

协同虚拟地理实验系统设计与初步试验^①

李 毅^② 龚建华^③ 周洁萍 李文航* 王伟星 汪东川 徐道峰

(中国科学院遥感应用研究所遥感科学国家重点实验室 北京 100101)

(* 北京大学信息科学技术学院 北京 100871)

摘 要 针对协同虚拟地理实验设计与实现中的问题,研究了网络技术和虚拟现实技术支持下的协同虚拟地理实验的理论原理与设计思路,提出了一种符合综合集成研讨厅体系以及空间认知和知识发现规律的协同虚拟地理实验框架和若干的实验关键技术。并以辽宁省浑河洪水演进模拟为案例搭建了协同虚拟地理实验环境,从实验可控度、实验尺度、协同性、成本和重用性角度对协同虚拟地理实验进行了评估,评估结果表明,协同虚拟地理实验能够弥补现场地理实验和物理模型实验的不足,为地理问题的决策提供有效的技术支持。

关键词 地理科学,实验地理学,虚拟现实,虚拟地理环境,协同虚拟地理实验

0 引 言

目前,我国的地理学发展仍面临着新的问题与挑战^[1-3]。任美镔院士与李吉均院士在 2002 年的地理学年会上,提出了振兴与发展地理学的问题。而要实现 21 世纪地理学的振兴,其中的一条重要途径是发展“实验地理学”。黄秉维与竺可桢生前均寄希望于借鉴其他学科的理论 and 先进技术,把地理学变成一门实验科学^[4]。

随着大规模科学计算、虚拟现实、人工智能、遥感科学、地理/地球信息科学、地学计算可视化与多维图解^[5]等学科的发展和“虚拟地理实验”、“虚拟地理实验室”概念的提出^[6,7],现实/虚拟地理实验将逐渐成为现代地理科学发展的重要推动力。钱学森曾经指出地理系统是一个“复杂巨系统”,并认为这种复杂性问题不能用“还原论”的方法进行求解,应当采用“从定性到定量的综合集成方法与研讨厅体系”^[8-10]。万庆也指出,地理学的问题,无论是学术研究、工程实践还是地理教育,在本质上都是一个群体协同工作的过程^[11]。因此,本文着重针对协同虚拟地理实验环境建设的基本要求,提出了协同虚拟地理实验的设计框架和关键技术,并通过辽宁省浑河洪水演进协同实验,对所提出的设计框架和关键

技术进行了验证,从实验可控度、实验尺度、协同性、成本和重用性等方面对虚拟地理实验与现场地理实验和物理模型实验进行了对比和评估。

1 协同虚拟地理实验环境设计

1.1 系统需求分析

协同虚拟地理实验环境应该具备以下需求:

(1) 以大型的多源、多尺度集成数据库和多领域地学知识库为基础,符合地学空间认知和知识发现规律的集成化计算与分析环境。地理科学实验的主要任务是建立地理数学模型,发挥科学的解释和预言职能,活跃思想、激励问题,开拓知识领域,因此系统应该具备知识发现和地学规律总结的机制。

(2) 具备一个贴近现实的地学实验环境,一个多维、多感知的虚拟地理环境,能够实施对过去地理现实的反演、对现在地理现实的放大以及对未来地理现实的预测,并使实验人员进入一个地学问题思考的状态。

(3) 能够为进入实验环境的研究人员提供接近自然的实时交流交互手段。同现实的交流类似,这种自然的交流交互手段应是多角度、多层次的,包括以化身为基础的视觉、听觉甚至感情等,从而实现在三维虚拟教学环境中接近于真实的面对面交流。

① 973 计划(2007CB714402)和国家科技重大专项(2009ZX10004-720,2008ZX10004-012)资助项目。

② 男,1981 年生,博士生;研究方向:虚拟地理环境,虚拟地理实验。

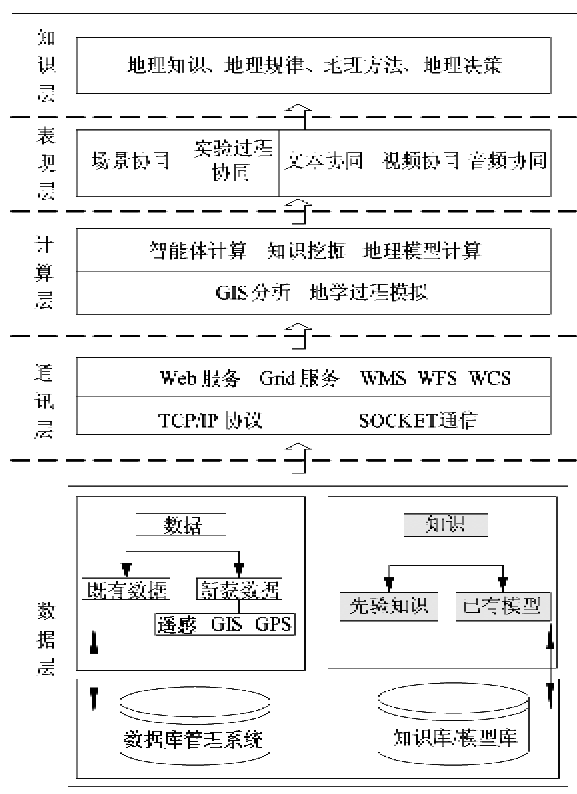
③ 通讯作者, E-mail: jhgong@irsa.ac.cn

(收稿日期:2008-12-29)

(4) 有一个协同实验控制的环境。不同研究领域、不同地域的研究人员能够针对同一个地学问题进行同步仿真实验,不同实验人员可以看到其他人的实验操作过程和结果,整个实验能够有条不紊地协同进行。

1.2 系统框架设计

基于以上考虑,本文设计出一种协同虚拟地理实验的系统框架。系统框架分为数据层、通讯层、计算层、表现层、知识层。数据层包括虚拟地理实验所需的各种数据以及诸多地学领域的知识库、模型库。通讯层采用 Web 服务,各种地理信息数据的服务以及网格(Grid)服务的方式建立协同体系的服务架构。计算层包括地理模型计算、地学过程模拟、GIS 分析、智能体模拟、知识挖掘等计算的手段及方法。表现层主要是采用协同技术、虚拟现实和多媒体技术展现一个贴近现实的地学实验环境和多实验者交流互动的平台。知识层是在三维人机交互的基础上加入地理知识、地理决策的表达方法,对实验中所获取的新知识进行表达。整个系统框架设计如图 1 所示。



WMS: Web 地图服务; WFS: Web 要素服务; WCS: Web 覆盖服务

图 1 协同虚拟地理实验系统框架

2 协同虚拟地理实验关键技术

2.1 实验数据获取与数据流技术

虚拟地理实验区别于 GIS 技术以及其他地学分析技术的本质在于虚实结合,即虚拟地理实验需要能与现实实验以及物理模型实验相结合,互相之间的数据流传输,以及实验结果之间的对比。从中不断对实验的手段、地学模型以及知识库进行改进。

因此虚拟地理实验需要对真实世界采用遥感技术以及大面积的无线网络传感器技术进行数据获取,以及对实验中的物理模型进行数据采集,最终经过压缩后以数据流的方式源源不断地传输到虚拟地理实验中来,以供实验与分析。

2.2 虚拟场景建模与可视化技术

虚拟地理实验需要针对真实世界的地理要素和模型进行虚化,并在虚拟地理环境中重现。采用虚拟场景建模与可视化技术建立逼真的虚拟实验场景,方便实验者进行仿真实验。

2.3 协同技术

虚拟地理实验中的协同技术既包括实验者之间自然的协同交流、协同实验控制技术,又包括计算协同技术。

实验者之间的协同交流指的是实验者能共处一个虚拟实验环境中,通过视频、音频、文本等方式就地学问题进行探讨。协同实验控制技术指通过工作流等技术手段,使实验者对实验进行同步协同控制。而计算协同技术指通过网络技术使实验者共享实验的计算资源和数据资源。

2.4 地学模型计算技术

地学模型计算技术是虚拟地理实验的核心技术,也是通过实验不断发展和完善的技术。目前既包括传统的 GIS 技术和空间分析技术,也包括新兴的人工智能技术。

目前人工智能领域中的元胞自动机、神经网络和智能体等技术发展十分迅猛。在对于大范围复杂地理问题的计算有很广泛的应用^[12,13],比如地形侵蚀、洪水模拟、城市交通与土地利用变化模拟等。

2.5 知识可视化与知识挖掘技术

知识可视化与知识挖掘技术能通过概念图、认知地图和语义网络等手段有效地从大量的地理数据和地理概念中提取新的地学知识。通过知识发现和地学规律的总结进一步完善地学模型计算技术。

3 原型系统应用及分析

3.1 案例介绍

浑河发源于抚顺市清源县长白山脉的滚马岭,流经抚顺、沈阳等重要城市,汇入大辽河,经营口入渤海,干流总长 415km,流域面积 11481km²。近百年来浑河流域发生较大洪水 10 次。其中 1995 年,暴

雨强度大、范围广、持续时间长,洪水量级超过百年。造成沈阳城区受灾面积 61km²,农作物受灾面积 25 万多亩,受灾人口 10 万人,直接经济损失达 26.61 亿元。随着沈阳浑南大开发建设,目前浑河成为沈阳城市的内河,与城市社会经济的结合更加紧密,已成为沈阳城市防洪的关键。浑河沈阳城市段河段分布图见图 2。



图 2 浑河沈阳城市段河段分布图

辽宁省水利厅为了研究浑河沈阳市段的洪水影响以及防洪工程规划,建造了一个与现实相似的洪水模拟物理模型,该物理模型主要针对模拟已建、拟建的堤防、桥梁、拦河坝、隧洞、水闸和丁坝等建筑物,观测在宣泄各设计频率洪水时沿程水位、水流流速、流态、淹没范围等水力要素,为确定合理的工程设计参数、河道工程治理方案以及近期、远景规划方案提供技术支撑。物理模型的规划施工过程严格按照实验的相似性准则,例如,建筑物模型高程误差控制 $\pm 0.3\text{mm}$;地形高程误差控制 $\pm 2\text{mm}$ 等。河道糙率根据天然河道的糙率情况按照糙率比尺进行模拟,并在完成加糙后的模型上施放相应的洪水,通过加糙情况调整,使模型实测水位与洪痕吻合,以确保物理模型的准确性。

由于物理模型建造成本昂贵,且利用率低,因此迫切需要高效准确的虚拟地理实验方法。

3.2 协同虚拟地理实验原型系统工作

3.2.1 实验数据收集

应用三维激光扫描仪扫描并建立了浑河城市段物理模型的三维数字地形模型,并收集了浑河城市段相关的地质与水文信息,以及百年一遇的洪水物理模拟试验数据。

3.2.2 系统网络模型

3.2.2.1 服务端

服务端选用 Java 作为开发语言。服务器是整个协同虚拟地理实验系统的信息中枢,其功能是将

某客户端激发的协同信息向其他客户端发送,能够对不同实验阶段的用户的操作权限进行限制,并维持所有客户端协同场景及内容的一致性。

服务器端部署了真实浑河地区的高程数据、影像数据,水文数据;以及物理模型的三维激光扫描地形,洪水物理模拟试验数据。另外还包括当地土地利用图、土壤侵蚀图以及大量相关领域的知识库和模型库。同时实现了消息、文字、数据、视频、音频等多种协同服务功能。每种服务都是一个多线程机制的 ServerSocket。能够在特定端口监听并响应客户端请求,启动新的服务线程,并在服务器端自行注册;服务线程启动后将按照自身的规则主动向客户端发送特定数据。

3.2.2.2 客户端

客户端是各专家用户参与虚拟地理实验的交互接口,采用 Java 和 C++ 结合开发的模式。以 Java 和 Java3D 作为三维交互界面与网络通信的系统架构,洪水模拟底层计算模块采用 C++ 语言实现。

三维场景以树状结构组织场景数据,同时响应各种鼠标、键盘事件,以允许操作者在三维场景中任意漫游,并对场景内容进行修改。客户端所有可能引发协同的键盘、鼠标、菜单事件如虚拟场景视点的改变、往虚拟场景中添加物体、删除物体等操作均被自动捕获,并编码成操作命令向服务器发送。

3.2.3 洪水演进模拟

圣维南方程是描述水道和其他具有自由表面的

浅水体中渐变不恒定水流运动规律的偏微分方程组。由反映质量守恒律的连续方程和反映动量守恒律的运动方程组成。

本文采用刘树坤^[14]推导的圣维南方程组的离散计算方式结合元胞自动机实现浑河百年一遇的洪水模拟。

3.3 协同虚拟地理实验分析

以上工作在基于分布式协同实验研讨室的浑河洪水模拟实验中得到了很好的应用。如图 3 所示,虚拟空间中的电子沙盘上表达浑河城市段的三维洪水模拟实验,墙壁的显示屏上显示物理模型的试验影像以及洪水演进的效果图,其中洪水的演进状态随时间变化,与物理模型的实验数据以及真实世界的历史数据实时比较。同时可供多用户参与者使用计算机网络相互联系,以化身的形式参与,并通过视频、音频、文字等多种途径进行实时交流与协同实验分析。

通过初步实验,浑河百年一遇的洪水河道水深分色和淹没范围如图 4 所示,虚拟实验中模拟结果与物理模型实验、洪水历史数据的比较如表 1 所示,其中,物理模型实验数据与历史数据大体一致。



图 3 协同浑河洪水演进实验

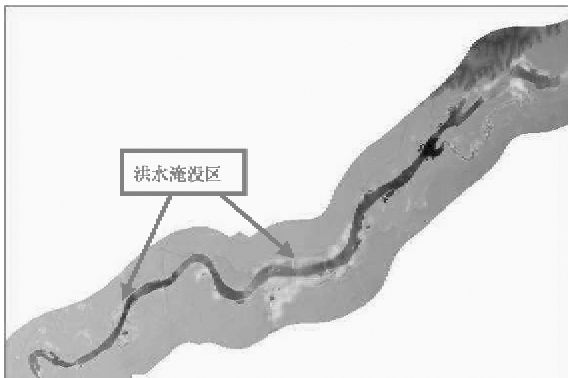


图 4 百年一遇的浑河洪水演进模拟结果

经过数据比较可以看出,虚拟实验中从浑河闸至浑南坝一段模拟的洪水水位偏低,而从浑南坝至旧东陵桥一段,模拟的洪水水位偏高。经过实验人员的视频语音讨论和文本的交流分析,初步认为该误差的一部分原因是河道糙率数据处理中的人为误差造成,另一方面原因是由基于圣维南方程组的元胞自动机计算中时空单元尺度选择不当造成,基于此分析大量后续工作正在开展中。

表 1 协同虚拟地理实验结果与物理模型实验数据(历史数据)的对比

断面号	累加距	实验结果水位(m)($P=1\%$)	
		物理模型实验 (历史数据)	虚拟实验
浑河闸	0	37.38	37.19
胜利桥下游	7700	39.33	39.15
胜利桥上游	7800	39.42	39.26
工农桥下游	9220	39.96	39.81
工农桥上游	9320	40.02	39.94
浑南坝下游	13755	40.55	40.56
浑南坝上游	13855	43.78	43.81
浑河桥下游	14770	43.80	43.84
浑河桥上游	14870	43.99	44.10
富民桥下游	17390	44.33	44.38
富民桥上游	17490	44.45	44.52
长青桥下游	20145	44.86	45.07
长青桥上游	20245	44.92	45.15
八三隧道	23715	45.36	45.40
王家湾桥	25905	47.18	47.22
新东陵桥下游	29795	47.35	47.37
旧东陵桥上游	29915	47.53	47.57
浑北坝下游	32525	48.34	48.33
浑北坝上游	32625	54.91	54.92
干河子坝上游	33665	55.15	55.18
高坎桥下游	38255	56.00	56.03
高坎桥上游	38355	56.06	56.09
下伯官坝下游	38914	56.18	56.22
下伯官坝上游	39014	59.77	59.78

从本文的原型系统以及应用案例中我们可以总结得到现场地理实验、物理模型实验与虚拟地理实验的差异和优缺点,见表 2。

4 结 论

本文介绍了协同虚拟地理实验的系统需求和体系框架,分析了开展实验所需的关键技术,并以辽宁省浑河洪水演进模拟实验为案例对协同虚拟地理实验进行了应用分析。

表 2 现场地理实验、物理模型实验与协同虚拟地理实验的对比

	实验环境	实验可控度	实验性质	实验尺度	人与人之间协同性	模型的成本	模型重用性
现场地理实验	现实野外地理环境	弱	直接实验	宏观尺度	较差	高 低	弱 强
物理模型实验	现实野外或室内环境	较强	间接实验	缩小尺度	强		
协同虚拟地理实验	虚拟地理环境	强	间接实验	无尺度	强		

经过初步的协同虚拟地理实验并结合现场以及物理模型实验工作可以看出,地理问题属于复杂性系统问题,对其进行精确的实验模拟需要以实验地理学的视角结合各地学领域专家协同完成。但是采用现场的地理实验往往难以将诸多专家及时地集中到现场,而且实验的可控性较差。并且建立物理模型进行实验需要较高的成本,模型的重复利用性差。

协同虚拟地理实验可以进行现场地理实验所无法进行的模拟实验,可以通过与历史数据和物理模型实验数据的比较和协同研讨分析不断改进实验技术与手段,并能与现实及物理模型实验互相补充和结合,有效地帮助地学规律总结以及地理问题决策。

参考文献

- [1] 任美镔. 地理学——大有发展前景的科学. <http://www.dlpd.com/dlqp/shownews.asp?newsid=413>; 中国地理学会, 2002
- [2] 陆大道. 中国地理学发展若干值得思考的问题. 地理学报, 2003, 58(1): 3-8
- [3] 史培军, 宋长青. 中国地理教育: 继承与创新. 地理学报, 2003, 58(1): 9-16
- [4] 唐登银. 实验地理学与地理工程学. 地理研究, 1997, 16(1): 1-10
- [5] 陈述彭(主编). 地学信息图谱的探索研究. 北京: 商务印书馆, 2001
- [6] 龚建华, 林琚. 虚拟地理环境——在线虚拟现实的地理学透视. 北京: 高等教育出版社, 2001
- [7] 龚建华, 鲁学军. 虚拟地理环境与虚拟地理实验. <http://www.vgelab.org>; 中国科学院遥感应用研究所虚拟地理环境实验室, 2003
- [8] 钱学森, 戴汝为. 论信息空间的大成智慧——思维科学、文学艺术与信息网络的交融. 上海: 上海交通大学出版社, 2007
- [9] 钱学森. 一个科学新领域——开放的复杂巨系统及其方法论. 城市发展研究, 2005, 12(5): 1-8
- [10] 钱学森. 谈地理科学的内容及研究方法 (在 1991 年 4 月 6 日中国地理学会“地理科学”讨论会上的发言). 地理学报, 1991, 46(3): 257-265
- [11] 万庆. 地理协同工作与群体空间决策支持系统研究: [博士学位论文]. 北京: 中国科学院研究生院, 2002
- [12] Ez-Zahraoui H, Benrihane Z, Benyoussef A. The effect of off-ramp on the one-dimensional cellular automaton traffic flow with open boundaries. *International Journal of Modern Physics B*, 2004, 18(16): 2347-2360
- [13] Brath A, Montanari A, Toth E. Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences*, 2002, 6(4): 627-639
- [14] 刘树坤, 李小佩, 李仕功等. 小清河分洪区洪水演进的数值模拟. 水科学进展, 1991, 2(3): 188-193

Design and preliminary test of a collaborative virtual geographic experiment system

Li Yi, Gong Jianhua, Zhou Jieping, Li Wenhong*, Wang Weixing, Wang Dongchuan, Xu Xiaofeng
(State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications,
Chinese Academy of Sciences, Beijing 100101)

(* School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)

Abstract

Aiming at the problems in design and realization of collaborative virtual geographic experiments, the paper researches the theory and design of the experiments under the support of network and virtual reality technologies, and presents a collaborative virtual geographic experiment framework conforming to the system of the Hall for Workshop of Metasynthetic Engineering and the laws of spatial recognition and knowledge discovery, and gives its corresponding several key techniques. Taking the flood simulation of the Hunhe River in Liaoning province as a case, the paper builds the environment for collaborative virtual geographic experiments, and evaluates the collaborative virtual geographic experiments from the perspectives of experimental controllability, experimental scale, synergy, cost and reusability. The evaluation draws a conclusion that virtual geographic experiments can make up the insufficiency of field experiments and physical model experiments, and can provide the effective technical support for the decision-making of geographic problems.

Key words: geographic science, experimental geography, virtual reality, virtual geographic environment, collaborative virtual geographic experiment

基于 PC 机群的多点源高斯大气污染扩散模型的并行计算研究^①

胡 亚^{②*} 林 琨^{**} 徐丙立^{**} 朱 军^{*} 胡明远^{**}

(* 西南交通大学土木工程学院测量工程系,成都 610031)

(** 香港中文大学太空与地球信息科学研究所,香港)

摘 要 针对基于单机的多点源高斯大气污染扩散模型计算效率较低的问题,设计了高斯模型并行算法,该算法是基于计算输出层设计的,它将计算输出层尽量平均地分配到各个计算节点上。针对并行算法,构建了分布式并行计算平台。该平台采用三层分布式体系结构,即用户层、控制层和计算资源层。实验结果表明,基于该算法的高斯并行计算大大提高了模型计算效率,分布系统结构合理,可以为准实时甚至实时的环境管理业务提供支持。

关键词 个人计算机机群,多点源,高斯污染扩散模型,并行计算

0 引 言

高斯模型是一种应用普遍的大气污染扩散公式,可以用来计算污染物扩散均匀后污染源下风向三维空间点的污染浓度值^[1]。当污染源个数增多,每个污染源下风向需要计算污染浓度值的空间点个数也很多的情况下,在单机上完成高斯模型污染物浓度计算需要花费很长时间^[2]。目前,为了提高大气污染扩散模型计算效率,大都采用并行计算方法,如 Martín 利用硅图公司(SGI)的图形工作站作为并行计算平台^[3],Mourino 利用计算机网格作为并行计算平台^[4],姚丽萍利用“自强 3000”超级计算机作为并行计算平台^[2],王茜利用高性能计算机系统“曙光 4000L”作为并行计算平台^[5]。这些已有的工作以“超级计算机”或“高性能计算机”作为并行计算平台取得了较高的计算效率,但是采用“超级计算机”或“高性能计算机”作为计算平台,其计算能力有一定的限制,并且价格昂贵,需要较高的一次性投入,许多单位或机构都不具备这种计算平台。随着计算机技术的普及以及计算机性价比的提高,在许多工作单位中都出现了大量的高性能个人计算机(PC)和由这些 PC 机组成的局域网,而这些 PC 机在大部分时间里其计算资源都处于空闲状态。针对前述的“超级计算机”计算资源不足而 PC 机计算资源空闲这一现状,本文尝试把多点源高斯大气污染扩散模

型计算移植到由普通 PC 组成的计算机机群上来。

根据计算运行平台不同,并行计算主要分成两类:基于大型并行机即超级计算机的运算;基于 PC 机群的运算。随着计算机技术的发展,基于 PC 机群的并行计算正在逐步取代一部分大型并行机的功能并在高性能计算领域取得领先地位^[6]。本研究的主要工作是以 PC 机群作为计算平台,采用并行方法以提高大气污染扩散模型的计算效率,在分析多点源高斯污染迭加模型的基础上,设计了基于计算输出层的并行算法,针对 PC 机群对并行分配和计算进行了框架设计和实现,实验表明该算法和系统构建基本合理,能够在一定程度上减少模型计算时间,提高模型的运行效率。

1 多点源高斯迭加模型

区域性大气环评高斯迭加模式是由烟气抬升公式和各类源的高斯扩散公式组成,多源模式一般只考虑点源与点源、点源与面源的迭加。多点源高斯迭加模型如下^[7]:

$$C(X, Y, Z) = \sum_{i=1}^n C_i(x, y, z) \quad (1)$$

其中:

$$C_i(x, y, z) = \frac{Q}{2\pi\sigma_y\sigma_z u} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \times \left\{ \exp\left[-\frac{(z - He)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z + He)^2}{2\sigma_z^2}\right] \right\} \quad (2)$$

① 863 计划(2009AA12Z207, 2006AA12Z207)和香港特区研究资助局基金(447807)资助项目。

② 男,1980 年生,硕士,讲师;研究方向:GIS 与虚拟地理环境;联系人, E-mail: huya@home.swjtu.edu.cn (收稿日期:2009-01-16)

在式(1)中, $G(X, Y, Z)$ 为 n 个污染源对空间点 (X, Y, Z) 的污染浓度值迭加, (X, Y, Z) 为空间点 p 在固定坐标系(固定坐标系中, X 轴指向东, Y 轴指向北)中的坐标表示, n 为污染源的个数, $C_i(x, y, z)$ 为第 i 个污染源在空间点 $p(X, Y, Z)$ 处的污染浓度值。

在式(2)中, $C_i(x, y, z)$ 为第 i 个污染源对其下风向空间点 (x, y, z) 处的空气污染浓度(mg/m^3)。这里的 (x, y, z) 为空间点 $p(X, Y, Z)$ 在风向坐标系(风向坐标系的坐标原点为第 i 个污染源所在的位置, x 轴方向为风向方向, y 轴与 x 轴形成右手坐标系)里面的坐标表示。 x 为下风向距离(m), y 为横截风向距离(m); z 为距地面高度(m); Q 为污染物源强, 即释放率(mg/s); u 为排气口处的平均风速(m/s); σ_y, σ_z 分别为水平方向和垂直方向扩散参数(m), 它们是下风向距离 x 及大气稳定度的函数; H_e 为污染源有效排放高度。

2 基于 PC 机群的并行计算

用于并行计算的机群系统主要有两类: 共享存储的机群系统(如超级计算机机群)和分布式存储的机群系统^[8]。PC 机群就是一种基于分布式存储器的多机系统, 它由多台同构或异构的普通计算机组成, 并通过网络进行连接。目前, 在我国, 很少有单位拥有一台或多台超级计算机, 而拥有多台 PC 机的单位已经十分普遍, 本文正是基于这种现状而展开研究工作的。

本节首先设计了并行计算的体系结构及流程, 然后详细介绍了本文所采用的并行算法, 最后讨论了并行计算中的网络通信等问题。

2.1 并行计算总体框架

基于 PC 机群的并行计算以网络为通信媒介, 采用三层体系结构, 如图 1 所示。在图 1 中, 从右至左分别为并行计算请求端、并行计算控制端和计算节点序列端。并行计算请求端、计算节点序列端都通过局域网与并行计算控制端进行连接。

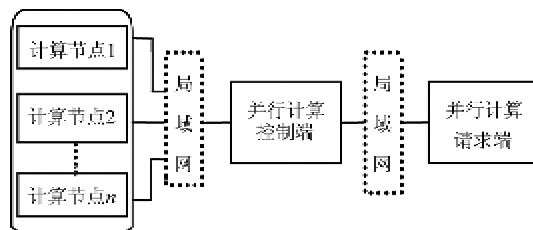


图1 基于 PC 机群的并行计算的三层体系结构

计算节点是由局域网内的 PC 计算机组成的计算机群, 它主要负责对分配的任务进行计算。并行计算控制端装配有并行算法, 它负责任务的划分与分配以及计算结果的汇总。并行计算请求端也是系统的客户端, 主要生成高斯模型的初始参数。并行计算的步骤为:

(1) 在并行计算请求端发起计算请求之前, 参与并行计算的计算节点应先与并行计算控制端建立连接以形成计算节点序列;

(2) 并行计算请求端向并行计算控制端发起计算请求, 当请求获得通过之后, 并行计算请求端向控制端发送与计算相关的参数和数据文件;

(3) 当控制端接收到从并行计算请求端发送过来的参数与数据文件之后, 根据并行算法计算各计算节点应该计算的污染层, 并向各计算节点发送相应的参数和数据文件;

(4) 当计算节点接收到与计算相关的参数和数据文件之后, 利用前述公式(1)、(2)计算分配到的空间点污染浓度值;

(5) 当计算完所有分配到的空间点污染浓度值后, 把各空间点的位置数据以及污染浓度数据写入文件形成结果文件, 然后把结果文件传送给控制端;

(6) 控制端接收各计算节点发送过来的结果文件, 并把结果文件发送给并行计算请求端;

(7) 并行计算请求端接收控制端发送过来的结果文件。当接收完所有的结果文件后, 整个并行计算结束。

以上的过程可以用图 2 表示。

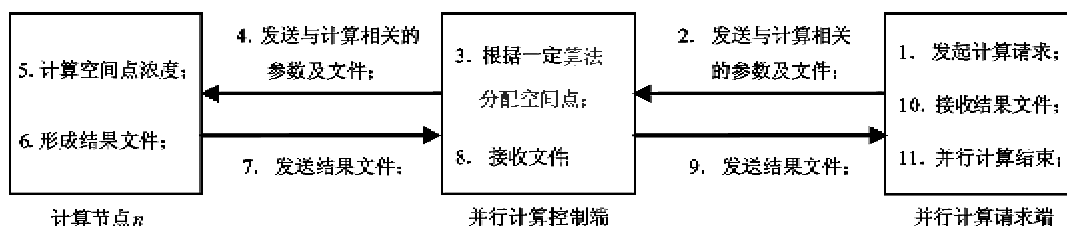


图2 基于 PC 机群的并行计算流程

2.2 基于污染层的并行算法

基于多点源高斯迭加模型的并行计算可以采用如下三种方法:

(1) 针对污染源并行: 认为一个计算点的浓度是多个污染源的叠加贡献, 如果参与计算的污染源很多, 可以把污染源分配到各个计算节点上去, 让每个节点只计算少量污染源对计算点的浓度分量, 最后把各个浓度分量叠加得到所有污染源对计算点的浓度。

(2) 针对计算点并行: 如果要求计算浓度的计算点数量很多, 可以把计算点分配到各个计算节点上去, 让各个节点只计算所有污染源对少量计算点的浓度。

(3) 针对计算点及污染源的并行: 如果参与计算的污染源和计算点都很多, 并且能够参与并行计算的计算节点也足够多时, 可以结合(1)、(2)两种并行方法以进一步提高模型计算效率。

本文由于要求计算浓度的计算点数量很多, 并且能够参与并行计算的计算节点数量有一定的限制, 所以采用针对计算点的并行方法。计算点的分配可以根据需要按任意方式进行, 这里采用按层的方式进行。具体的并行算法如下所述。

设污染区域有 L 层, 每层采用相同的划分方式, 即在 Y 轴方向上有 R 个格点, 在 X 轴方向上有 C 个格点。污染区域的设置如图 3 所示。

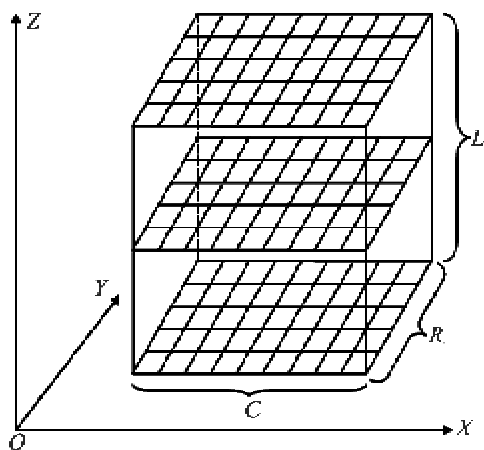


图 3 污染区域设置

假设参与并行计算的计算节点数为 $ComputerNodeNum$ 。为了让各个计算节点的计算量均衡, 针对污染层的并行算法是让每个计算节点需要计算的污染层数尽可能相等。具体算法如下: 首先, 可以计算每个计算节点至少可以分配到的污染层数为 $AverageComputerLayerNum = (int)(L /$

$ComputerNodeNum)$ 。如果计算节点数不能整除 L , 则有 $HeavyComputerNodeNum = L \bmod ComputerNodeNum$, 规定编号小的计算节点可以分得更多的层, 这些计算节点将分得 $AverageComputerLayerNum + 1$ 层。这样就可以得到编号为 $flag$ (计算节点的编号从 0 开始) 的计算节点分得的层数:

$$ComputerLayerNum = \begin{cases} AverageComputerLayerNum + 1 & (flag < HeavyComputerNodeNum) \\ AverageComputerLayerNum & (flag \geq HeavyComputerNodeNum) \end{cases}$$

例如, 假设参与并行计算的计算节点数 $ComputerNodeNum$ 为 3, 需要计算的污染层数 L 为 8, 则每个计算节点至少可以分配到的污染层数 $AverageComputerLayerNum = (int)(8/3) = 2$ 。由于计算节点数 3 不能整除污染层数 8, 此时剩余的污染层数 $HeavyComputerNodeNum = 8 \bmod 3 = 2$ 。根据上述并行算法, 第一台计算节点(编号为 0)、第二台计算节点(编号为 1), 由于编号均小于剩余的污染层数 2, 它们分配到的污染层数为 $AverageComputerLayerNum + 1 = 2 + 1 = 3$; 第三台计算节点(编号为 2), 由于编号与剩余的污染层数 2 相等, 所以它分配到的污染层数为 $AverageComputerLayerNum = 2$ 。

通过上述的基于污染层的并行算法, 各计算节点计算的污染层数相差最大的为 1 层, 可以最大限度地平衡计算节点间的计算负荷, 有利于整个并行计算尽快结束。当计算机数量增加时, 计算时间将减少。当计算机数量达到和层数相同时, 计算时间最小。但是当计算机数量超过计算层数时, 计算时间将不再减少, 增加的计算机将造成浪费。

2.3 网络通信

并行计算中的网络通信通常采用如下三种方式:

(1) 利用并行计算平台, 如消息传递接口 (message passing interface, MPI)、并行虚拟机 (parallel virtual machine, PVM), 实现计算机之间的网络通信^[9]。这种方法适合于大型、复杂的并行计算, 多用于 Linux 操作系统。

(2) 借助于专门的软件实现计算机之间的网络通信。如在计算流体力学领域里广泛使用的 FLUENT 软件需借助于专业远程服务软件 RSHD 实现并行计算^[10]。这种方法适用于某些特定领域。

(3) 利用网络编程实现计算机之间的通信^[11]。这种方法适用于并行流程比较简单的并行计算。

本文的并行计算其并行流程比较简单,所以计算机之间的通信采用网络编程实现。本文采用 Windows 网络编程标准 Windows Sockets,借助于 VC. Net 2005 平台的强大网络通信开发环境,编写程序实现计算机之间的通信。

由于各计算节点只计算分配给它的污染层的污染浓度值,计算节点之间没有联系,所以网络通信只在并行计算请求端与并行计算控制端、并行计算控制端与计算节点之间进行。在计算节点开始计算之前,网络通信的方向是:并行计算请求端→控制端→计算节点,传输与计算相关的参数与文件,即高斯模型计算参数、选择的污染源数据等。在计算节点完成计算之后,网络通信方向是:计算节点→控制端→并行计算请求端,传输结果文件与相关的控制命令等。

3 并行测试与分析

根据上述方案,采用 C++ 语言在 VC. Net 2005 平台上编写并行计算的 3 个程序,即并行计算请求端程序、并行计算控制端程序、计算节点端程序。并行计算请求端程序运行界面如图 4 所示,控制端与计算节点端程序运行界面省略。在图 4 中,左边为高斯模型部分参数设置区域,右边的列表为参与并行计算的计算节点序列。

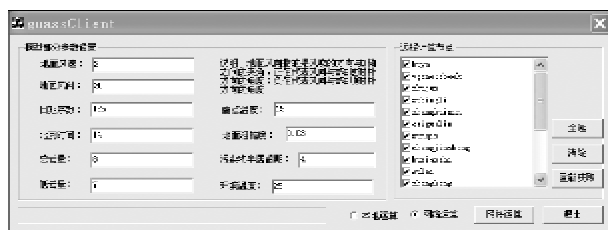


图 4 并行计算请求端程序运行界面

并行计算中,计算节点端的硬件环境为 HP dx7200 系列,其处理器为英特尔双核 3.20GHz,内存为 2GB,操作系统为 Windows XP;并行计算采用的局域网网速为 100.0Mbps,在该局域网内传输文件的速度大约为 10Mbps。实验区域为珠江三角洲,区域内的点污染源分布(共 3179 个)。本次并行计算选择的空点分布如图 5 所示。其中,空点平面位置分布在红线框范围内,其个数的表现形式为 $220 \times 247 \times 45$ (45 是 Z 方向上空点的层数);同一层之间,空点与空点的 X 方向间距、Y 方向间距为 1000m。基于上述硬件和实验数据,图 6 是并行

计算用时与计算节点关系测试结果。

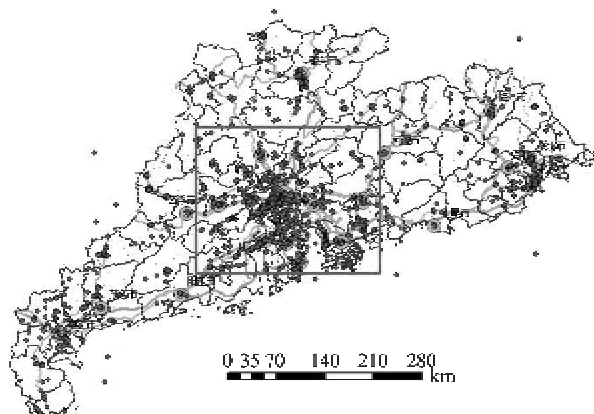


图 5 实验区域污染源分布及空间点分布区域

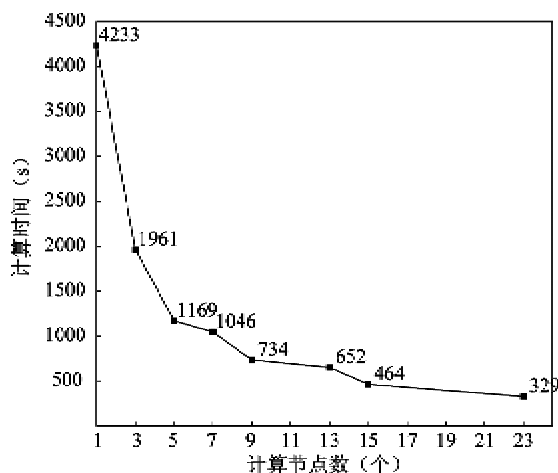


图 6 并行计算用时与计算节点关系测试结果

图 6 中的计算时间是由图 2 中整个并行计算流程的用时所组成,包括模型计算时间和数据在网络上的传输时间。从图 6 中的测试结果可以看出,当使用 15 台并行节点时,计算能够在 10min 内完成,大大缩短了模型计算所需要的时间,这说明基于 PC 机群的多点源高斯大气污染扩散模型的并行计算可以提高模型的计算效率。测试结果同时表明,随着计算节点的增加,虽然在并行计算控制端与计算节点端的通信量会增加(每增加一个计算节点,只会多增加一次网络通信,即从并行计算控制端向并行计算节点端传输的与计算相关的初始参数信息,而在其它地方没有再增加网络通信。对于本文所采用的实验数据,每增加一个计算节点,需要增加约 220kB 网络通信量,这对于每秒能传输大约 10MB 文件大小的网络速度,增加的网络通信时间大约为 0.0215s),但增加的通信主要在高速的局域网内完成,通信增加的时间远小于并行计算减少的时间,甚至可以忽略不计,所以并行计算完成的时间会随着计算节点的增加而进一步减少。针对本次计算选择

的空间点数量,如果参与计算的计算节点足够多(如超过 45 个计算节点),并行计算完成的最短时间应约等于在本地计算完成一层空间点(即 $220 \times 247 \times 1$)的计算时间 115s 与并行计算网络通信时间 30s (每增加一个计算节点,针对本文所采用的实验数据,增加的网络通信时间大约是 0.0215s。由于增加的通信时间很少,所以这里认为并行计算网络通信时间与计算节点的数量没有关系,大约为 30s 左右)之和,即 145s。

为了验证计算结果的正确性,对并行计算结果的第 5 层(空间点距地面高 500m)进行了可视化,见图 7。可视化采用 GIS 二次开发平台 ArcEngine 实现。从图中可以看出,污染浓度分布符合高斯形态,浓度渐变(颜色黑的地方代表污染浓度高)方向与风向一致(模型计算时,风向与 X 轴逆时针方向形成 30 度)。

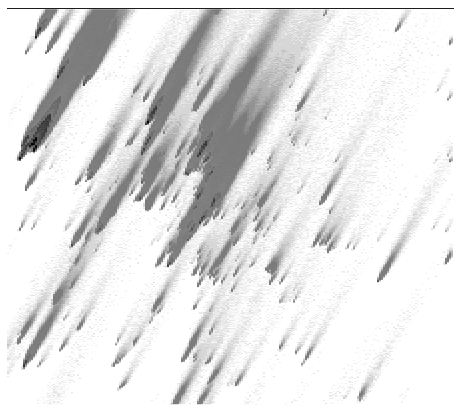


图 7 高斯大气污染扩散模型结果可视化

4 结论

(1)本文以普通 PC 机群作为计算平台,实现了多污染源高斯大气污染扩散模型的并行计算。实验结

果表明,本文所采用的技术路线大大提高了模型的计算效率,缩短了模型计算完成的时间,可以为准实时甚至实时的环境管理业务提供支持。(2)本文使用的大气污染扩散模型是较为简单的多污染源高斯迭加模型,但就其实现过程而言,可以推广到更为复杂的大气污染计算模式,进而可以有效解决环境管理业务对大气污染扩散模拟提出的效率需求。(3)随着时间的推移,PC 机的性价比会进一步的提高,PC 机的计算能力会进一步增强,许多单位都会拥有更多更高性能的 PC 机。因此,本文所采用的技术体系会更适合于没有超级计算机或超级计算机资源紧张而又需要快速完成大量计算的单位或组织。当然,基于 PC 机群的并行计算需要编写较多的程序代码,尤其是需要对网络编程比较熟悉。

参考文献

- [1] 魏振钢,郭遵强,张琳等.基于高斯模式的大气污染模型的应用.中国海洋大学学报,2008,38(2):327-330
- [2] 姚丽萍,王远飞.基于 MPI 的大气污染扩散模型的并行计算研究.计算机工程,2005,32(22):54-57
- [3] Martín M J, Parada M, Doallo R. High performance air pollution simulation using OpenMP. The Journal of Supercomputing, 2004, 28(3): 311-321
- [4] Mourino J C, Martín M J, González P, et al. High performance air quality simulation in the European crossagic project. Computing and Informatics, 2006, 25(4): 1001-1019
- [5] 王茜,张法,邱振戈.城市小区气象污染扩散并行算法优化.高性能计算发展与应用,2008,24(3):38-42
- [6] 谢晓东,王勇,王忠群.基于网络的并行计算系统模型.计算机技术与发展,2007,17(8):198-204
- [7] 吴兆清,邓国胜,许国强等.高斯模式在株洲市城区 SO_2 分布规律研究中的应用.湖南有色金属,2004,20(2):33-35
- [8] 花峰,傅游.基于 PC 机群并行环境的构建方法.山东科技大学学报(自然科学版),2002,21(3):28-30
- [9] 刘伟军,谭功.基于消息传递的网络并行计算环境.四川理工学院学报(自然科学版),2004,17(3):67-71
- [10] 罗挺,许雅娟.山谷污染物扩散过程 CFD 并行计算方法的研究.四川环境,2005,24(4):77-80
- [11] 刘勇,朱国强.基于 TCP/IP 的网络并行计算的实现.计算机工程,2000,26(8):187-189

A study on PC cluster based parallel computation of Gauss plume model for multi point sources

Hu Ya* **, Lin Hui**, Xu Bingli**, Zhu Jun*, Hu Mingyuan**

(* Surveying Engineering Department, School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031)

(** Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong)

Abstract

To solve the problem that the Gauss plume model (GPM) widely used for air pollution modeling is of low efficiency when applied to large number of point sources, the study presented in this paper tried to improve GPM computation efficiency by designing a parallel algorithm and developing a distributed system based on personal computer (PC) cluster. The parallel algorithm was designed according to the levels of output, which allocates multiple levels to multiple computers evenly as far as possible. The architecture of the distributed system based on PC cluster was designed to have the three tiers of computation sources, task allocation and user interface for command generation. The test results show that the algorithm is reasonable and the computation time can be significantly decreased to one-tenth of that of one computer. The tests also prove that the PC cluster based distributed system is reliable and rational.

Key words: PC cluster, multi-point sources, Gauss plume model, parallel computation