

基于大数据的水环境风险业务化评估与 预警研究

王永桂, 夏晶晶, 张万顺*, 张潇

(武汉大学资源与环境科学学院, 武汉 430079)

摘 要 随着生态环境大数据的不断建设, 国家级、流域级和行政区级别的大数据中心相继产生。快速、高效地分析和利用实时产生、类型多样的环境大数据, 提取环境大数据在风险评估与预警的价值, 具有重要的意义, 也充满挑战。本文基于环境大数据的数据特点和不同层次的组织管理特征以及环境风险评估预警的业务化需求, 构建基于大数据的流域环境风险评估与预警技术及其业务化系统的体系、分析环境风险智能识别模式、研究环境风险高效模拟预测和评估的方法, 并提出一套满足各级管理部门需求, 高效利用环境大数据的多中心业务化系统。

关键词 大数据; 风险评估; 预警预测; 业务化

中图分类号: X824; X52

文章编号: 1674-6252 (2017) 02-0043-08

文献标识码: A

DOI: 10.16868/j.cnki.1674-6252.2017.02.043

Research of Water Environment Risk Assessment and Early Warning Based on Big Data

WANG Yonggui, XIA Jingjing, ZHANG Wanshun*, ZHANG Xiao

(School of Resources and Environmental Science, Wuhan University, Wuhan 430079)

Abstract: With the continuous construction of ecological environment big data, national, basin and administrative level of big data centers appeared one after another. It is of great significance and challenge to quickly and efficiently analyze and make use of the real-time generation, and various types of environmental big data, extract its value in environmental risk assessment and forecasting. Based on the characteristics of ecological environment big data, its organization and management at different levels and the operational requirements of water environmental risk assessment and early warning, this paper built the big data based operational system for basin water environmental risk assessment and early warning. Intelligent identification model for environmental risk and the method of efficient simulation and prediction of environmental risk were researched, then a set of multi-center operational system, which can meet the needs of multilevel business management and efficient use of environment data, was put forward.

Keywords: big data; risk assessment; early warning and forecasting; operational

引言

随着大数据上升为国家战略, 我国环保、军事、医疗等各个领域大力开展大数据的建设, 环保部发布《生态环境大数据建设总体方案》, 从总体上明确了环境大数据的建设、管理、运维和应用的方向^[1]。大数据有别于传统数据, 是具有海量规模 (Volume)、实时产生

(Velocity)、类型多样 (Variety) 的巨价值 (Value) 数据云。2015年国务院发布的《生态环境监测网络建设方案》, 确定了2020年我国全面生态环境监测网络, 实现各级各类监测数据系统互联共享和监测预报预警的目标^[2]。这预示着我国环境大数据的规模将不断增大, 由此带来的环境大数据分析、计算和应用的能力建设与科学研究, 将成为环境领域的重点和热点。

资助项目: 国家水体污染控制与治理科技重大专项——三峡库区及上游流域水环境风险评估与预警技术与示范课题 (2013ZX07503-001); 中国博士后科学基金 (2016M602364)。

作者简介: 王永桂 (1987—), 男, 博士, 主要从事流域环境模拟与仿真、大数据、高性能计算技术研究, E-mail:

wangyonggui@whu.edu.cn。

* 责任作者: 张万顺 (1965—), 男, 教授, 博士, 博士生导师, 主要从事河流动力学、环境管理学、生态环境水力学、环境数学模拟仿真、流域水环境演化等领域的研究, E-mail: wszhang@whu.edu.cn。

我国环境管理理念正从总量管理向风险管理转变，环境风险评估与预警是风险管理的关键技术和基础^[9]。国内外针对环境风险开展了大量的研究，涉及污染源识别、风险评价、风险模拟预测以及环境风险管理平台等各个方面，产生了一系列的研究成果。在技术上，形成了针对突发事故风险、累积性风险源识别技术、水动力水质模拟技术、水生态模拟技术和风险处置专家决策支持等技术成果^[6,7]；在系统构建上，形成了三峡库区、无资料地区和太湖流域水环境风险预警系统等成果^[8,9]。这些成果为我国环境风险评估与预警能力的提升提供了支撑，但仍然存在诸多问题。这些问题表现在，污染源识别方法依然依赖传统的人工调查，风险评估方法对数据利用率低，风险模拟预测模拟速度和模拟精度达不到业务需求，环境风险管理平台无法实现业务化，难以适应和满足新时期生态环境保护工作需求。

环境大数据构建、分析和计算技术的发展，为解决流域水环境风险评估与预警技术和应用的问题，提供了新的思路 and 工具。本文基于环境大数据的采集、分析方法和环境风险评估的业务需求，研究环境风险评估模型体系的基本结构、环境模型的高性能集群计算方法和多中心业务集成的环境风险管理预警平台，提出了适应于现代环境风险评估与预警的多中心的软件系统架构。

1 我国环境大数据特点和发展趋势

环境大数据拥有大数据的 4V 特征，未来我国环境大数据将围绕环境数据搜集—处理—公开—技术支持一体化方向不断发展。随着我国数据共享程度的不断加大，消除数据孤岛的政策不断发布，气象、水利、环保、测绘等部门的环境数据共联共享体系的构建，将为水环境数据的来源提供更多、更便捷的途径。以环保部级各个流域管理机构和各个行政单位为中心的数据中心将相继形成，多层次、多目标的环境大数据中心的建设对环境数据的分析、计算提出了更高的需求。为适应环境大数据的发展趋势，如何让沉睡的数据最大程度地发挥价值，其核心在于数据分析。基于云计算、智能计算、智能数据挖掘等理论提出的数据存储、计算、处理、展现的新一代数据分析技术，是环境大数据的必然选择。

2 环境大数据风险评估与预警的特征和需求

传统的环境风险模拟评估，基于有限的数 据，开展环境风险因子评价、环境风险等级的划分；以模糊计算、回归分析、相关分析和线性计算等方法确定环境风险的主要风险源、风险等级和风险范围^[10]。而在风险预警预测上，通过选取流域典型污染物，采样测定污染物分布特征，并应用数学模型计算开展预测和预警是当前的主要研究方式^[11]。但由于计算方法与数据的局限性，环境风险的预测预警限于局部区域、针对特定物质，且大量的工作投入到模型的不确定性和计算效率上，在非特定

区域发生环境风险时，已有的方法和技术显得捉襟见肘、束手无策。环境大数据的发展，为风险评估与预警的研究带来了新的机遇，在环境风险识别、风险评估与预测以及风险管理等方面带来了新的解决方案和思路，也提出了更大的挑战。

（1）环境风险识别是环境风险评估预警的前提，能为环境风险评估提供风险的来源、风险发生的时间、风险可能的程度和风险的责任单位、管理部门等众多信息。传统的环境风险识别，主要依赖于人工水质监测和不定期的排查。由于人工监测周期长、效率低、范围小，传统的环境风险识别时效性差、识别范围有限，且极为被动。如老大难的偷排问题，在没有民众举报或发生重大水质污染事件时，管理部门往往无法确切地获知偷排事件。即使知晓了偷排事件，对偷排责任主体的排查也极其困难，甚至无能为力。环境大数据的发展，得益于环境数据产生手段的不断进步，水文水质自动监测站、遥感监控、视频监控等环境数据的监测和环境监控方法，将产生大量实时的水文水质数据、遥感影像数据和视频监控数据，这些数据的产生频率高、产生数量多和覆盖范围广。基于实时、全面的环境大数据，应用数据挖掘方法、数据关联分析法、人工智能等方法，能实时、准确地识别环境风险，并根据环境风险中风险物质的种类等信息，迅速判断出环境风险的责任主体，将环境风险的人工识别转向机器智能识别，将固定时间的调查统计转为实时的主动发现。

（2）环境风险评估与预测是在环境风险识别的基础上，评估预测环境风险的发展趋势、判断环境风险的影响范围的过程。基于回归分析、模糊预测等统计学方法和基于水动力水质模型的机制模型方法是进行环境风险评估与预测的主流方法；统计学方法无法精确地得到风险在空间上的发展趋势和影响范围；机制模型依赖于丰富的数据支撑^[12]。传统的数据管理方法，无法快速地获取用于环境风险评估的数据，效率较低，且数据在时间上和空间上的局限性较大。环境大数据，更新了环境数据的存储、管理、索引和共享方式，从整体、全局的角度对数据进行收集和管理，一方面能充分利用区域已有的数据，迅速查找到用于环境风险评估的数据；另一方面，基于环境大数据的网络数据分析、检索和挖掘功能，在数据缺乏的地区，可通过数据类比、扩充、延长等方式，得到环境风险评估预测的数据。这就是说，环境大数据时代的环境风险评估与预测，不仅能分析风险发生的局部区域的特点，还能高效、快速地计算并预测全流域环境风险的时空分布状况和变化规律，追踪风险的来源，并能给出不同外界条件下环境风险的发展趋势。同时，对数据丰富地区，能开展精确的风险预报预测，对数据缺乏甚至无数据地区，基于大数据的分析评估，也能通过大数据的计算手段，开展趋势性的风险预报预测，突破环境风险评估与预警对特定物质、特定区

域的限制。

（3）环境风险管理是集环境风险识别、评估、预测和处置于一体的系统过程。随着信息化的发展，环境风险管理信息系统成为辅助决策的核心工具。在环境大数据的背景下，环境风险管理信息系统应该具备数据管理模块、数据分析模块和辅助决策模块。其数据管理模块既能全面对接环境数据监测体系，自动接收监测站数据，对数据进行预分析与处理和入库，又能自动搜集互联网相关的数据，充分利用网络共享数据，进行数据预判和处理与利用；其数据分析模块，应该能自动开展环境模拟预测、风险评估和警情通报；而其辅助决策模块，则能针对所评估的风险，智能地提供风险决策处置方案。

3 基于大数据的环境风险评估与预警体系

3.1 总体架构设计

大数据最根本的特点是挖掘隐藏在其大体量后的大价值。环境风险模拟评估与预警所需求的价值是在国家中心、流域中心、行政中心不同层次的环境大数据基础上，快速、准确地识别风险特点和风险来源，揭示风险时空变化规律，确定风险危害程度以及提出风险的最优处置方案。通过基于大数据采集、挖掘、分析和智能优化技术体系与环境大数据，形成业务化的软件平台系统，以提高国家中心、流域中心和行政中心的环境风险管理能力，辅助业务部门开展环境风险处置与决策。整套体系的架构如图 1 所示。

国家中心在大数据层次，拥有所有流域的全套上报数据，进行宏观整体的水环境风险掌控，保障全国水环境安全，开展全国性水环境规划；而流域中心拥有特定流域的环境数据，实现流域综合管理，开展流域联动预警；国家与流域的水环境风险管理目标的实现，需要依

赖各行政中心的具体落实，行政中心拥有辖区内所有监控位置的水环境信息，开展环境监控和风险应急决策，并将决策下放到各个业务中心进行执行，业务中心按照决策指示，在可靠的技术和便捷产品的支持下，快速、有效地进行监测、应急处置等工作。基于大数据的环境风险模拟评估与预警体系应用不同中心的数据，解决不同中心的业务需求，满足多中心、多层次、多目标的服务需求。

3.2 环境大数据中心构建

随着我国环境监测体系的不断完善和环境监测数据的积累，目前，我国各级环境管理部门、研究机构已经掌握了大量的监测数据，并初步具备了通过物联网、移动互联网等新渠道，采集环境大数据的能力；在数据的获取方式、获取类别、获取数据量上已经具备构建环境大数据的条件，已拥有较为成熟的大数据挖掘、人工智能、分布式计算，大数据关联分析、趋势分析、空间分析等技术。但由于我国信息化的标准规范体系不健全，生态环境信息资源目录标准尚未构建，各部门数据开放和共享不够等问题，我国环境大数据中心建立仍需要破除数据孤岛、建立数据开放目录，形成从下到上的完整数据开放和共享计划。在国家和部门软硬件和政策标准等条件足够的条件下，构建环境大数据中心的框架如图 2 所示。

构建的环境大数据中心，通过市行政区的各级业务部门的环境大数据监测网络体系，如自动站、人工实测、卫星遥感等和业务中心自带的网络数据采集方式收集服务于业务部门职能的数据，打破行业壁垒，进行不同职能部门的数据共享，并接入省级环境大数据中心；省级环境大数据中心打破行政壁垒，开展不同行政区数据共享，同时将数据接入流域环境大数据中心；各流域环境

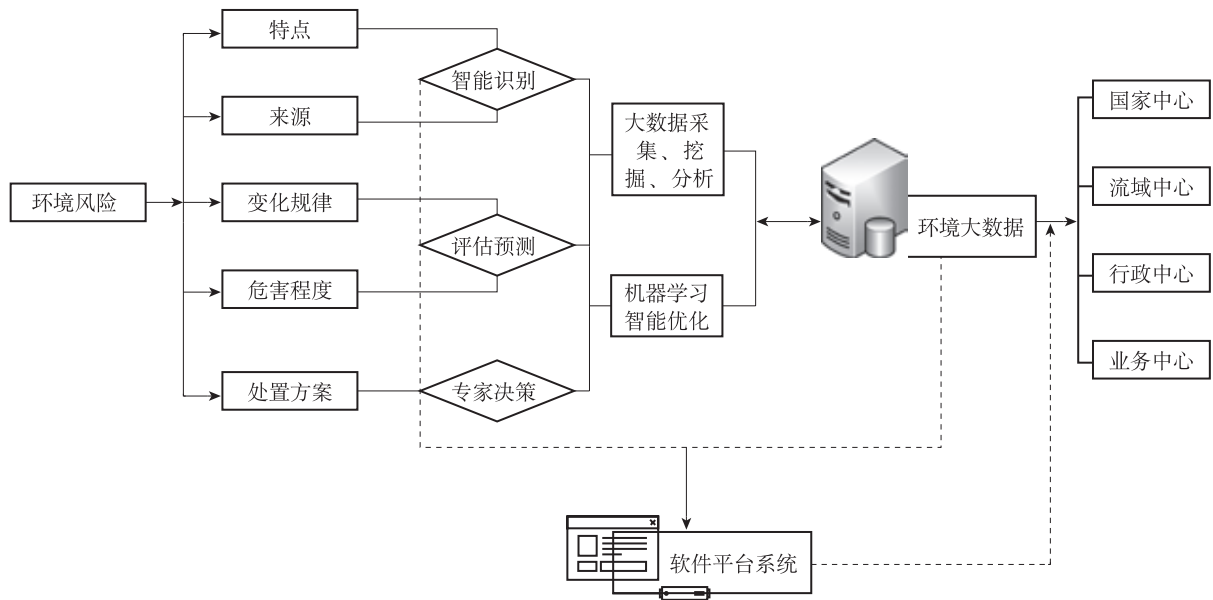


图1 基于大数据的环境风险模拟评估与预警体系总体架构

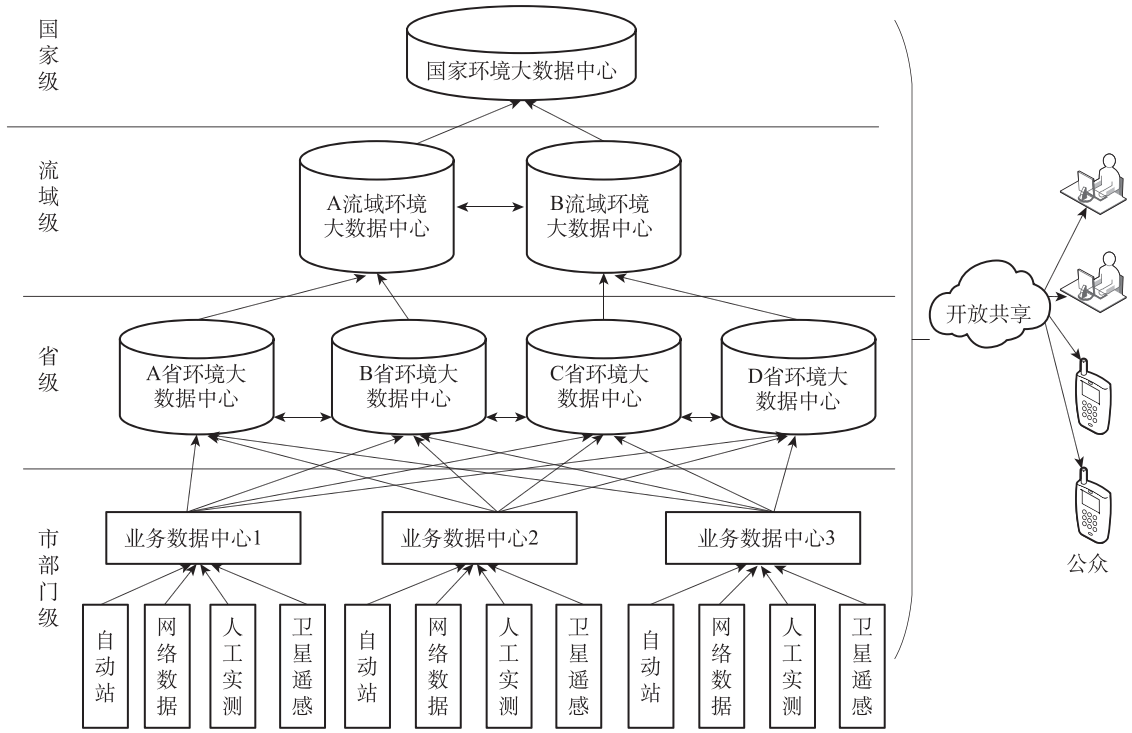


图2 环境大数据中心构建框架

大数据中心在进行数据共享的同时，接入国家环境大数据中心。各级数据中心实行一定程度上的生态环境数据互联互通和开放共享。

3.3 环境风险智能识别

环境风险的识别是环境风险模拟评估与预警的前提，包括造成环境风险的风险物质、涉及区域以及风险来源的识别与追踪。在水环境中，风险包括突发水污染风险、累积性水环境风险两个方面。其中，突发水污染风险指短时间内突然对水体水质产生污染的事故，如排污口偷排、危险品运输船泄漏等；而累积性水环境风险则是指由于环境长期的不健康状态所累积的，渐进造成水体污

染的水环境风险，包括水华、生物重金属累积等。基于大数据的环境风险智能识别，以实时环境监测数据为基础，通过对水质水生态数据的识别，判断环境风险是否存在，确定环境风险存在的位置、时间、主要风险污染物；通过对实时遥感影像数据的知识学习、水色分析、图像识别，追踪环境风险发生的源头、风险源过程，为环境风险的评估与预警提供支撑。环境风险智能识别技术体系如图 3 所示。

3.4 环境风险模拟评估预测

环境风险模拟评估预测包含两方面的内容：一是预测未来可能发生的环境风险、开展先期预警，以阻止或

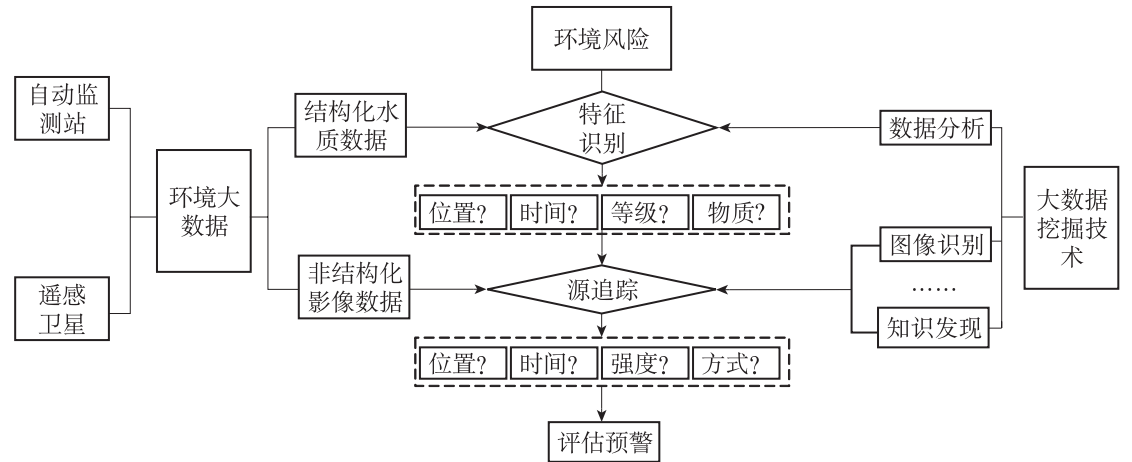


图3 环境风险智能识别技术体系

减缓环境风险的发生；二是在环境风险发生后进行风险发展趋势、影响程度和影响持续时间的预测预报，为环境风险处置提供支撑。大数据环境风险评估与预测应该对整体的区域和样本进行分析，是一个关乎整体流域完整水循环和水质演化的问题。数值模型是环境风险评估预测的主要计算工具，而与模型相关的辅助计算方法则是提高数值模型可用性和实时性的关键。

3.4.1 数值模型体系

完整的水循环过程涉及气象、水文、水动力水质和水生态过程，对应的数值模型包括气象模型（空）、陆地模型（地）、水动力模型、水质模型和水生态模型以及突发事件模型（水）。本文依据水环境风险发生前后的管理需求，分别构建流域的空、地、水的模型模块，将其分为条件模型和应用模型。其中，气象模型为其余模型提供气象条件；陆地模型为水动力模型、水质模型水生态模型提供气象条件；水动力模型为水质模型、水生态模型的条件模型；突发事件模型包含自由的水动力和污染物扩散模拟模块。各子模型之间既相互耦合又能独立模拟，模型间的耦合状态如图 4 所示。

在水环境模型体系中，条件模型是针对全流域模拟需求而设定的，其模拟区域固定，只对全流域进行整体的模拟；模型的模拟边界范围、数据网格是固定的。条件模型一方面是针对水环境管理业务化的需求，实现对整个流域全局范围的在线实时的水环境模拟；另一方面为局部精细模拟，即模型应用节点的模拟提供边界条件。应用模型是针对流域水环境管理中具体区域内的实际应用而设定的，用于评价分析流域内局部区域气象水文水质状况、开展突发事件应急决策。应用模型按需进行模拟分析，其模拟的水域范围不固定，可由用户自行指定；其模拟的网格精度较高、相对于条件节点为细网格；水动力水质水生态的模型参数、水环境指标在模型中会有预置值，但用户可以根据实际情况进行修改。在模型开展精细全局模拟的过程中，当前最重要的问题是要突破模型计算效率与计算精度之间的矛盾，而这需要一套高效能的计算方法予以支撑。

3.4.2 模型高效能计算

大数据理论体系下，解决计算效率问题的最优模

式是 MapReduce 模式，通过将大区域计算的大问题划分为大量互不干扰的小区域计算的小问题，利用集群计算予以解决。为提高模型的模拟速度和精度，基于 MapReduce 模式，本文构建的模型采用 OpenMP+MPI 的混合并行计算模式进行模拟计算，混合并行计算模式如图 5 所示。

将模型计算范围进行分块，并分配到不同的机器（进程）上进行计算，利用 MPI 将链接不同进程进行数据交换，模型内部以 OpenMP 的 fork-join 模式对循环体进行并行，形成模型内部、模型计算分块间的混合并行计算方式，以提高模型的模拟速度。

3.5 风险预警决策支持系统

风险预警决策支持系统的核心是满足业务化的管理需求，既能开展常规业务化的风险预测评估，也能实时在线地进行业务情景分析。具体的需求即软件系统能在日常情况下，定时预测并发布既定敏感位置的未来几天的水质状况和风险等级等信息；在检测到突发风险的情况下，能迅速预报预测风险事故在一天内的变化趋势，提出最优的应急处置方案。本文依据业务化环境风险评估与预警的业务化管理需求，针对大数据产生、收集、存储和共享的特点，构建了流域水环境风险预警决策支持系统，系统包括控制中心、计算中心、数据中心和业务中心，系统架构如图 6 所示。

控制中心是整个系统的头脑，管理系统运行流程，分配系统资源，使各中心协同分工合作，快速、有序地完成常规业务和情景业务；同时监控系统运行，处理系统故障，实现集群管理系统的自动化、智能化业务处理。数据中心单独分布在国家、流域和行政单位的数据云中，能自动收集、处理水利、环保、气象数据，实现对历史数据、模型运算数据的存储管理，为计算中心的评估计算和业务中心数据共享发布提供数据支撑。计算中心包括条件模型和应用模型体系，条件模型自动定时开展全流域模拟计算，生成全流域水文水质水生态数据和敏感位置预测预报结果。应用模型随时待命，在有情景业务需求时，获取条件模型结果，依据情景条件进行应急评估与决策运算。业务中心提供业务管理部门所需的、可视化的水环境监测、预报和预警信息，完成水环境警情

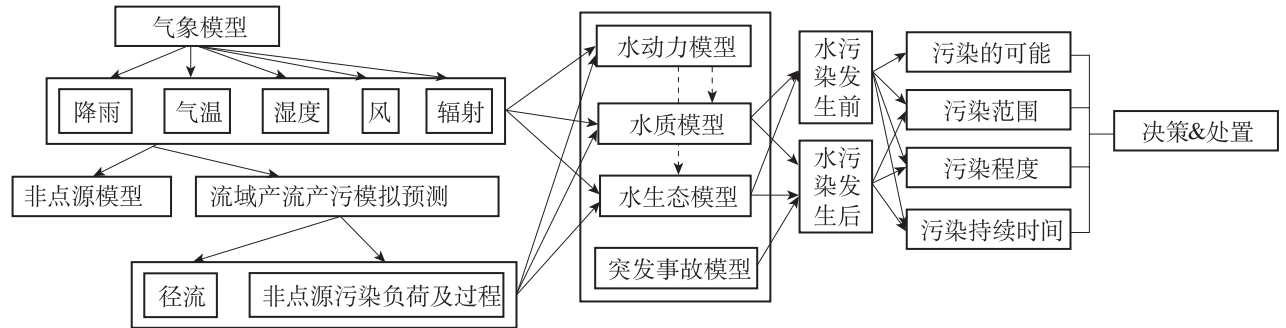


图4 流域水环境模型体系

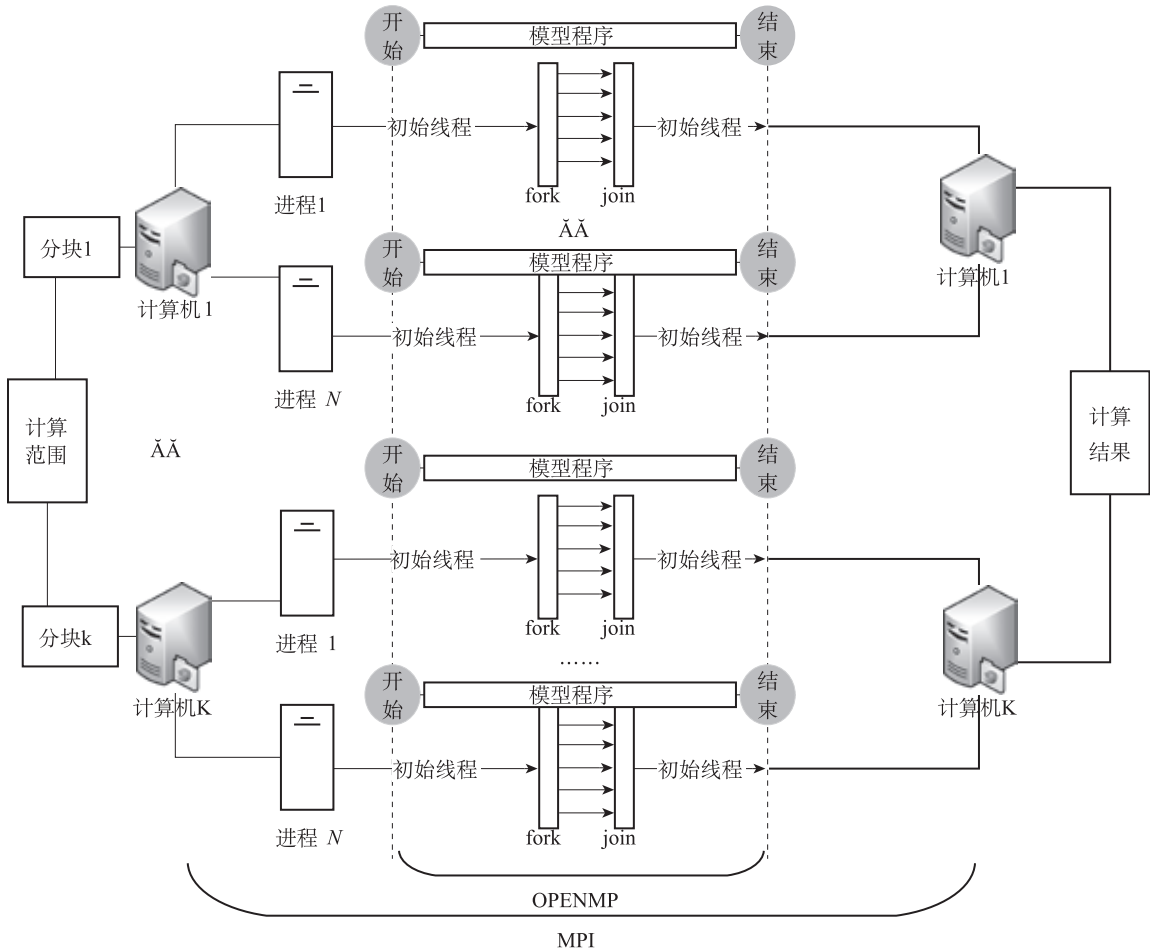


图5 模型的混合并行计算模式

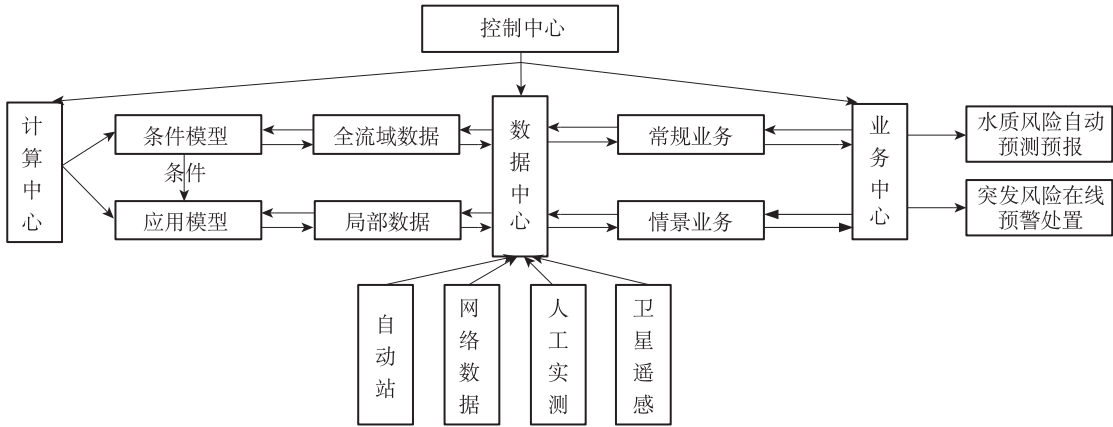


图6 风险预警决策支持系统架构

信息的发布和报告等服务。

4 三峡库区水环境风险评估与预警示范系统

4.1 系统简介

随着经济建设的快速发展,我国水环境风险隐患不断增加,与之相随的突发水污染事故、水华、生态破坏等一系列水环境风险在众多流域频频发生。三峡库区的

库一水一人在我国流域具有典型的代表性,库区水环境面临点源、面源、突发污染等一系列的风险隐患。为保护三峡库区的生态环境,三峡库区相关的水利、水质管理部门和各级政府机构在三峡设置了众多的监测站,形成了三峡库区气象、水文、水质的立体监测网络。不同的科研单位近年来在三峡库区开展了众多的研究,取得了一系列的科技成果。库区已具备构建环境大数据的监测条件、数据条件和技术条件。本文在国家水体污染控

制与治理科技重大专项——三峡库区及上游流域水环境风险评估与预警技术与示范课题的支撑下,基于大数据的环境风险评估与预警体系,构建了三峡库区水环境风险评估与预警示范系统。该系统具有常规水质预测、突发水污染事故风险识别与预测、水华风险评估预测和信息发布等功能。

4.2 业务化应用与案例

三峡库区水环境风险评估与预警示范系统成功部署在四川省环境监测总站、重庆环境科学研究院、湖北省环境监测中心站、长江水利委员会水文局等地方环境管理职能部门和流域管理机构。示范系统自 2015 年业务化运行以来,每天通过自动作业的方式,自动获取水文监测站、气象监测站的最新水文气象信息,进行水文气象预报,并模拟预测未来 72 小时的水动力水质状况,进行常规水质预报,开展水污染风险评估和水华风险评估与预警。自业务化运行以来,示范系统共获得 120 多万条水质自动监测数据、36.5 万多条气象数据,向业务管理人员推送了大约 14 850 条水质预报信息。四川和重庆示范区平台常规水质预报和风险评估的以及短信推送如

图 7 所示。

业务化系统每天预测未来的水质浓度,并对比评估水质自动监测值与模拟值,当模拟值与实测值存在较大误差,且实测值大于先前三天的平均值时,系统自动发出水质浓度超标预警信息。如 2016 年 7 月 1 日 10 时,釜溪河邓关断面总磷实测值为 0.70mg/L,超过平台总磷预测值(0.34mg/L)100.06%,大于先前三天的平均值,系统发出水质超标风险预警。环保部门依据平台信息数据判断可能是上游企业在 6 月 30 日降雨期间有短时高浓度总磷废液偷排所致,当即通知相关单位展开排查。图 8 为四川示范平台中系统预测浓度与水质自动站实测浓度对比。

5 结论

本文分析了环境大数据的实时、大量、非结构化等特点和业务化水环境风险评估与预警的需求。在环境大数据不断发展的趋势下,国家层次、流域层次、行政区层次和业务部门层次的大数据中心将相继建立,并各自完成不同的管理和业务功能。基于此,本文提出了多层次、多目标的环境风险模拟评估与预警的体系架构,确



四川示范系统



重庆示范系统

图7 常规水质预报与风险评估

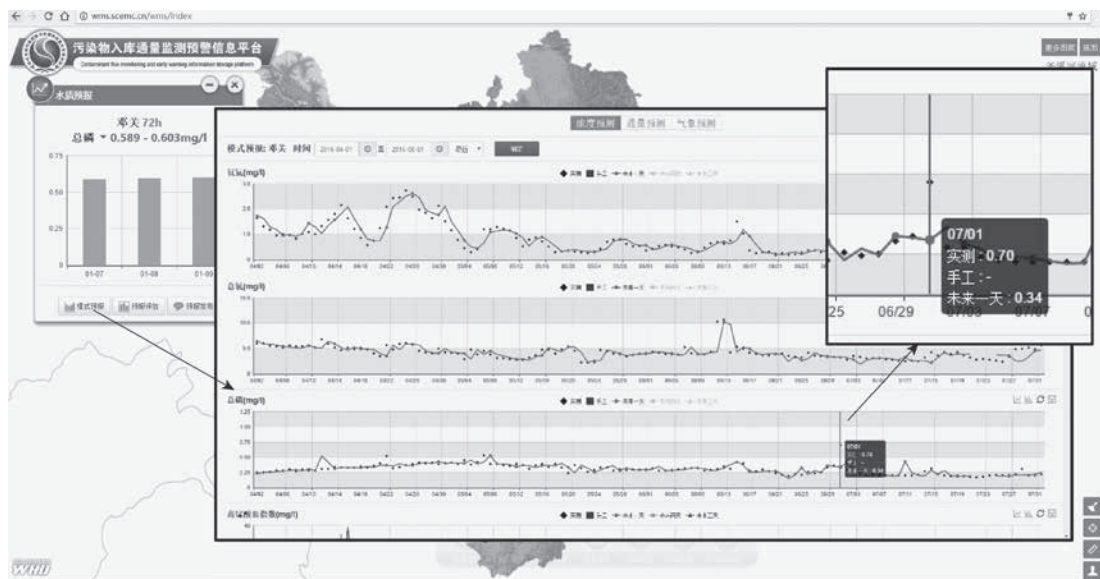


图8 2016年7月1日水质预测浓度与实测浓度对比

定了环境风险的智能识别模式，构建了环境风险模拟预测的空地水一体化模型体系架构，并探讨了基于 MPI 和 OpenMP 混合并行的模型高效能计算方法，提出了一套包含控制中心、计算中心、数据中心和业务中心的风险预警决策支持系统架构。最后，本文以三峡库区水环境风险评估与预警示范系统为例，对其业务化的效果进行了介绍。本文所研究的在大数据背景下，流域水环境风险模拟技术与业务化系统的体系框架，对新形势下流域风险评估与预警技术的研发、技术体系的构建具有一定的借鉴意义。

参考文献

- [1] 环境保护部办公厅. 关于印发《生态环境大数据建设总体方案》的通知 [EB/OL]. (2016-03-08)[2016-03-08]. http://www.zhb.gov.cn/gkml/hbb/bgt/201603/t20160311_332712.htm.
- [2] 佚名. 国务院办公厅印发《生态环境监测网络建设方案》[J]. 中国环境管理, 2015(4): 56-56.
- [3] KOSSENKO M M, DEGTEVA M O. Cancer mortality and radiation risk evaluation for the Techa river population[J]. Science of the total environment, 1994, 142(1-2): 73-89.
- [4] 刘新华, 洪浩林, 吕春玲, 等. 建设项目环境风险评价及应用 [J]. 江苏环境科技, 2004, 17(4): 37-38.
- [5] 孙钰. 落实应急预案: 提升环境应急能力 [J]. 环境保护, 2006(5): 38-41.
- [6] XU Y H, PENG H, YANG Y Q, et al. A cumulative eutrophication risk evaluation method based on a bioaccumulation model[J]. Ecological modelling, 2014, 289: 77-85.
- [7] 张万顺, 徐艳红. 基于水质目标的水环境累积风险评估模

型 [J]. 环境影响评价, 2013(5): 51-54.

- [8] WANG Y G, ZHANG W S, ENGEL B A, et al. A fast mobile early warning system for water quality emergency risk in ungauged river basins[J]. Environmental modelling & software, 2015, 73: 76-89.
- [9] LI W X, ZHANG Y C, LIU Z, et al. Outline for establishment of the Taihu-Lake Basin early warning system[J]. Ecotoxicology, 2009, 18(6): 768-771.
- [10] STOREY M V, VAN DER GAAG B, BURNS B P. Advances in on-line drinking water quality monitoring and early warning systems[J]. Water research, 2011, 45(2): 741-747.
- [11] BURCHARD-LEVINE A, LIU S M, VINCE F, et al. A hybrid evolutionary data driven model for river water quality early warning[J]. Journal of environmental management, 2014, 143: 8-16.
- [12] 张颖, 高倩倩. 基于灰色模型和模糊神经网络的综合水质预测模型研究 [J]. 环境工程学报, 2015, 9(2): 537-545.