一组数据的离散程度，如公式 4-1 所示。

(2) 均方根误差 (Root Mean Square Error, 记为 $RMSE$)

均方根误差又称标准误差，表示有限观测数据中，观测值与真值的偏差。本文使用均方根误差表示网格点模拟值与观测值的偏差，从而反映模型的模拟

能力。均方根误差可表示为公式 4-3，其中 $RMSE$ 表示均方根误差， MSE 表示均方误差， $y_i (i = 1, 2, \dots, N)$ 为模拟数据， Y_i 为观测数据。

(3) 效率系数 (Nash-Sutcliffe coefficient, 记为 NSE)

效率系数用来评价模型模拟的好坏^[63]，如公式 4-4所示， NSE 等于 1 减去均方误差与观测值变异的比值。 NSE 值的范围从负无穷大到 1，越接近 1 表示模型效果越好。如果模拟值的均方误差与观测值的方差相同，则 NSE 等于 0；如果模拟值的均方误差大于观测值的方差，则 NSE 小于 0；如果模拟值的均方误差小于观测值的方差并趋近于后者，则 NSE 趋近于 1，表示模型具有很好的模拟能力。

(4) 超级集合

超级集合由 Krishnamuti^[64] 等提出，通过多元线性回归分析对每个观测站点的数据进行分析，各模型的权重系数，如公式 4-5所示：

$$F_t = \bar{O} + \sum_{i=1}^N a_i (F_{i,t} - \bar{F}_t) \quad (4-5)$$

式中， F_t 是超级集合合成后的预测值； \bar{O} 是训练期的平均观测值； $F_{i,t}$ 是第 i 个模型的预测值； \bar{F}_t 是第 i 个模式在训练期的平均预测值； n 是参加建立超级集合的模型数量； t 是时间； a_i 是回归系数，代表各个模型的权重。在建立超级集合时，将数据分成训练集和验证集，通过训练集获取式中的各个参数，通过验证集验证超级集合的模拟水平。本文中参与建立超级集合的模型有 MODIS MOD17 A3、IBIS、Biome-BGC、LPJ，观测值数据是 FLUXNET La Thuile，具体对比结果在第。。。节介绍。

4.3.1.2 可视化对比方法

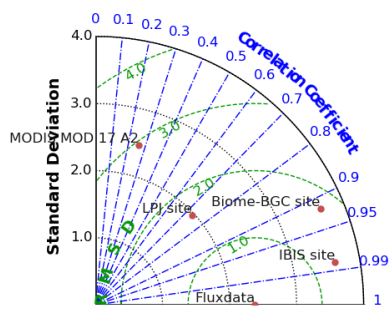
(1) 泰勒图

泰勒图常用于评价模型的精度，如图 4.11(a)所示，泰勒图中的散点表示模型，辐射线表示相关系数，横轴线表示标准差，虚线表示均方根误差^[65]，泰勒图能够在二维平面上同时呈现三个指标，在 IPCC、CMIP5、CMIP6 中被广泛应用于多模型模拟能力评估对比。

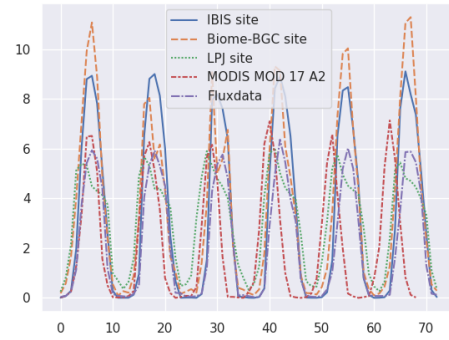
(2) 时间序列折线图

如图 4.11(b)所示，时间序列折线图反映的是模拟指标的年度变化趋势，并能够对比模拟值与观测值的趋势差异。

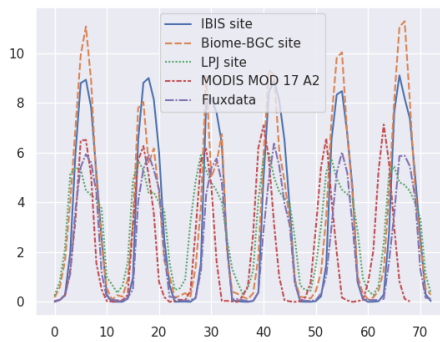
(3) 热力图



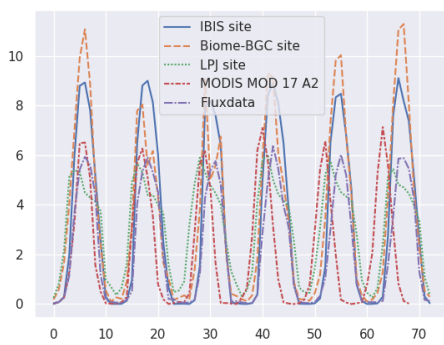
(a) 泰勒图



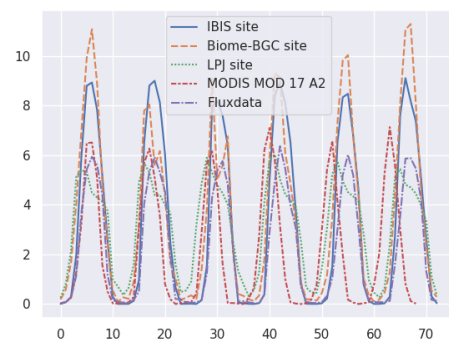
(b) 时间序列折线图



(c) 热力图



(d) 等值线图



(e) 偏差等值线图

图 4.11 可视化对比图表

如图 4.11(c)所示,横轴是观测站点,纵轴是模型,横着看能直观地发现同一个模型在不同站点的模拟能力差异,纵着看能直观地发现不同模型对同一个站点的模拟能力差异。

(4) 等值线图和偏差等值线图

如图 4.11(d)和 4.11(e)分别是等值线图和偏差等值线图,等值线图可以直观地表现 GPP 的空间分布格局,偏差等值线图则表示的是模型误差的空间分布。

4.3.2 开放式碳循环模型对比方法接入方法

4.4 本章小结

本章针对数据、模型和对比方法这三种对比资源在框架下的开放式服务化接入进行了详细的阐述。对于数据资源,从数据格式、尺度、编排三个角度分析了其特征,通过本文设计的结构化描述接口来描述 NetCDF、Shapefile 和 CSV 数据,实现了 OGC WMS、WFS、WCS、数据下载服务、数据处理服务的发布,为模型成功运行和结果对比提供了基础;对于模型资源,分析了其异构性特征,设计了面向人和机器理解的模型描述接口,并通过运行接口的设计实现了模型服务的封装、部署和发布;对于对比方法,首先详细介绍了本文用到的统计学对比方法和可视化对比方法,在采用模型封装接口将对比方法也封装为开放式服务资源。

第 5 章 原型系统与案例验证

5.1 原型系统构建

5.1.1 系统网站组织

5.1.2 系统开发运行环境

5.1.3 网络架构和系统功能模块设计

5.1.4 陆地生态系统碳循环对比资源库

5.1.5 开放式对比任务执行引擎

5.2 实验案例

5.2.1 模型资源和数据资源

5.2.2 对比方案

5.2.3 对比结果

第 6 章 结论与展望

6.1 研究结论

6.1.1 开放式对比系统

6.1.2 IBIS、Biome-BGC、LPJ 模型的验证和评价

6.1.3 IBIS、Biome-BGC、LPJ 模型对气候变化的响应

6.1.4 IBIS、Biome-BGC、LPJ 模型的不确定性分析

6.1.5 GPP、NPP 的时空分布格局

6.2 创新点

6.2.1 开放式、可共享、可重用的对比系统架构

6.3 不足与展望

6.3.1 不足

- (1) 模型参数的校准
- (2) 更多对比情景的实验
- (3) 更多对比模型的参与

6.3.2 展望

- (1)

参考文献

- [1] Lieth H, Whittaker R H. Primary productivity of the biosphere[M]. [S.l.: s.n.], 1975
- [2] Potter C S, Klooster S, Brooks V. Interannual variability in terrestrial net primary production: Exploration of trends and controls on regional to global scales[J]. *Ecosystems*, 1999, 2(1):36-48.
- [3] Potter C, Klooster S, Myneni R, et al. Continental-scale comparisons of terrestrial carbon sinks estimated from satellite data and ecosystem modeling 1982–1998[J]. *Global & Planetary Change*, 2003, 39(3):201-213.
- [4] Prince S D, Goward S N. Global primary production: A remote sensing approach[J]. *Journal of Biogeography*, 1995, 22(4/5):815-835.
- [5] Goetz S J, Prince S D, Small J, et al. Interannual variability of global terrestrial primary production- results of a model driven with satellite observations[J]. *Journal of Geophysical Research Atmospheres*, 2000, 105(D15):20077-20091.
- [6] Running S W, Coughlan J C. A general model of forest ecosystem processes for regional applications i. hydrologic balance, canopy gas exchange and primary production processes[J]. *Ecological modelling*, 1988, 42(2):125-154.
- [7] Running S W, Gower S T. Forest-bgc, a general model of forest ecosystem processes for regional applications. ii. dynamic carbon allocation and nitrogen budgets[J]. *Tree Physiology*, 1991, 9 (1_2):147.
- [8] Thornton P E, Running S. User' s guide for biome-bgc, version 4.1. 1[J]. University of Montana, Missoula, MT: Numerical Terradynamic Simulation Group, School of Forestry, 2000.
- [9] Gerten D, Schaphoff S, Haberlandt U, et al. Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model[J]. *Journal of Hydrology*, 2004, 286(1): 249-270.
- [10] Sitch S S B, Prentice I C, Arneth A, et al. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the lpj dynamic global vegetation model [review][J]. *Global Change Biology*, 2010, 9(2):161-185.
- [11] Oliver J E. Intergovernmental panel in climate change (ipcc)[J]. *Encyclopedia of Energy Natural Resource & Environmental Economics*, 2013, 26(2):48-56.
- [12] Meehl G A, Boer G J, Covey C, et al. The coupled model intercomparison project (cmip)[J]. *Bulletin of the American Meteorological Society*, 2000, 81(2):313-318.
- [13] C mip6-endorsed mips[Z/OL]. <https://www.wcrp-climate.org/modelling-wgcm-mip-catalogue/modelling-wgcm-cmip6-endorsed-mips>.
- [14] Taylor K E, Stouffer R J, Meehl G A. An overview of cmip5 and the experiment design[J]. *Bulletin of the American Meteorological Society*, 2012, 93(4):485-498.
- [15] 胡迪. 地理模型的服务化封装方法研究[J]. *测绘学报*, 2015, 44(11):1298-1298.
- [16] Yue S, Wen Y, Min C, et al. A data description model for reusing, sharing and integrating geo-analysis models[J]. *Environmental Earth Sciences*, 2015, 74(10):7081-7099.

- [17] 毛留喜, 孙艳玲, 延晓冬. 陆地生态系统碳循环模型研究概述[J]. 应用生态学报, 2006, 17(11):2189-2195.
- [18] 袁文平, 蔡文文, 刘丹, 等. 陆地生态系统植被生产力遥感模型研究进展[J]. 地球科学进展, 2014, 29(5):541-550.
- [19] 谢馨瑶, 李爱农, 靳华安. 大尺度森林碳循环过程模拟模型综述[J]. 生态学报, 2018, 38(1): 41-54.
- [20] 王绍刚, 何国金, 刘定生, 等. 森林碳循环模型方法研究进展[J]. 科技导报, 2008, 26(9):72-77.
- [21] 王萍. 森林碳循环模型概述[J]. 应用生态学报, 2009, 20(6):1505-1510.
- [22] 基于数据同化的哈佛森林地区水、碳通量模拟[J]. 应用生态学报, 2013, 24(10):2746-2754.
- [23] 朱文泉, 陈云浩, 徐丹, 等. 陆地植被净初级生产力计算模型研究进展[J]. 生态学杂志, 2005, 24(3):296-300.
- [24] 杨延征, 马元丹, 江洪, 等. 基于 IBIS 模型的 1960-2006 年中国陆地生态系统碳收支格局研究[J]. 生态学报, 2016, 36(13):3911-3922.
- [25] 刘曦, 国庆喜, 刘经纬. IBIS 模拟东北东部森林 NPP 主要影响因子的敏感性[J]. 生态学报, 2011, 31(7):1772-1782.
- [26] Niu B, Zhang X, He Y, et al. Satellite-based estimation of gross primary production in an alpine swamp meadow on the tibetan plateau: A multi-model comparison[J]. 资源与生态学报 (英文版), 2017, 8(1):57-65.
- [27] Kim J, Waliser D E, Mattmann C A, et al. Evaluation of the surface climatology over the conterminous united states in the north american regional climate change assessment program hindcast experiment using a regional climate model evaluation system[J]. Journal of Climate, 2013, 26(15):5698-5715.
- [28] Kim J, Waliser D E, Mattmann C A, et al. Evaluation of the cordex-africa multi-rcm hindcast: systematic model errors[J]. Climate Dynamics, 2014, 42(5-6):1189-1202.
- [29] Kim J, Guan B, Waliser D E, et al. Winter precipitation characteristics in western us related to atmospheric river landfalls: observations and model evaluations[J]. Climate Dynamics, 2017, 50(1):1-18.
- [30] 刘敏, 江志红. 13 个 IPCC AR4 模式对中国区域近 40a 气候模拟能力的评估[J]. 大气科学学报, 2009, 32(2):256-268.
- [31] Stoner A M K, Hayhoe K, Yang X, et al. An asynchronous regional regression model for statistical downscaling of daily climate variables[J]. International Journal of Climatology, 2013, 33(11): 2473-2494.
- [32] Eyring V, Bony S, Meehl G A, et al. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization[J]. Geoscientific Model Development, 2016, 9(5):1937-1958.
- [33] Gutowski W J. The coordinated regional downscaling experiment (cordex): A framework for mitigation and adaptation information (invited)[C]//Agu Fall Meeting. [S.l.: s.n.], 2009.
- [34] 赵宗慈, 罗勇, 黄建斌. CMIP6 的设计[J]. 气候变化研究进展, 2016(3):258-260.
- [35] 袁爽. 空间数据 Web 处理服务平台设计与实现[博士学位论文]. [出版地不详: 出版者不详], 2010.

- [36] 吴楠, 何洪林, 张黎, 等. 基于 OGC WPS 的碳循环模型服务平台的设计与实现[J]. 地球信息科学学报, 2012, 14(3):320-326.
- [37] 毛曦, 李琦, 张子民. 基于 Web Service 的地球系统碳循环建模框架研究[J]. 地理信息世界, 2012, 10(2):57-60.
- [38] 宋东泽, 罗万明, 阎保平, 等. 一个生态传感网的 Web 处理服务平台[J]. 科研信息化技术与应用, 2015, 6(2):000037-46.
- [39] Yue S S, Wen Y N, Lu G N, et al. Key technology research on open architecture for the sharing of heterogeneous geographic analysis models[J]. ISPRS - International Archives of the Photogrammetry, 2013, 40(4):133-138.
- [40] Lieth H. Modeling the primary productivity of the world[M]//Primary productivity of the biosphere. [S.l.]: Springer, 1975: 237-263
- [41] SEINO H, UCHIJIMA Z. Agroclimatic evaluation of net primary productivity of natural vegetation[J]. Journal of Agricultural Meteorology, 1985, 41(2):139-144.
- [42] Hai-yan D, et al. Change characteristics of phytoclimate productivity and climate in Hohhot in nearly 50 years[J]. Journal of Anhui Agricultural Sciences, 2012, 19.
- [43] 王萍. 基于 IBIS 模型的东北森林净第一性生产力模拟[J]. 生态学报, 2009, 29(6):3213-3220.
- [44] Foley J A, Prentice I C, Ramankutty N, et al. An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics[J]. Global biogeochemical cycles, 1996, 10(4):603-628.
- [45] Kucharik C J, Foley J A, Delire C, et al. Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure.[J]. Global Biogeochemical Cycles, 2000, 14(3):795-825.
- [46] Thornton P E. Regional ecosystem simulation: combining surface-and satellite-based observations to study linkages between terrestrial energy and mass budgets[J]. 1998.
- [47] Kimball J S, Thornton P E, White M A, et al. Simulating forest productivity and surface-atmosphere carbon exchange in the boreas study region[J]. Tree Physiology, 1997, 17(8-9): 589-599.
- [48] White M A, Thornton P E, Running S W, et al. Parameterization and sensitivity analysis of the biome-bgc terrestrial ecosystem model: net primary production controls[J]. Earth interactions, 2000, 4(3):1-85.
- [49] Thornton P E, Rosenbloom N A. Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model[J]. Ecological Modelling, 2005, 189 (1-2):25-48.
- [50] Pietsch S A, Hasenauer H, Thornton P E. Bgc-model parameters for tree species growing in central European forests[J]. Forest Ecology and Management, 2005, 211(3):264-295.
- [51] 王绍武, 赵宗慈. 未来 50 年中国气候变化趋势的初步研究[博士学位论文]. [出版地不详: 出版者不详], 1995.
- [52] 秦大河. 气候变化的事实与影响及对策[J]. 中国科学基金, 2003, 17(1):1-3.
- [53] Griggs D J, Noguer M. Climate change 2001: the scientific basis. contribution of working group I to the third assessment report of the intergovernmental panel on climate change[J]. Weather, 2002, 57(8):267-269.

- [54] Fowler M, Lewis J. Microservices a definition of this new architectural term[J]. URL: <http://martinfowler.com/articles/microservices.html>, 2014.
- [55] Ludäscher B, Altintas I, Berkley C, et al. Scientific workflow management and the kepler system [J]. Concurrency and Computation: Practice and Experience, 2006, 18(10):1039-1065.
- [56] Zhao Z, Belloum A, Bubak M. Special section on workflow systems and applications in e-science [J]. Future Generation Computer Systems, 2009, 25(5):525-527.
- [57] Inc O G C. Opengis web processing service (version 1.0.0)[J]. 2007-6-8.
- [58] Moore R V, Tindall C I. An overview of the open modelling interface and environment (the openmi)[J/OL]. Environmental Science & Policy, 2005, 8(3):279 - 286. <http://www.sciencedirect.com/science/article/pii/S1462901105000456>.
- [59] 乐松山. 面向地理模型共享与集成的数据适配方法研究[博士学位论文]. [出版地不详]: 南京师范大学, 2016.
- [60] Inc O G C. Opengis map server implementation specification (version 1.3.0)[J]. 2006-3-15.
- [61] Inc O G C. Opengis feature service 2.0 interface standard[J]. 2010-11-2.
- [62] Inc O G C. Ogc wcs 2.0 interface standard-core[J]. 2010-10-27.
- [63] Gordon W S. Climate change, hydrology, and ecological models: Intercomparison and validation [D]. [S.l.: s.n.], 2003.
- [64] Krishnamurti T, Kishtawal C, LaRow T E, et al. Improved weather and seasonal climate forecasts from multimodel superensemble[J]. Science, 1999, 285(5433):1548-1550.
- [65] Taylor K E. Summarizing multiple aspects of model performance in a single diagram[J]. Journal of Geophysical Research: Atmospheres, 2001, 106(D7):7183-7192.