



马金锋、饶凯锋、王子健

中国科学院生态环境研究中心
中科水质物联网研发中心
无锡中科水质环境技术有限公司

北京 2019.03.06

计算赋予决策智慧

Modelling Empower Wisdom



Problem 解决的痛点

监测

- 能自动生成日常办公报表?
- 能及时预警水质超标事件?

预测

- 未来水质会变成什么样?
- 黑臭水体治理工程效果怎么样?

决策

- 水质超标到底什么原因造成?
- 采用什么措施保障水质达标?



Solution 解决的方案



水环境模拟大数据平台

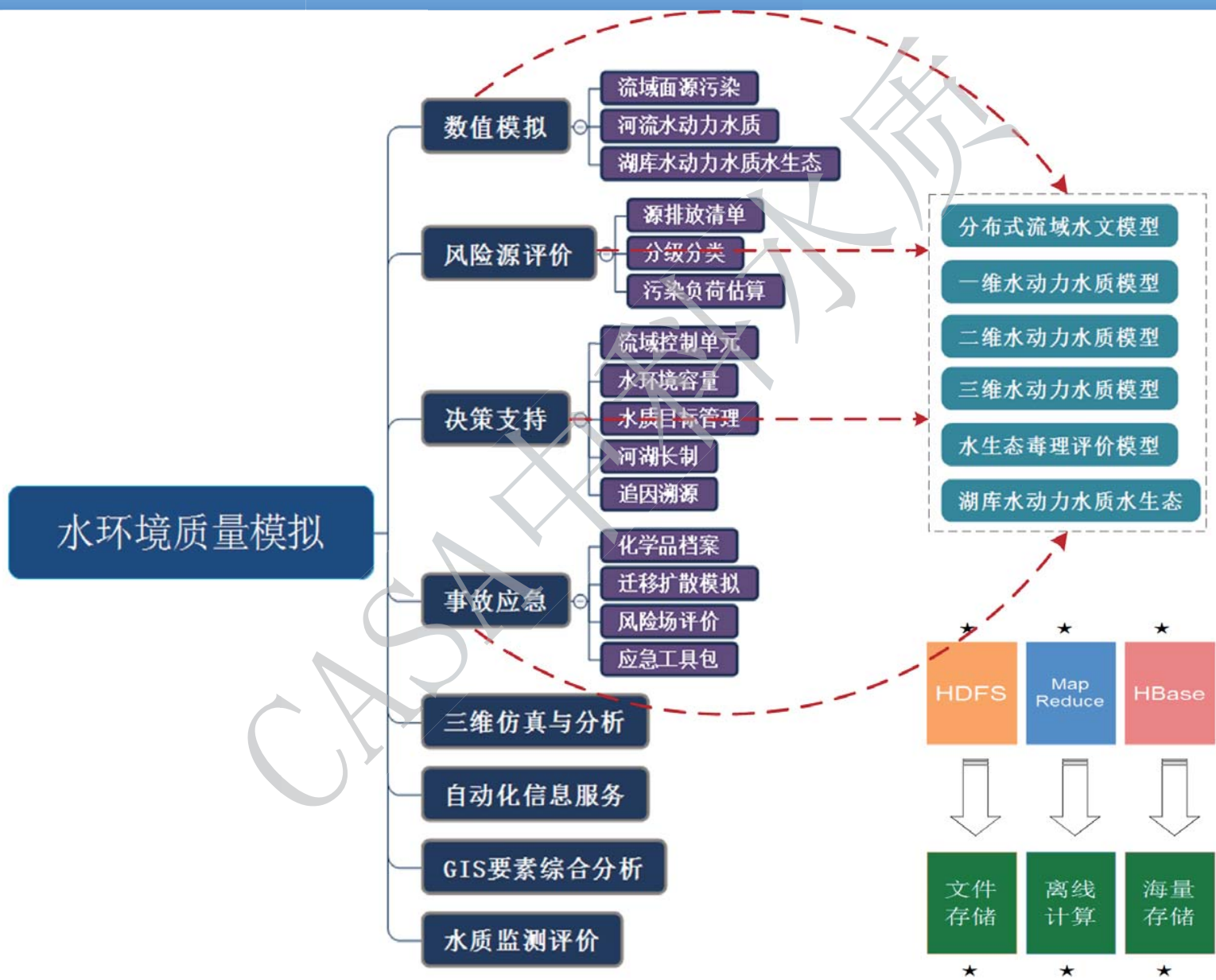


Domain 应用领域





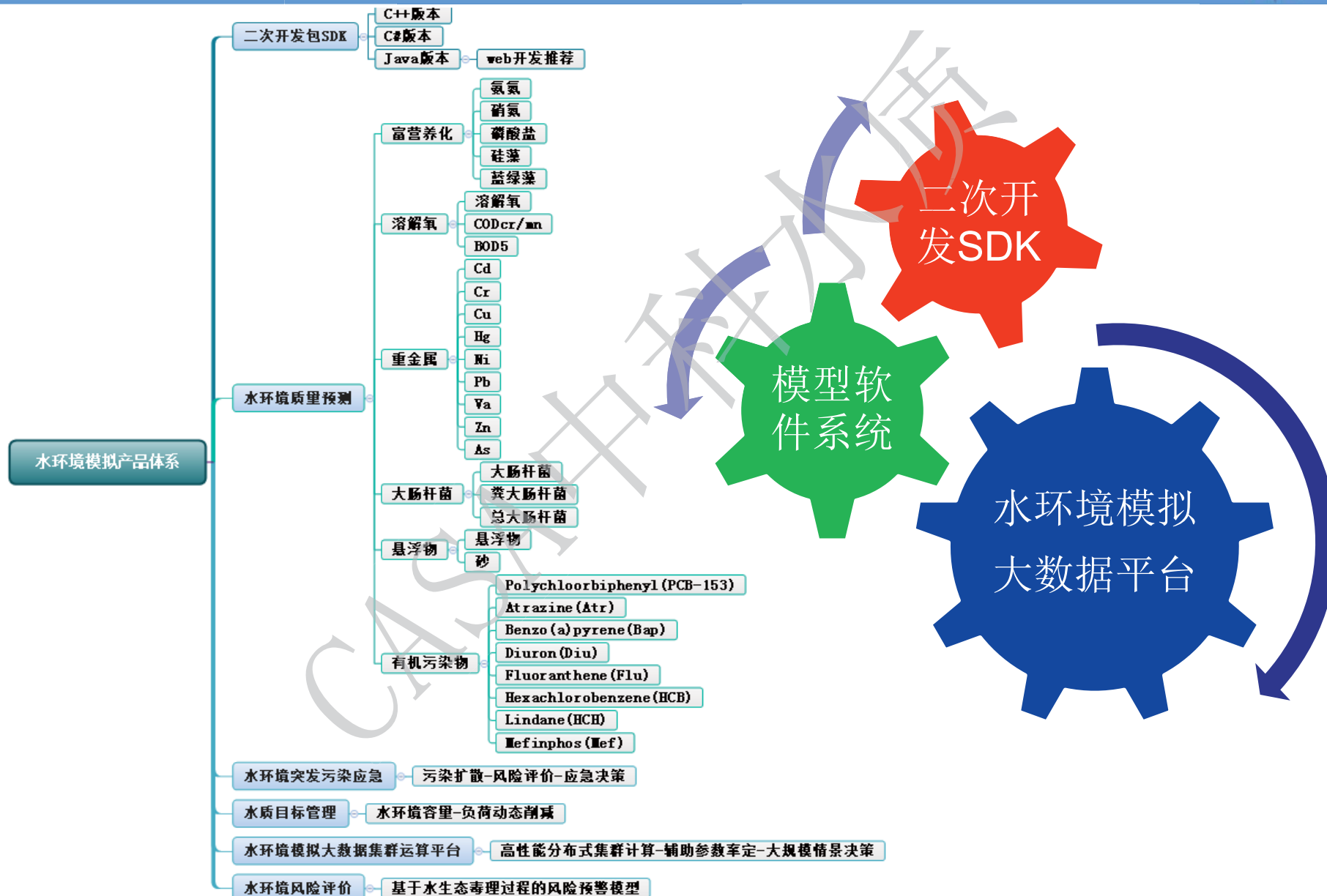
Framework 服务框架



Architecture 服务体系



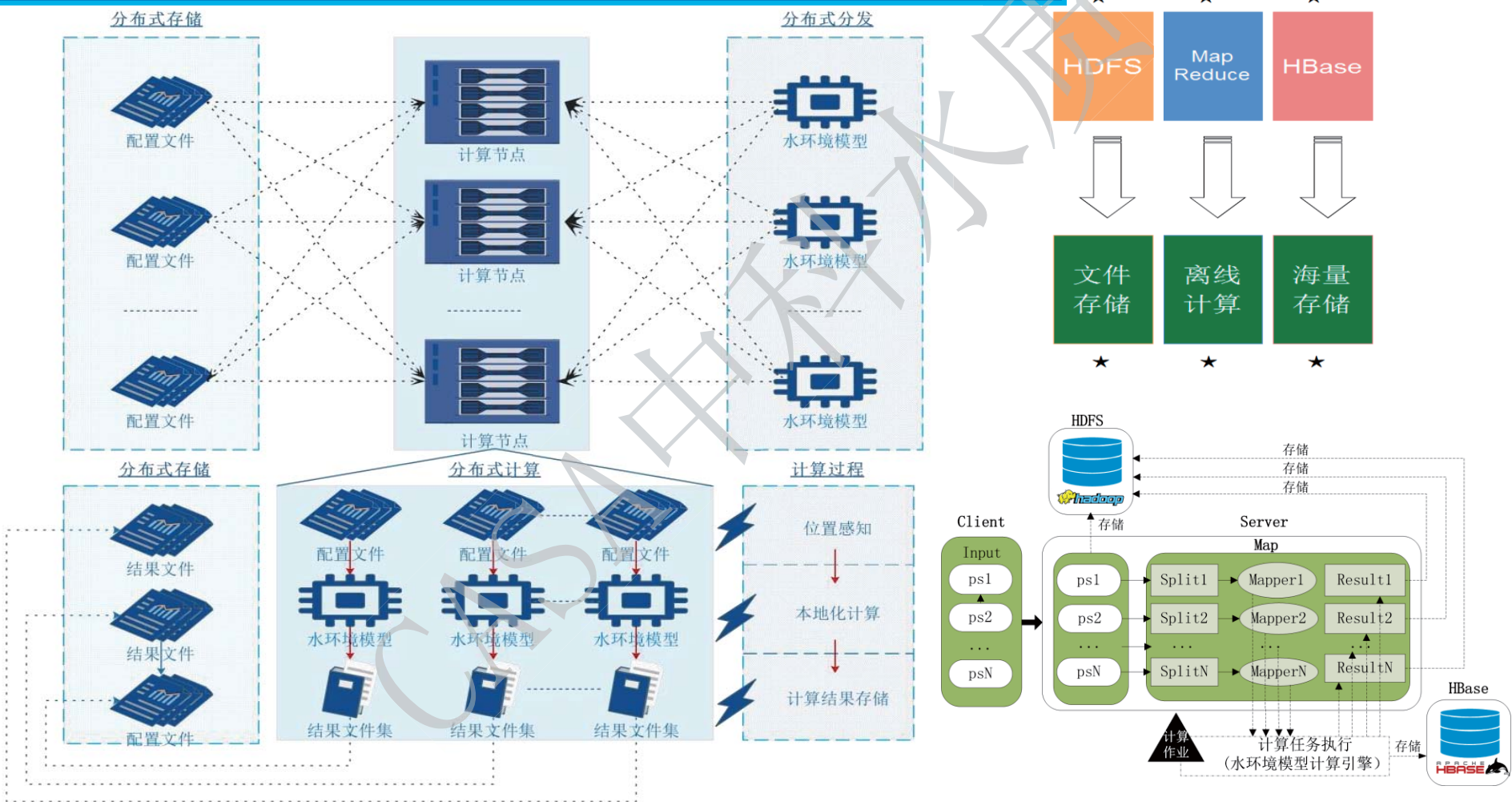
Product 产品体系





Technology 技术框架

解决：数值模拟模型在大数据环境下适应性问题



应用： 参数率定 敏感分析 不确定性 情景预测

Platform 高性能计算平台



分布式水文模型

产流汇流
产污汇污
控源减排

水生态毒理模型

水环境生态风险评价

水动力-水质模型

水环境质量预测
水质目标管理
水环境污染应急



计算平台





Service 决策服务

- (1) 水环境监测评价分析服务
- (2) 水环境模拟预测服务
- (3) 水环境污染风险动态预警服务
- (4) 水环境水质目标管理服务
- (5) 突发水污染事故应急处理处置服务
- (6) 水环境污染定位溯源服务





Application 应用案例

流域水质目标管理

- 海河流域北运河水系水环境实时调控与决策支持系统

水环境质量预测

- 贵州桐梓新桥水库水环境质量预测系统

水污染应急决策

- 北京市副中心水环境应急处置平台

环境工程影响评价

- 水文流场优化设计：巴音郭楞蒙古自治州博斯腾湖水体达标方案

其他

- 河北省唐山市陡河水库水质监测与生物预警系统
- 河北省石家庄市岗南水库水质监测与生物预警系统

①流域水质目标管理-立论及内涵



立论：粗放式治理和粗放式管理是我国水环境治理进展缓慢的一个重要原因。

内涵：

- ①掌握流域水系产汇污过程与水环境演变规律；
- ②建立水环境与污染负荷实时调控与决策支持系统；
- ③实现河流水质从“静态总量控制”到“动态目标控制”

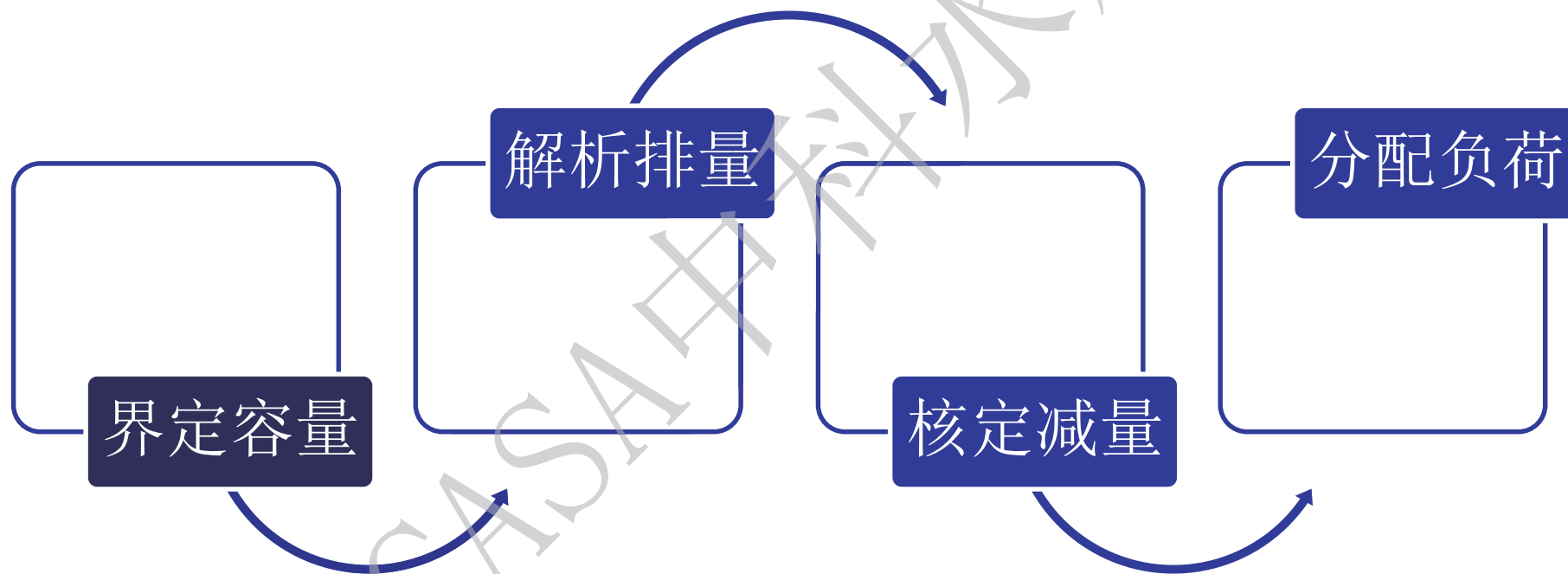
内容：

- ①乡镇级行政产汇污单元入河污染负荷量化计算；
- ②河流水环境承载力计算及入河负荷动态分配计算；
- ③建设流域水环境实时调控决策支持系统平台

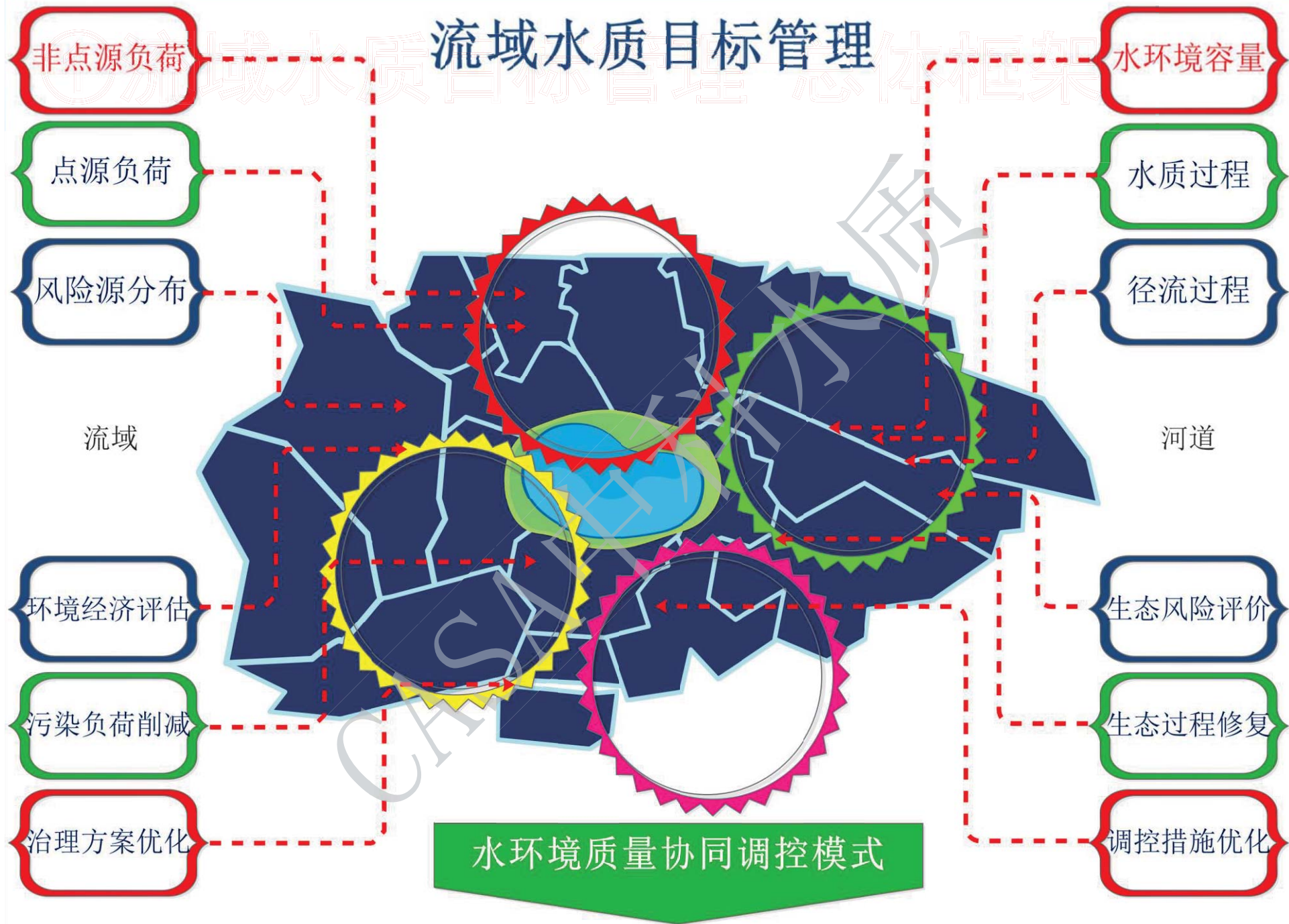
特色：

- ①建立河道水质目标与入河源强之间动态响应关系
- ②实现面向容纳水体水质目标的水环境承载能力计算
- ③基于社会经济利益相关者的负荷动态优化分配
- ④通过多元数据融合及数据模型同化实现水环境治理方案优化和实时调控

①流域水质目标管理-总体思路



流域水质目标管理



①流域水质目标管理-项目简介



项目名称：海河流域北运河水系水环境实时调控与决策支持系统

项目目标：构建北运河水系水环境质量的协同调控模式，开发针对重要水质指标的调控模型，集成在线监测网络和流域数字化技术，建立水环境实时调控的决策支撑系统，辅助优化示范工程方案，为流域“控源减排”提供技术支持。

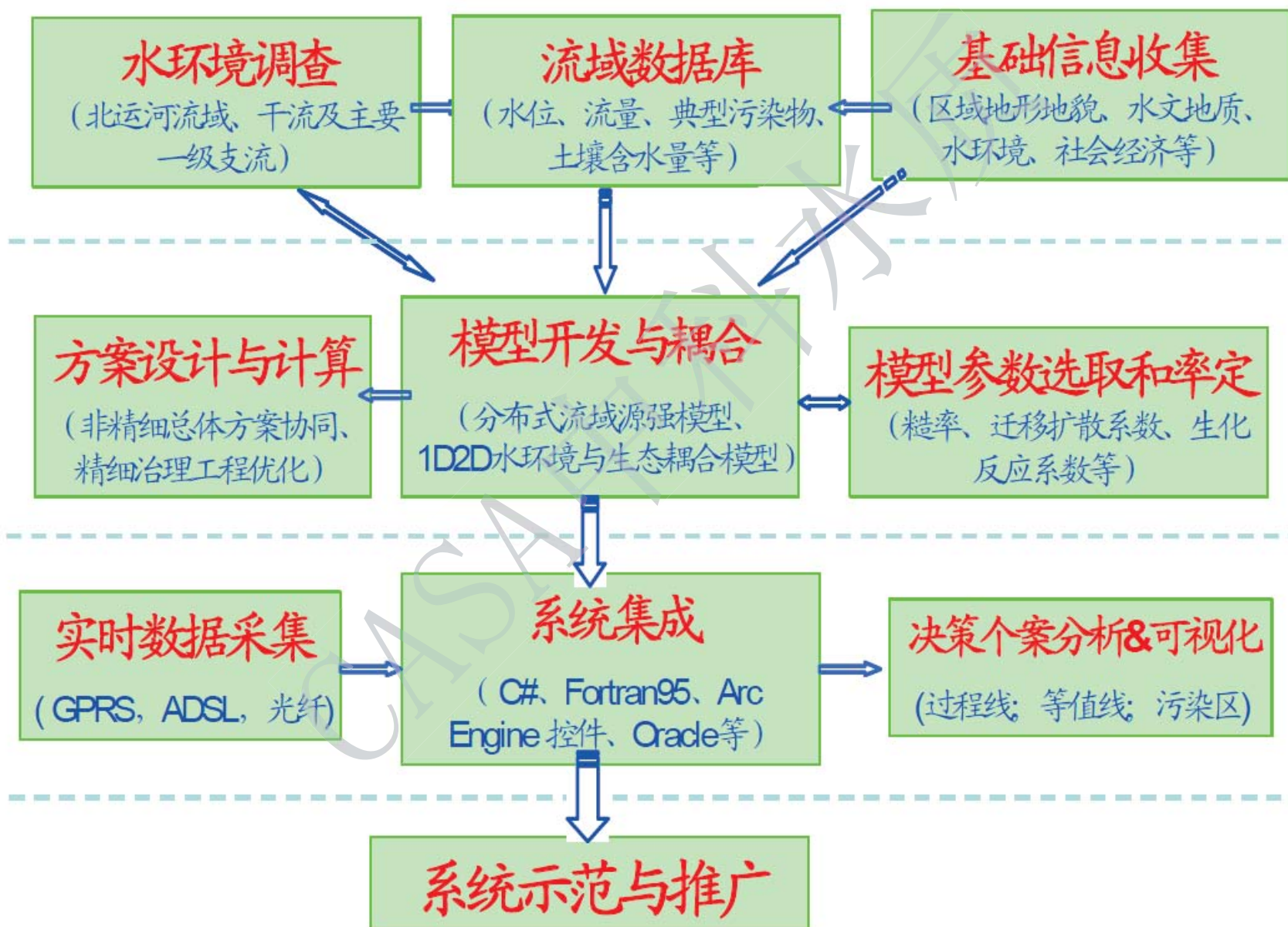
建设内容：围绕氨氮，COD 和重金属，

(1) 开发北运河流域分布式水文模型、入河污染负荷模型、河道一维二维耦合水质模型；

(2) 计算北运河水系各河段特征污染物的**负荷削减量**及其安全区间，实现非常规水源与受纳水体之间水质的动态匹配；

(3) 建立北运河水系水环境经济评估模型，实现以模型为基础的**负荷削减量测算和治理方案优化**

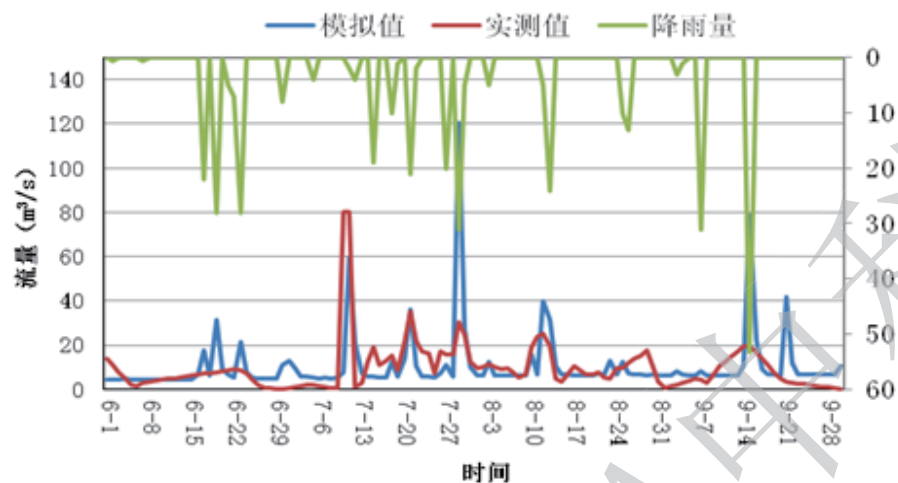
①流域水质目标管理-技术路线



①流域水质目标管理-模型率定

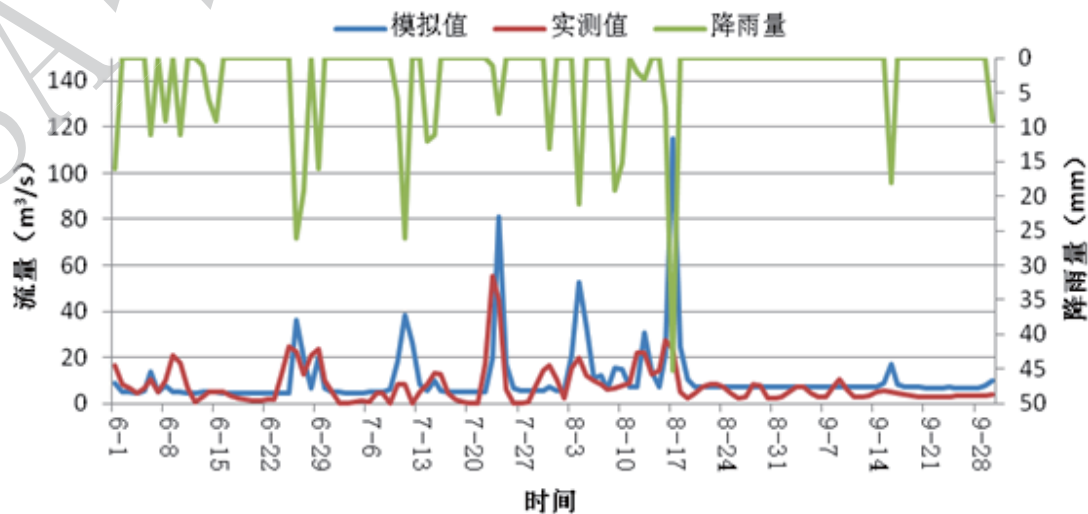


□ 分布式水文模型-模型率定



➤ 榆林庄6月1号到9月30号实测流量

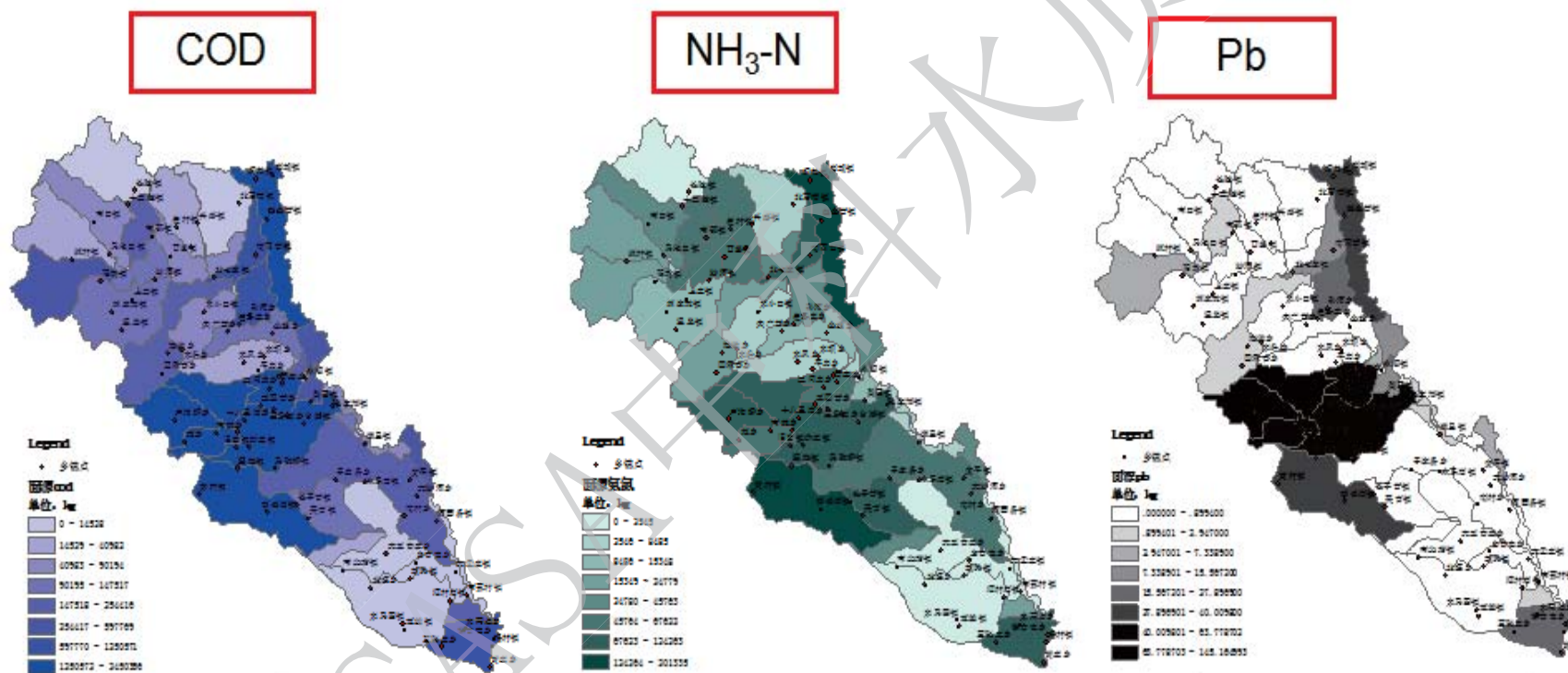
➤ 榆林庄6月1号到9月30号实测流量



①流域水质目标管理-负荷估算



□ 非点源产污模型-估算结果



| 污染物种类 | 城镇生活 | | 工业 | | 地表径流 | |
|-------|-----------|--------|-----------|--------|----------|--------|
| | 入河量 (t) | 百分比 | 入河量 (t) | 百分比 | 入河量 (t) | 百分比 |
| COD | 2748003.5 | 12.43% | 4095922.9 | 18.52% | 15267164 | 69.05% |
| 氨氮 | 120277.35 | 6.77% | 1111745.9 | 62.62% | 543418.4 | 30.61% |
| Pb | 0 | 0% | 0 | 0% | 646.9 | 100% |

①流域水质目标管理-水环境容量



| 河流 | 水质类别 | COD | | 氨氮 | |
|----------|------|---------------|-----------|---------------|--------------|
| | | 环境容量 (t/a) | 削减量 (t/a) | 环境容量 (t/a) | 削减量 (t/a) |
| 沙河闸-沙子营 | IV | 2573 | 4569.62 | 111.26 | 105.97 |
| 沙子营-北关闸 | V | 9103.8 | -4149.82 | 226.50 | 32.54 |
| 北关闸-老米店闸 | V | 41808.2 | -31284.41 | 2084.94 | -1435.03 |

根据污染物降解规律计算各单元的水环境容量及削减量

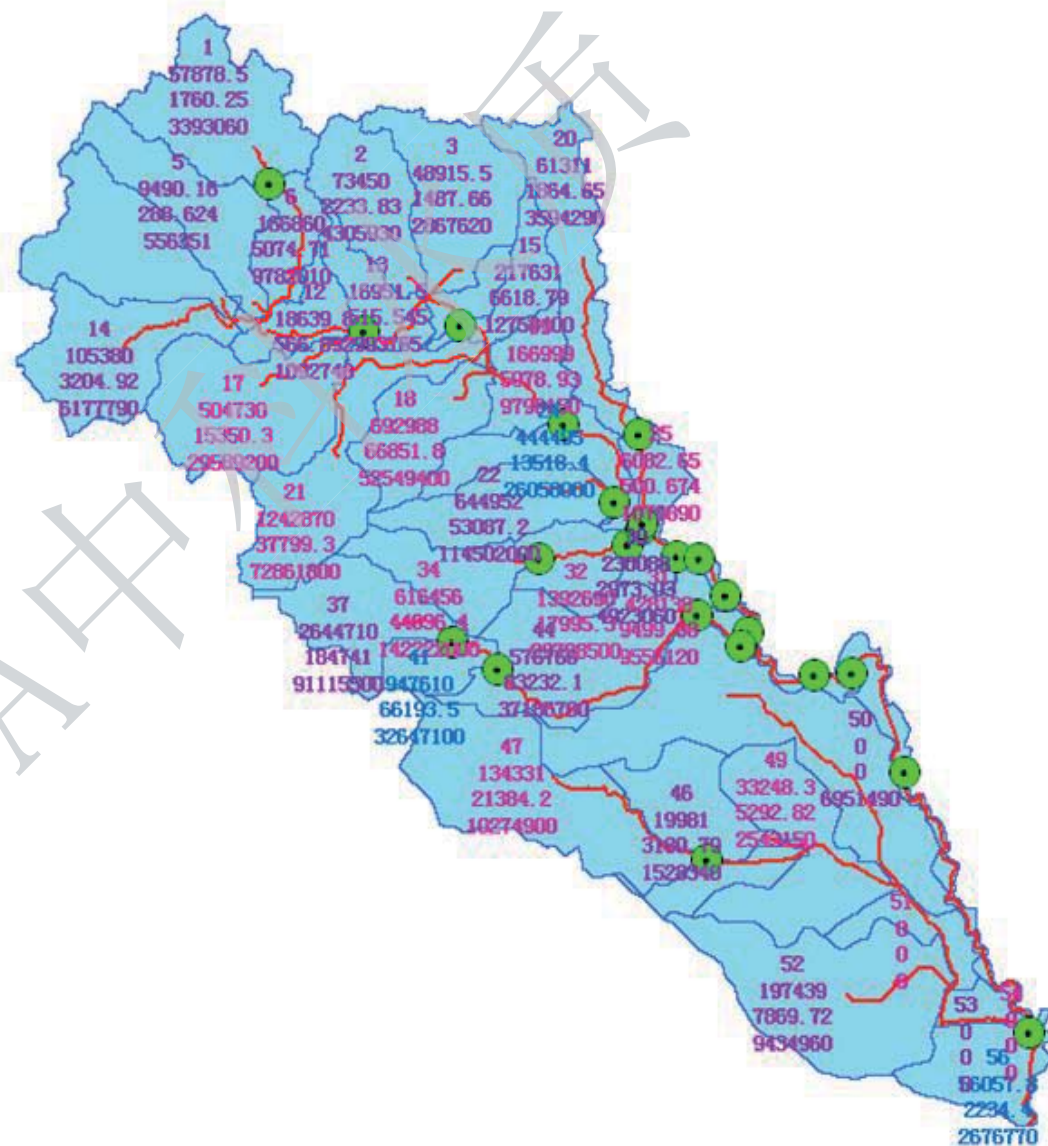
①流域水质目标管理-负荷削减



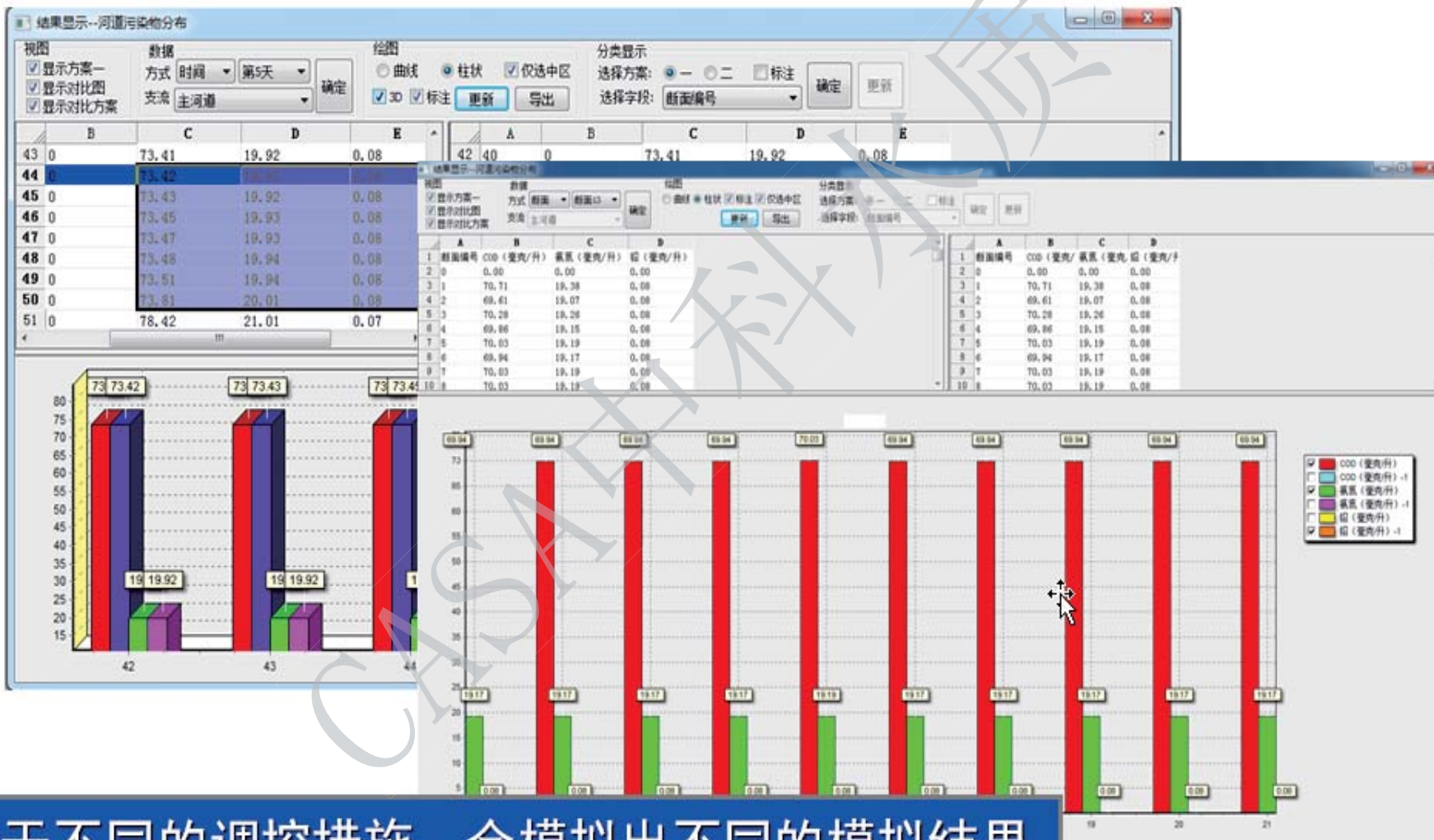
■ 负荷核算结果

各子流域污染物入河量
单位: t/a

每个子流域数据分别为:
第一行: 子流域编号
第二行: COD入河量
第三行: 氨氮入河量
第四行: 污水入河量



①流域水质目标管理-决策系统

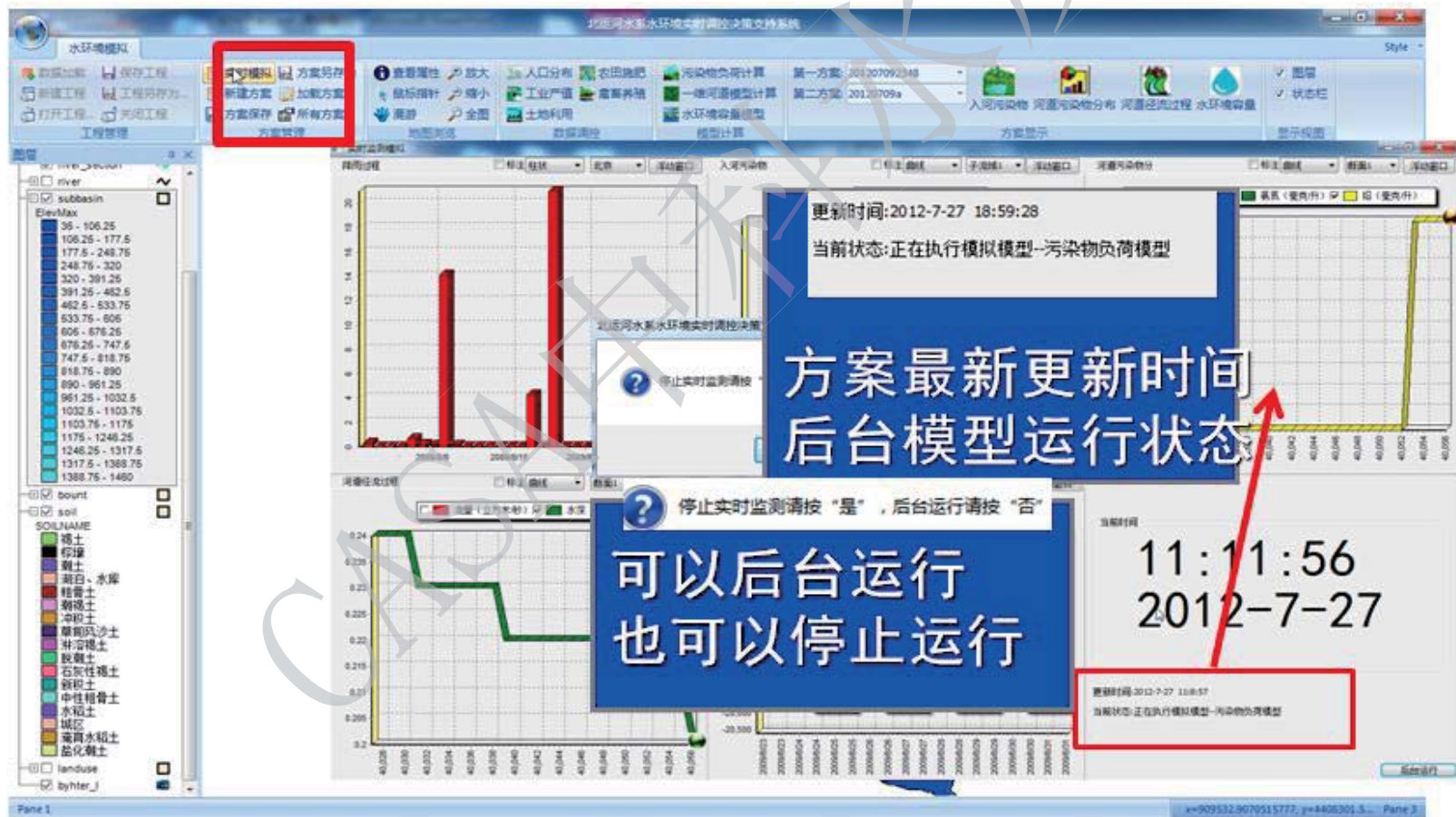


---方案的对比

①流域水质目标管理-负荷调控

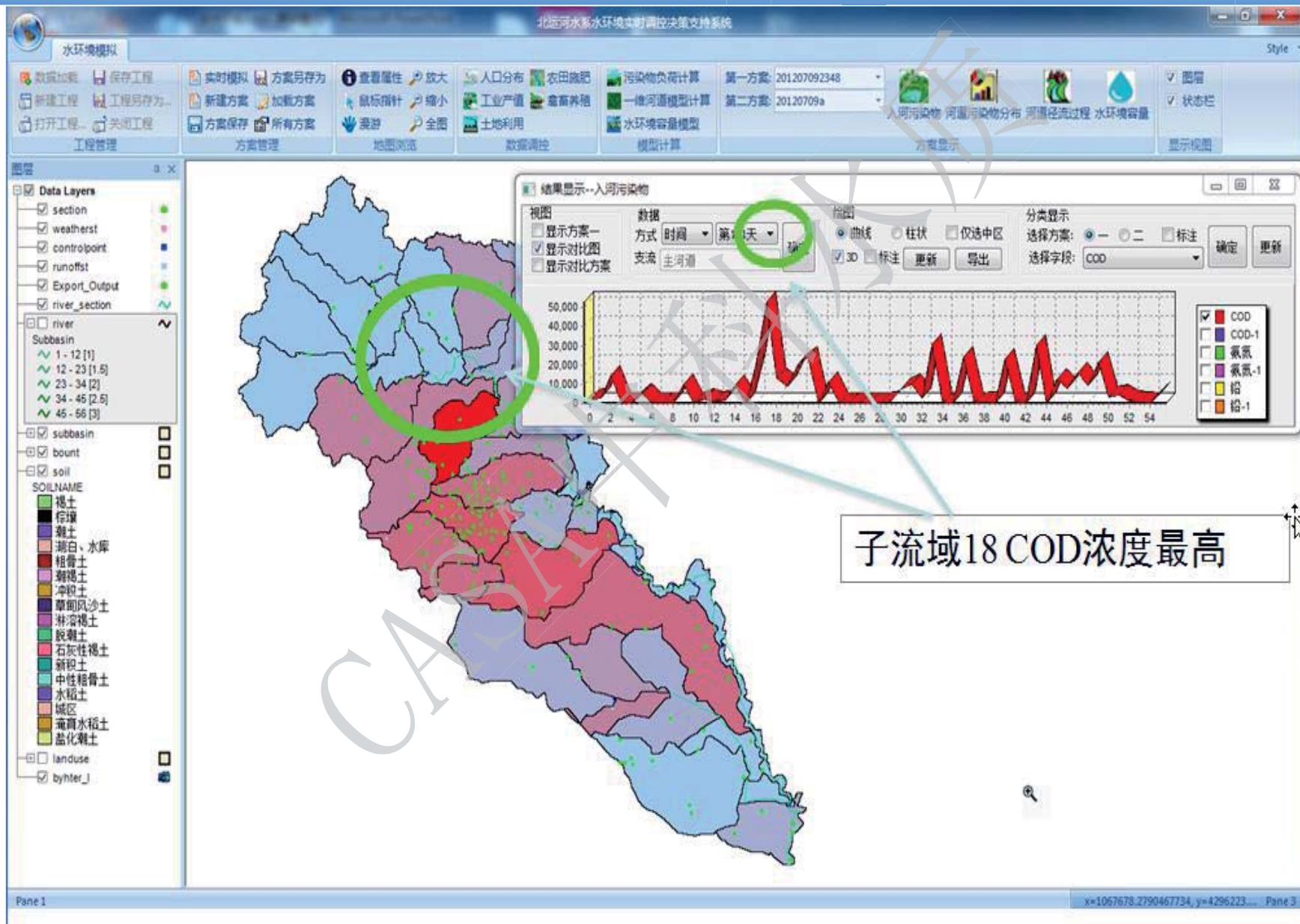


实时模拟和逐日负荷调控



根据实时监测数据进行滚动模拟，辅助逐日调控

①流域水质目标管理-业务化运行



②水环境质量预测



Browser tabs: Hadoop NameNode m, 金桥水库水环境模拟大数

Address bar: file:///C:/tomcat7/webapps/water/index.html

Navigation: 应用 | hadoop | 文章检索 | matlab

Menu: 首页 | 模型展示 | 实时模拟 | 辅助决策 | 客户端产品

Search: 设置 | 管理员

水库基础数据

2018年1月15日 9时40分47秒

| Parameter | 7-10-01 00:00:00 | 2017-10-03 08:00:00 | 2017-10-05 16:00:00 | 2017-10-08 |
|------------|------------------|---------------------|---------------------|------------|
| waterdepth | ~5 | ~15 | ~25 | ~35 |
| waterlevel | ~10 | ~20 | ~30 | ~40 |
| averaged-x | ~5 | ~15 | ~25 | ~35 |
| averaged-y | ~5 | ~15 | ~25 | ~35 |

场景参数

| Parameter | 0 | 160 | 320 | 480 | 640 | 800 | 960 | 1120 | 1280 | 1440 | 1600 |
|-----------|-----|-----|------|-----|-----|-----|-----|------|------|------|------|
| inflow | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |
| naoh | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |
| nh3 | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |
| outdiba | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |
| outflow | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |
| phenol | ~20 | ~80 | ~100 | ~80 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 | ~20 |

污染排放浓度

2017-11-01 17:00:00 none=0

水环境模拟大数据平台

2018年1月15日 9时40分47秒

Map controls: 主页 | 模拟 | 测面 | 测距 | 比例 | 图例 | 取消 | 影像图

Map labels: 田沟, 鹿水坝, 石老婆, 陈溪沟, 米粮坝, 龙会杨村

水库场景 值

| Parameter | Value |
|-------------|--|
| 场景数量 | 1728 |
| 网格 | 标准 |
| 出入口 | inflow, outdiba, outflow |
| 上游入口流速 | [60, 100] |
| 堤坝出口流速 | [-50, -70] |
| 下游出口流速 | [-10, -30] |
| 观测点 | 观察点A, 观察点B, 观察点C |
| 释放点 | 无 |
| 干燥点 | 无 |
| 薄坝 | 无 |
| 浮标 | 无 |
| 断面 | 无 |
| 时间周期 | 2017-10-01 00:00:00 2017-10-08 00:00:00 |
| 平均水位 | 10m |
| 模拟时间步长 | 1min |
| 输出时间频率 | 30min |
| 当地时区 | 8 + GMT |
| 盐度 | 标准 |
| 温度 | 标准 |
| 风 | 标准 |
| 污染物 | phenol, NH3, NaOH |
| 苯酚 (phenol) | [1.2, 3.0] |

Taskbar: 金桥水库水环..., water_前台, gisflow, water - [C:\to..., 年度工作总结(...), 水流模拟, 可视化系统, 192.168.0.100..., 6.jpg - 画图, 9:40 星期一, 2018/1/15

贵州桐梓新桥水库水质预测系统

③水环境污染应急决策支持系统



北京城市副中心水环境应急处置管理平台

11:34:19 星期二 2018-7-10

区域简介

| | | |
|-------|---------|-------------------|
| 河道长度 | 河道水面宽 | 水深 |
| 3.9km | 20~156m | 1.5~2m |
| 正常蓄水位 | 最高蓄水位 | 调蓄库容 |
| 18m | 20m | 55万m ³ |

常规五参数

| | | |
|-----|----|-------|
| PH | 6 | 无量纲 |
| 水温 | 24 | °C |
| 浊度 | 3 | NTU |
| 电导率 | 16 | us/cm |
| 溶解氧 | 30 | mg/L |

气象水文

城市: 北京

日期: 10日星期二 南风

温度: 24 高温 30°C 低温 22°C

注意: 各项气象条件适宜, 发生感冒机率较低。但请避免长期处于空调房中, 以防感冒。

指数: 空气污染指数 55

生物预警

- (1) 鱼类生物预警技术
- (2) 藻类生物预警技术
- (3) 菌类生物预警技术
- (4) 藻类生物预警技术

检测方法

- 胺类
- 有机金属类
- 盐类
- 无机金属类

法律法规

| 名称 | 发布日期 | 文号 |
|---------------|------------|-----------------|
| 行政处罚听证告知书 | 1974-06-11 | 环法字[2014]号 |
| 北京市水资源管理条例 | 1992-01-01 | 市人大常委会第23号 |
| 北京市水污染物排放标准 | 1985-11-01 | 京政发(1985)143号 |
| 中华人民共和国水污染... | 1985-09-24 | 环法字[2010]号 |
| 北京市水环境区域补偿... | 2014-10-31 | 京政办发 [2014] 57号 |
| 北京市政府信息网上公... | 1974-06-11 | 环法字[2010]号 |

历史案例

| 名称 | 日期 | 地点 | 污染物 | 类别 | 原因 |
|---------|--------------|---------|---------|------|---------|
| 广西排污... | 1974-06-1... | 广西龙江河 | 镉 | 重度污染 | 广西金河... |
| 山西长治... | 1974-06-1... | 山西省长... | 苯胺 | 重度污染 | 某化工厂... |
| 松花江重... | 2005-11-13 | 松花江 | 苯、苯胺... | 重度污染 | 中石油吉... |
| 河北白洋... | 2006年2... | 白洋淀 | 水体污染... | 重度污染 | 大量未经... |
| 太湖水污... | 2007-5 | 江苏省无... | 蓝藻大量... | 重度污染 | 排污严重... |

应急预案

- 发生 I 级 事件 --- 立即采取隔离封堵等
- 发生 II 级 事件 --- 警戒疏散隔离抢救等
- 发生 III 级 事件 --- 交通管制医疗救护等

地图资料

北京市副中心水环境应急决策系统

③水环境污染应急决策支持系统



北京城市副中心水环境应急处置管理平台

11:37:23 星期二 2018-7-10

参数设置

| 参数名称 | 值 |
|-------------|-------------------------------|
| 污染物 | cod |
| 入流-排放(m3/s) | 30 |
| 出流-水位(m) | 0 |
| 排放口位置 | 8,76 |
| 排放浓度(kg/m3) | 1.0 |
| 排放量(m3/s) | 20 |
| 开始日期 | 2018-08-01 00:00:00 |
| 结束日期 | 2018-09-01 00:00:00 |
| 初始浓度(kg/m3) | 0.0 |
| 颜色条 | |
| 扩散模拟 | <input type="checkbox"/> 空间分析 |
| 计算完成 | 已完成 |

方案选择

方案一：建设多道紧邻坝的拦截技术

拦水工程

方案二：污水排出技术

排水工程

方案三：污水引出隔离技术

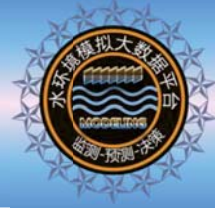
引水工程

辅助决策

播放 加速 减速 停止

北京市副中心水环境应急决策系统：水质预测

③水环境污染应急决策支持系统



DASA 中科水质 研发

北京市副中心水环境应急决策管理平台

11:38:32 星期二 2018-7-10

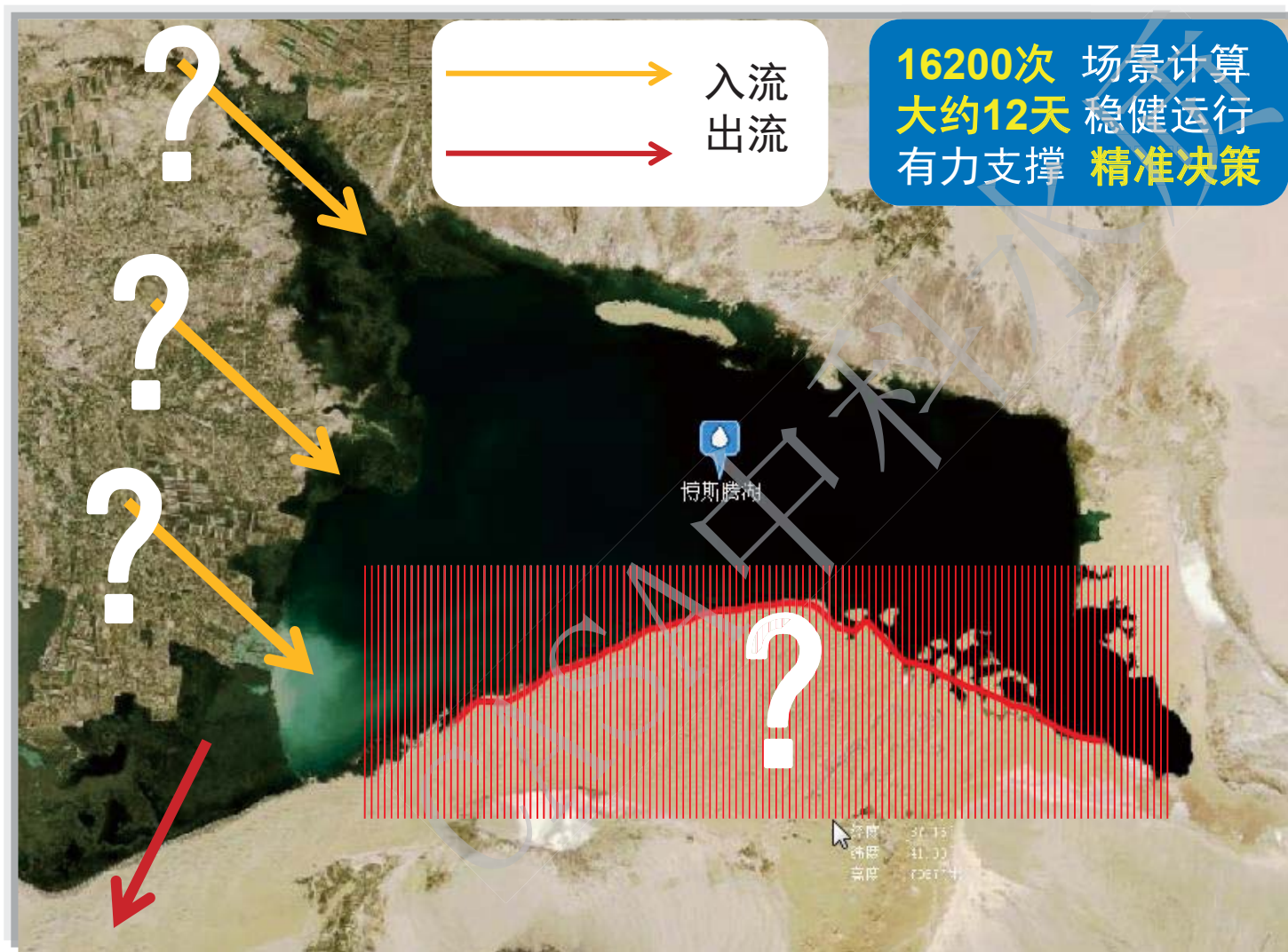
| 方案一：拦水工程 | 方案二：排水工程 | 方案三：引水工程 |
|------------------|------------------|------------------|
| 污染物: cod | 污染物: cod | 污染物: cod |
| 入流: 30 | 入流: 30 | 入流: 30 |
| 出流: 0 | 出流: 0 | 出流: 0 |
| 排放量: 20 | 排放量: 20 | 排放量: 20 |
| 排放浓度: 1.0 | 排放浓度: 1.0 | 排放浓度: 1.0 |
| 开始日期: 2018-08-01 | 开始日期: 2018-08-01 | 开始日期: 2018-08-01 |
| 结束日期: 2018-09-01 | 结束日期: 2018-09-01 | 结束日期: 2018-09-01 |

重新播放 播放 加速 减速 停止 一件加载

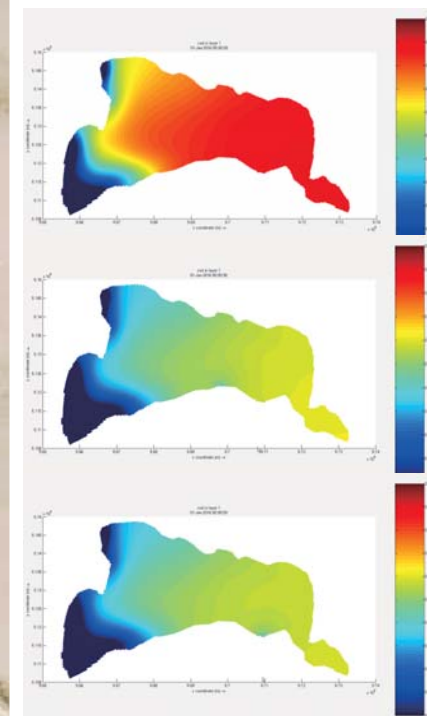
比选方案

北京市副中心水环境应急决策系统：多情景决策

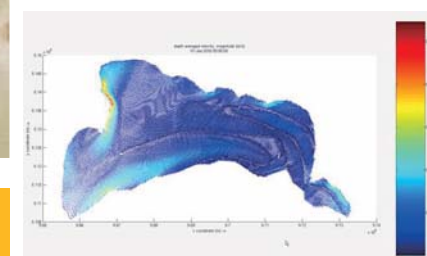
④环境工程影响评价-流场优化



COD



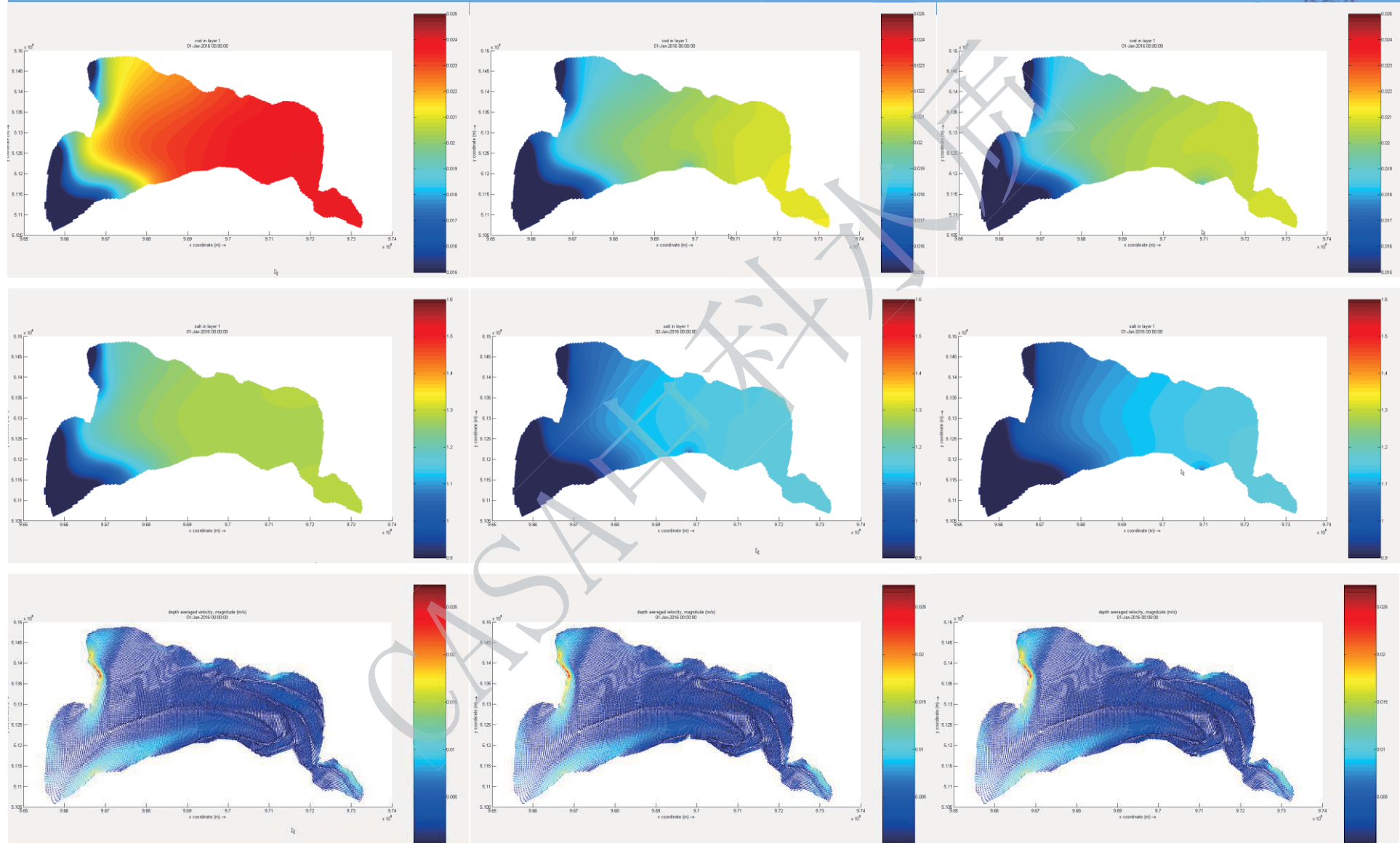
流场



多情景模拟:

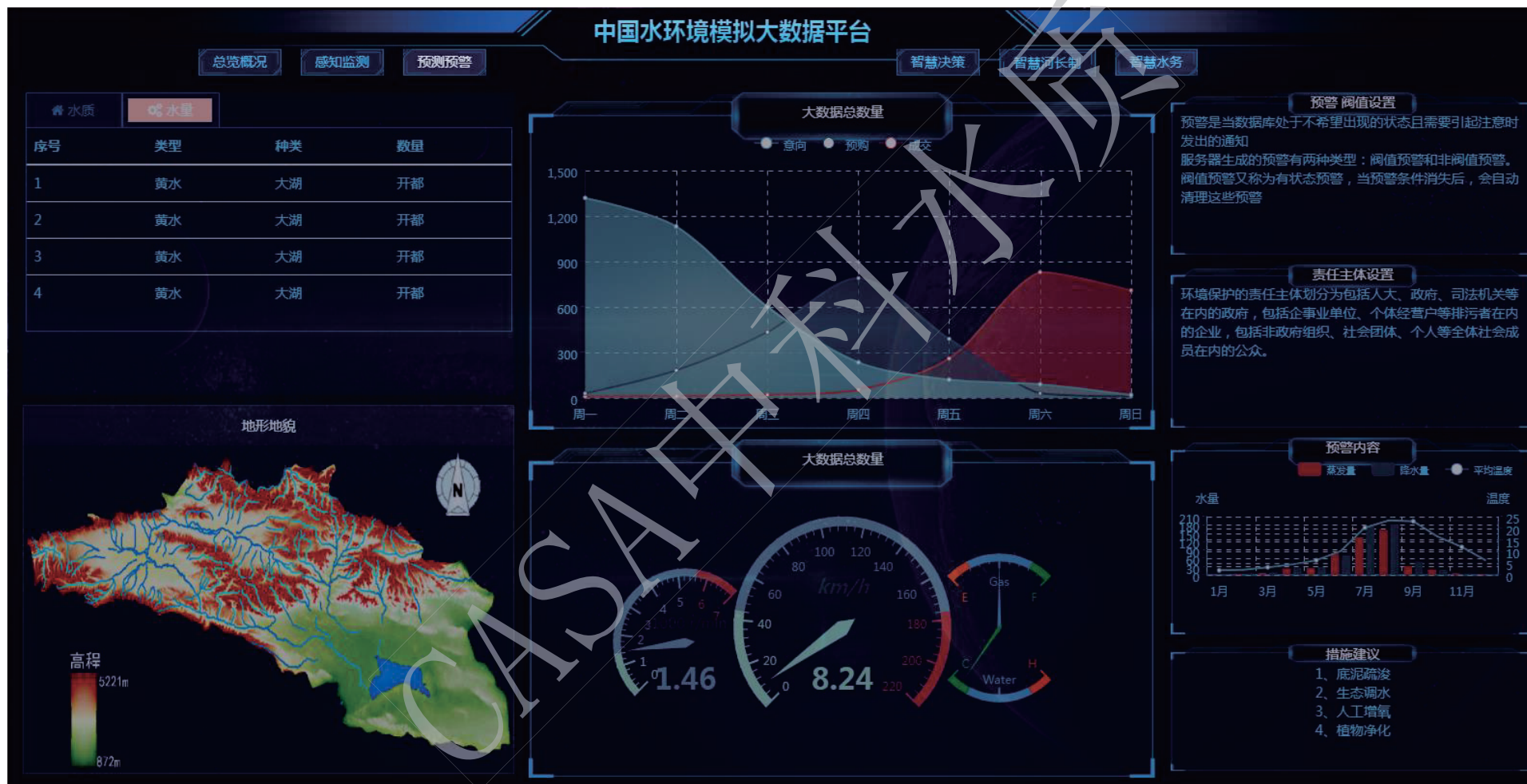
粗红线处-选择最佳排水位置-改善湖体循环能力

④环境工程影响评价-流场优化



不同情景下博斯腾水库COD、矿化度、流场演变过程

④环境工程影响评价-流场优化



感知监测

④环境工程影响评价-流场优化



中国水环境模拟大数据平台

总览概况

感知监测

预测预警

智慧决策

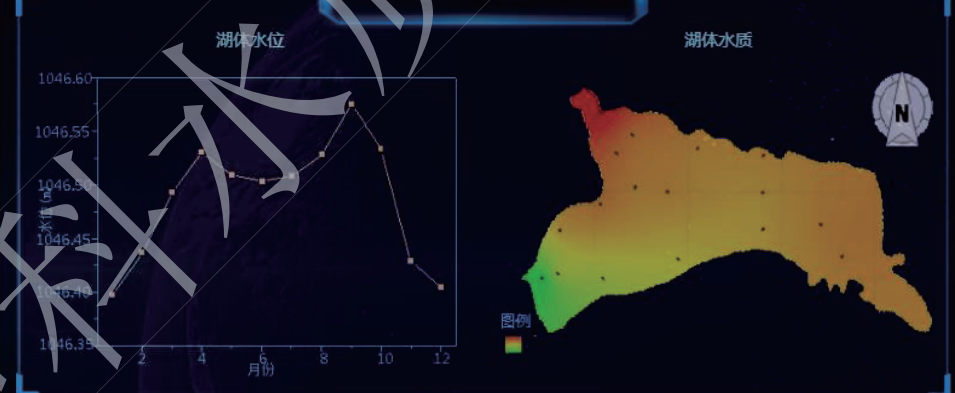
智慧河长制

智慧水务

流域水文

| 河流 | 水文站 | 年均径流量 /10 ⁸ m ³ | 月平均流量 (m ³ /s) | | | | | | | | | | | |
|-----|------|---------------------------------------|---------------------------|------|-----|-------|------|-------|-------|-------|-------|------|------|------|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 开都河 | 大山口 | 34.73 | 49.5 | 46.0 | 48. | 107.9 | 139. | 186.6 | 211.7 | 192.1 | 126.2 | 88.8 | 66.7 | 52.7 |
| | 焉耆站 | 25.94 | 45.6 | 50.5 | 52. | 76.2 | 88.6 | 126.9 | 156.7 | 144.1 | 74.6 | 53.0 | 62.6 | 52.3 |
| | 东支 | 17.04 | 36.0 | 35.7 | 31. | 32.3 | 46.1 | 86.3 | 111.2 | 110.7 | 51.6 | 17.2 | 43.0 | 44.5 |
| | 西支 | 7.60 | 13.9 | 13.0 | 21. | 24.7 | 26.7 | 37.2 | 42.1 | 34.3 | 20.4 | 20.0 | 18.9 | 15.0 |
| 黄水沟 | 黄水沟 | 2.91 | 3.7 | 3.5 | 3.5 | 3.8 | 7.4 | 17.3 | 26.0 | 20.6 | 10.2 | 6.0 | 4.8 | 4.0 |
| 清水河 | 克尔古提 | 1.19 | 1.6 | 1.6 | 1.4 | 1.3 | 1.6 | 4.8 | 10.3 | 9.4 | 5.7 | 3.6 | 2.8 | 2.0 |

湖体



水利工程

| 引水渠首工程 | 蓄水工程 | 输排水工程 | 提水工程 |
|--------------------|---------------------|--------------------------------------|--------------------------|
| 27座 | 8座 | 3647条 | 1221眼 |
| 扬水工程 2座 10800kw | 大小湖隔堤工程 全长12.2km | 黄水沟夏尔乌提分洪闸 泄流360m ³ /s | 孔雀河向塔里木河应急输水工程 实现应急输水 |

地表径流

| 开都河东支-水量 | 开都河东支-水质 | 黄水沟 |
|--|--|--|
| 开都河东支是博斯腾湖大湖区最主要的地表水补给源,2015年、2016年的入湖水量已突破20亿m ³ /年。 | 开都河东支水质状况情况良好,各项指标满足地表II类水标准。 | 黄水沟原本汇入博斯腾湖西北湖区,1964年黄水沟分洪闸建成后,黄水沟径流改道进入开都河。 |
| 农排渠 | 清水河 | 扬水站 |
| 博斯腾湖大湖区周边农排渠共计有19条,分布在西侧和西北侧,排渠水量受人为活动控制范围大、水量不稳定。 | 清水河仅在洪水季节才有极少量洪水汇入博斯腾湖,其它水量出山后均被用作灌溉用水,经地表下渗后补给博斯腾湖。 | 地表水出湖水量的唯一出口为博斯腾湖西南隅的扬水车站,各月调出水量取决于下游需水量。 |

业务报表

④环境工程影响评价-流场优化



中国水环境模拟大数据平台

总览概况 | 感知监测 | 预测预警 | 智慧决策 | 智慧河长制 | 智慧水务

博斯腾湖示意图

范围

地址：甘肃省 | 城市：新疆省

博斯腾湖： 地表径流 水质水文

地形地貌： 湖底地形 污染源 地下水

提交 重置 普通按钮

图

索引掌 视频监控 联盟 家访问 邮件营销

表

4屏 2屏 8屏

观测点

时间：时间

指标：指标

选择我

提交

演变一

条件一
条件二
条件三
条件四
条件五
条件六
条件七

演变二

条件一
条件二
条件三
条件四
条件五
条件六
条件七

演变三

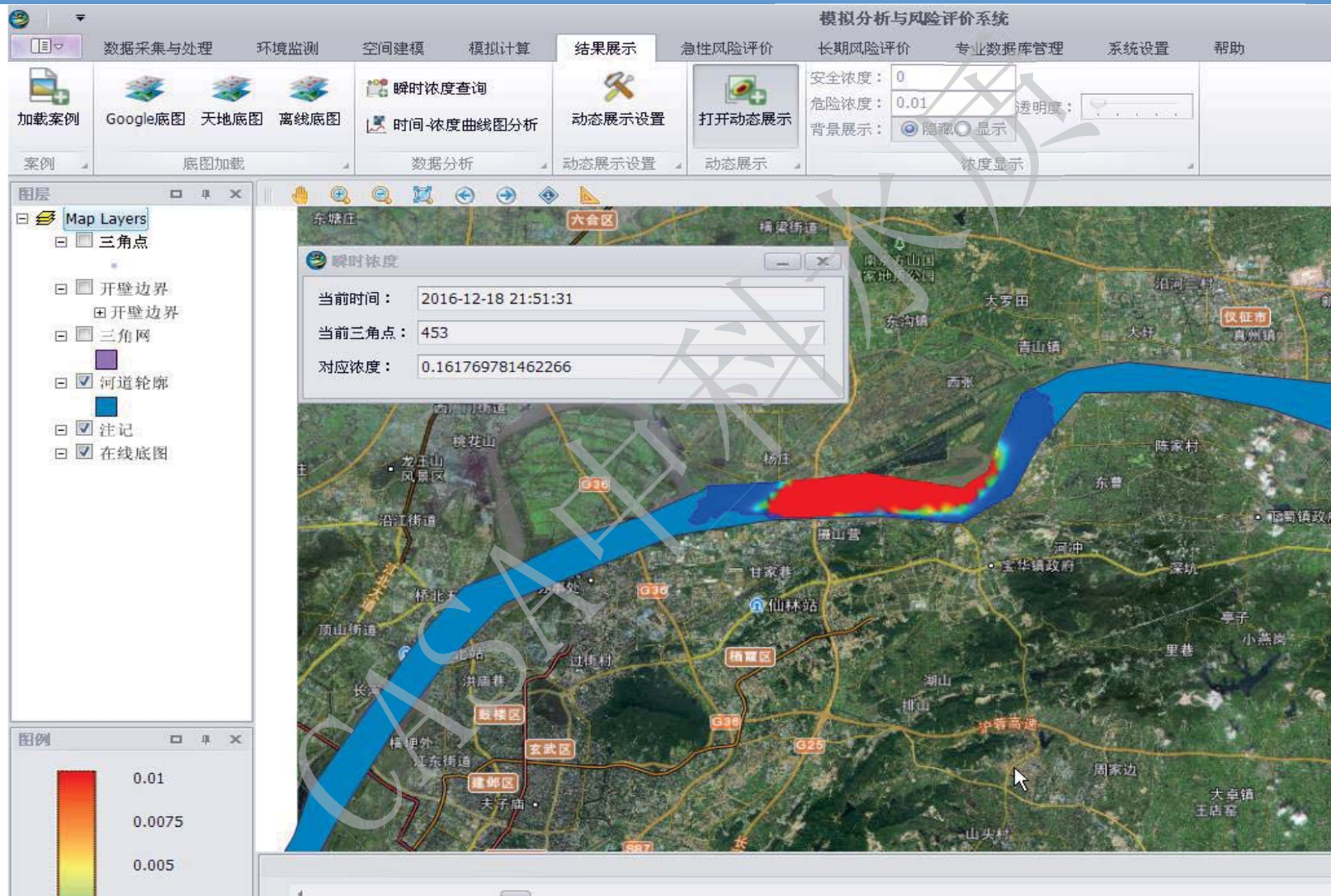
条件一
条件二
条件三
条件四
条件五
条件六
条件七

演变四

条件一
条件二
条件三
条件四
条件五
条件六
条件七

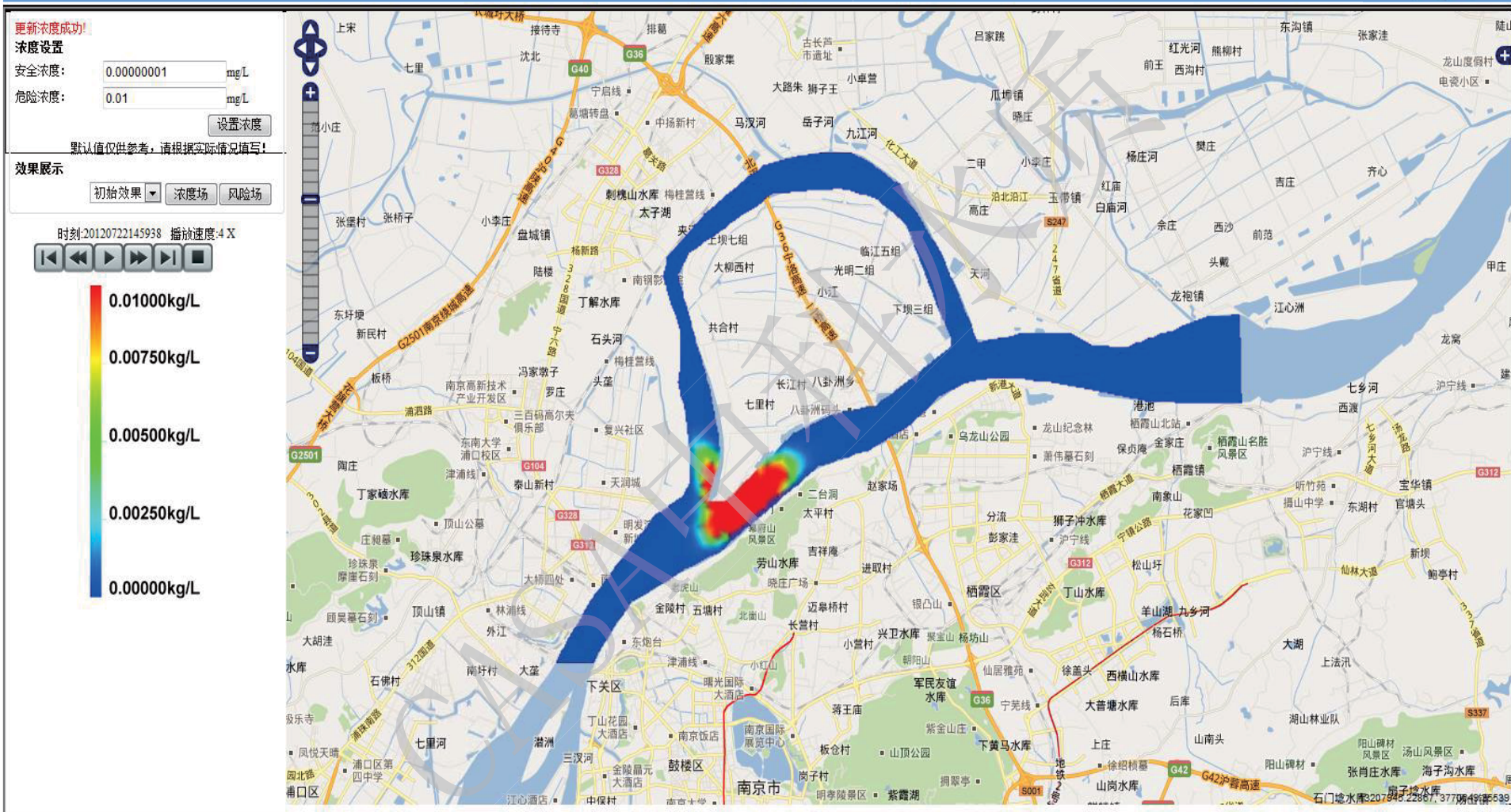
智慧决策

⑤早期案例：长江



长江水环境模拟分析与风险评价

⑤早期案例：长江



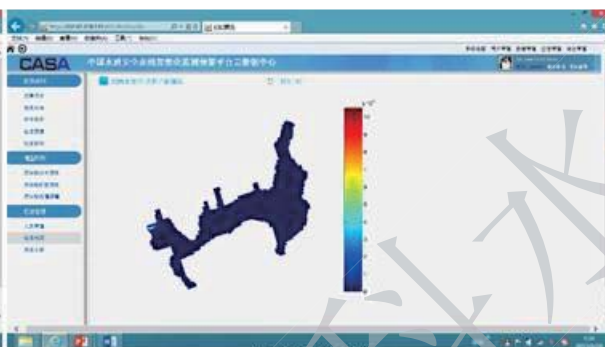
Copyright © 2011 中国科学院生态研究中心 All Right Reserved.

长江水环境模拟分析与风险评价

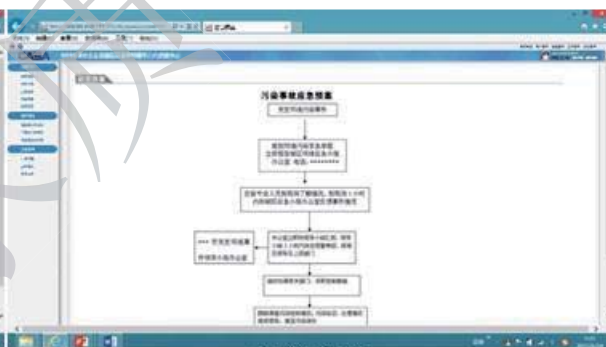
⑤早期案例：岗南水库



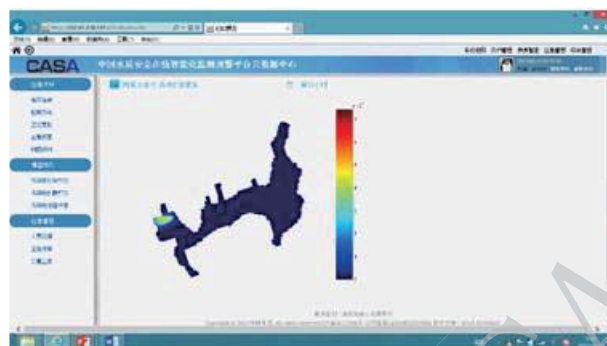
应急资料



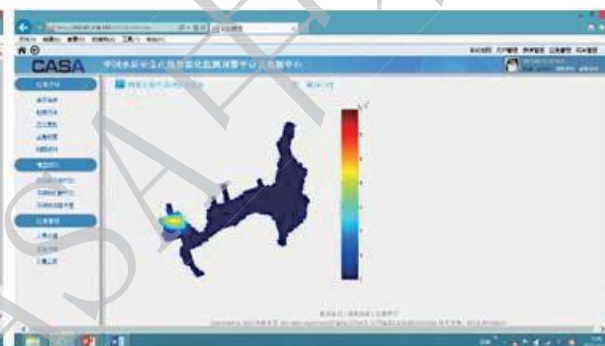
模型预测



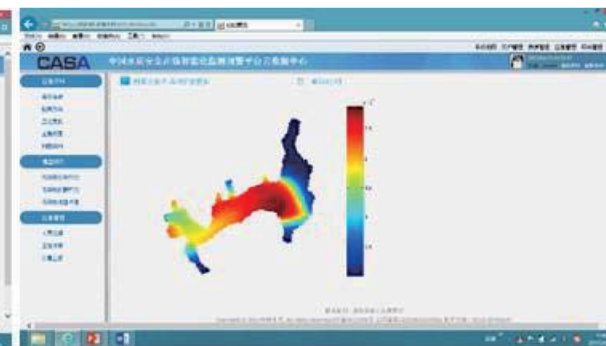
应急管理



污染扩散1小时



污染扩散24小时

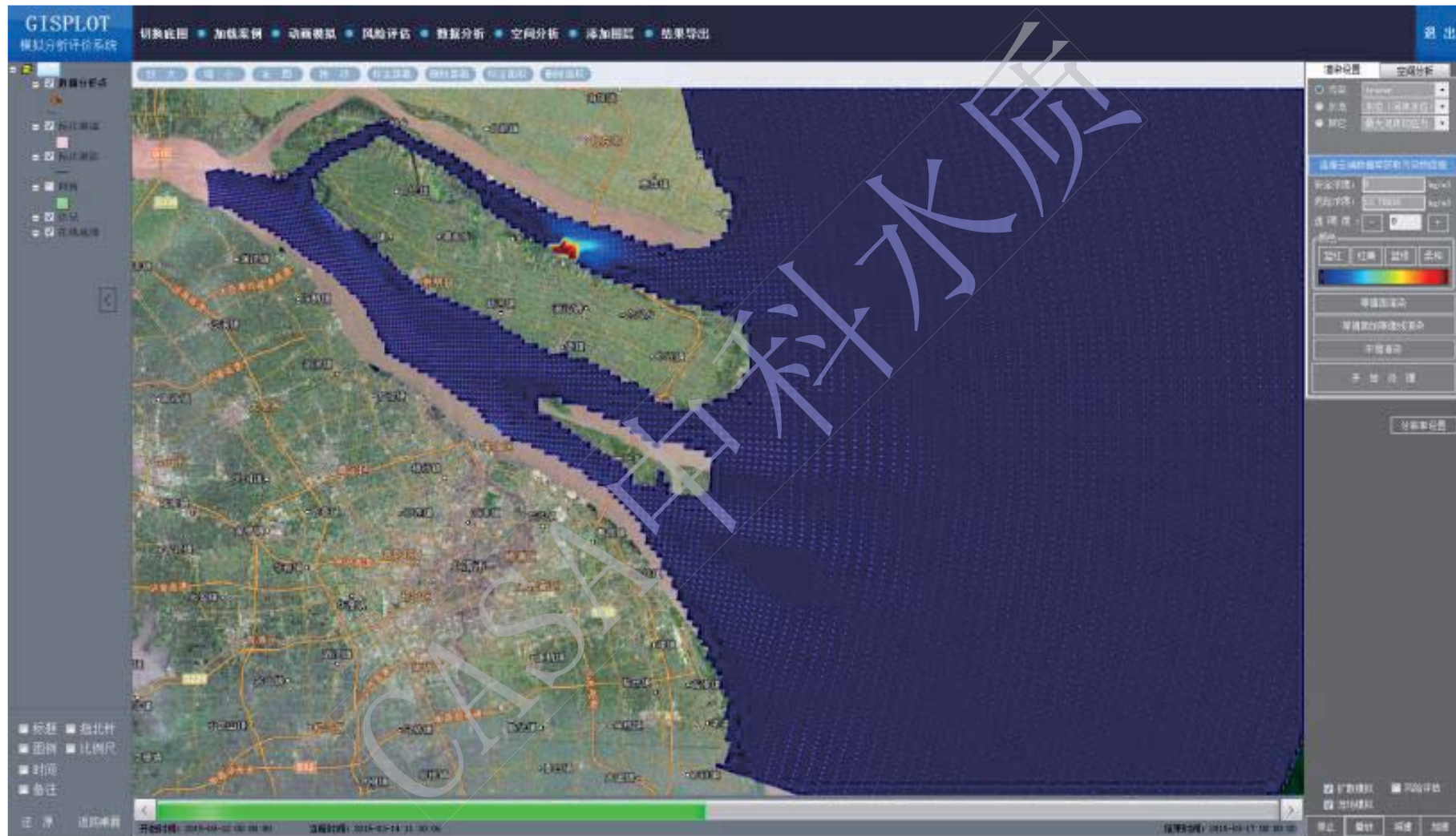


污染扩散320小时

水源地水质监测与生物预警系统



⑤早期案例：河口



长江口水环境模拟分析与风险评估

⑤早期案例：水土气应急集成系统



重大环境污染事件应急处理处置技术

登录帐号:admin 登录时间:2011-10-21 10:30

切换

- GIS功能
- 查询
- 多介质处置
- 设置

数据库管理

- GIS模块
- 危险化学品
- 日常预防管理
- GIS模块
- 知识库管理
- 污染源监管
- 化学品监管
- 敏感目标保护
- 事故应急处置
- GIS模块
- 风险评价
- 应急处置
- 处置方案

水体污染处置工具包

大气污染处置工具包

场地污染处置工具包

多介质处置工具包

源头堵截

扩散控制

就地处置

敏感目标

安全防护

应急调度

GIS

系统首页

事故应急处置

数据库管理

日常预防管理

系统管理

帮助

模型库管理

技术库管理

知识库管理

方案库管理

系统库管理

事故接警

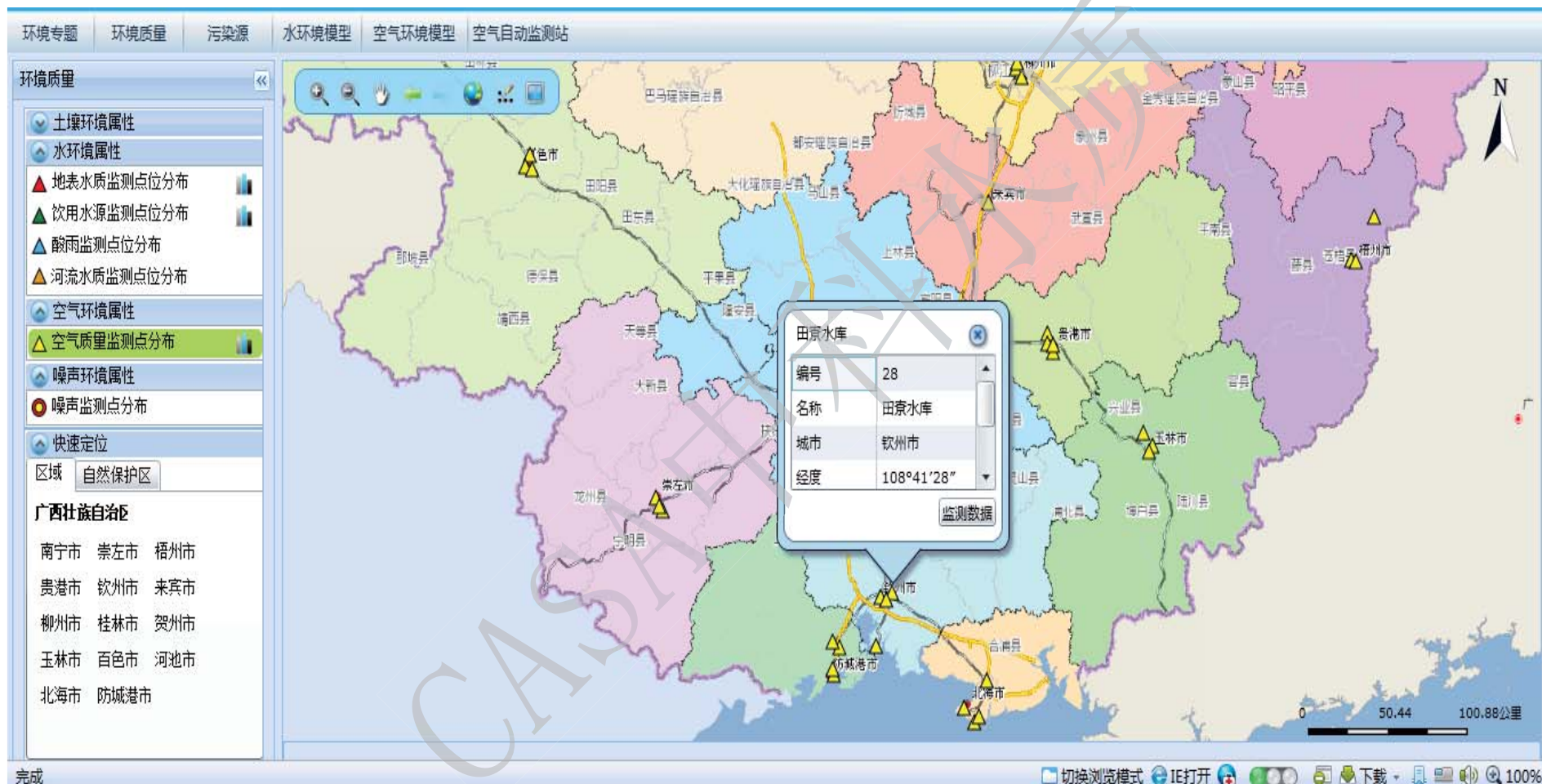
扩散模拟

风险评价

方案推荐

重大环境污染事件应急处理处置

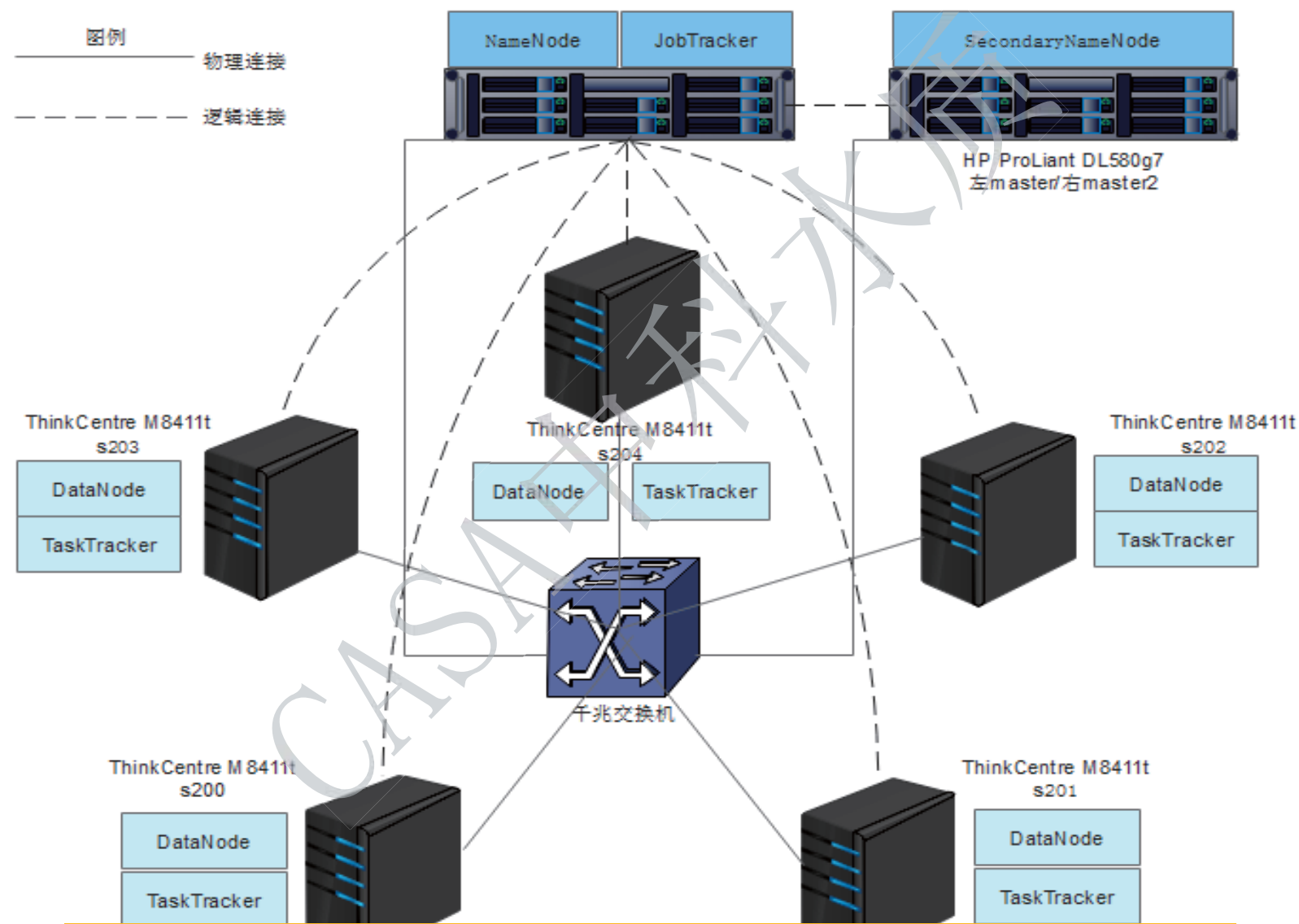
⑤早期案例：环保局系统集成



水环境模型应用—广西北部湾

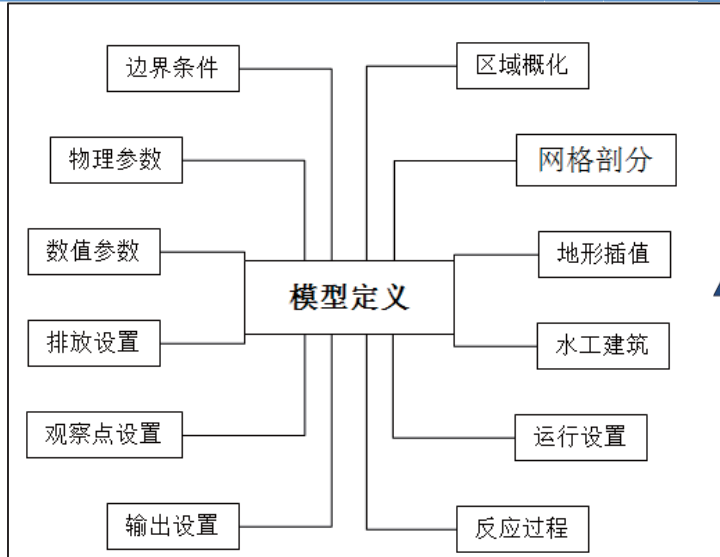


⑥计算平台-大数据集群

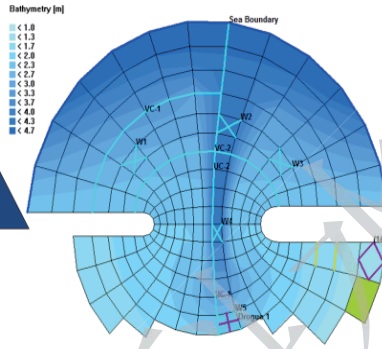


集群服务器 9台 120核cpu 352G内存 18T磁盘容量 千兆交换机

⑥计算平台-大数据集群



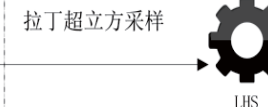
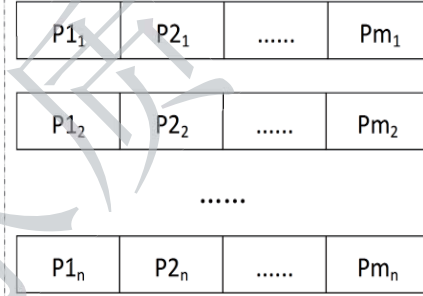
①模型定义（空间建模）



②参数提取



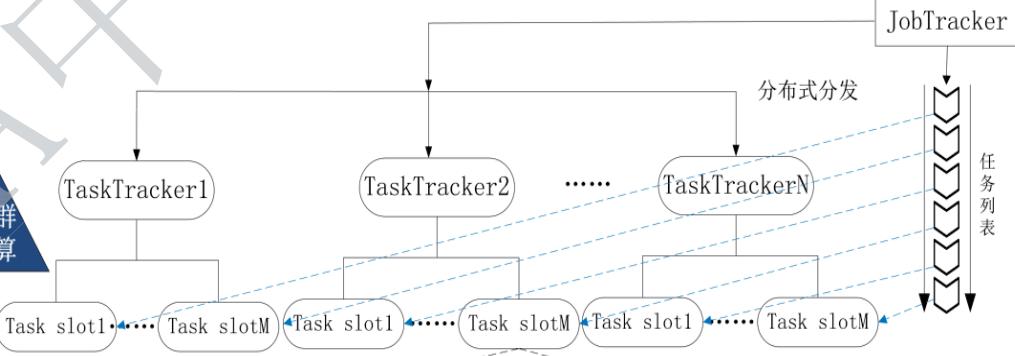
③参数集定义（情景定义）



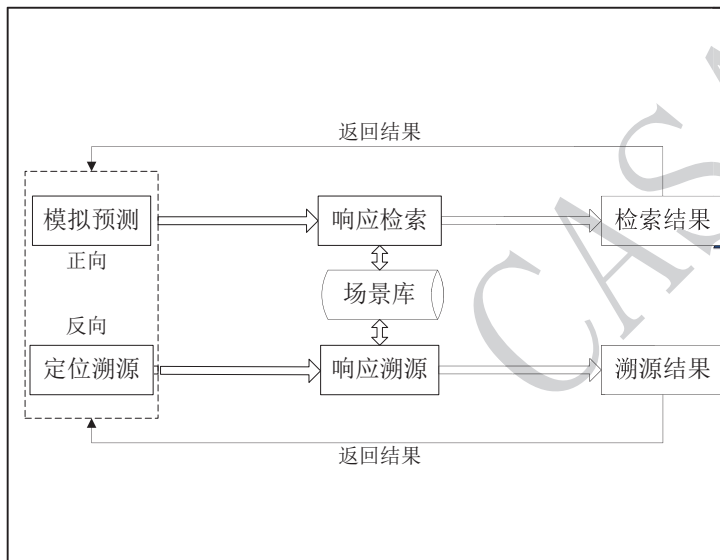
提交

参数集文件列表

④计算任务分发



⑤计算任务执行



集群计算



⑥ 计算平台-大数据集群性能

分布式计算

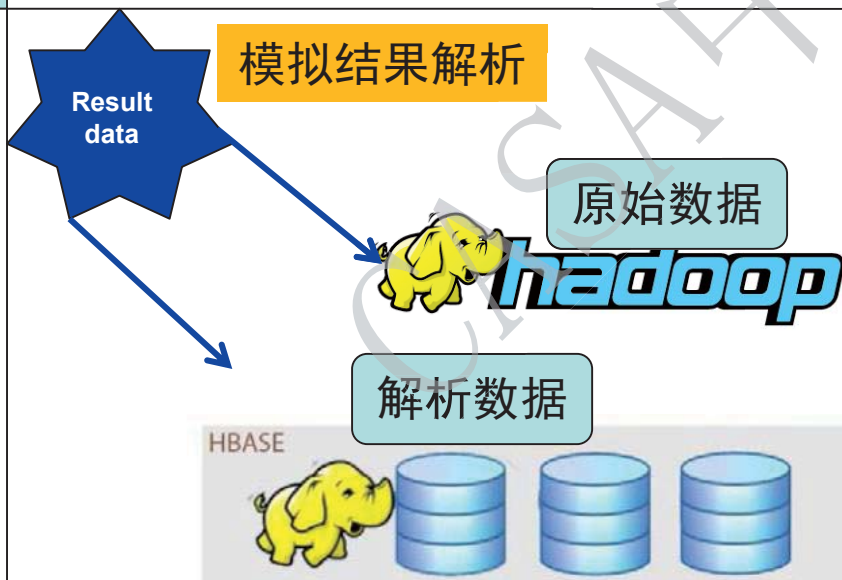


- 以南京八卦洲区域Delft3d模型为例
- 创建**10000**次模拟情景进行运算
- 耗时**187小时**，约8.7天成功计算完成
- 优点：过程简单；可扩展；可容错

Job Setup: Successful
 Status: Succeeded
 Started at: Thu Oct 19 11:28:30 CST 2017
 Finished at: Fri Oct 27 06:23:42 CST 2017
 Finished in: 186hrs, 55mins, 11sec
 Job Cleanup: Successful

| Kind | % Complete | Num Tasks | Pending | Running | Complete | Killed | Failed/Killed Task Attempts |
|--------|------------|-----------|---------|---------|----------|--------|-----------------------------|
| map | 100.00% | 10000 | 0 | 0 | 10000 | 0 | 0 / 0 |
| reduce | 100.00% | 0 | 0 | 0 | 0 | 0 | 0 / 0 |

海量存储



- 原始结果存入Hadoop HDFS文件系统
- 解析结果存入Hbase列式数据库
- 耗时**165小时**，约6.8天写入**73.5亿条**
- 优点：高速随机读写；大容量

| | | | |
|------------------------------------|-----------------|---|-----------------|
| ROWS | 7,349,996,740 | 0 | 7,349,996,740 |
| HDFS_BYTES_READ | 201,060 | 0 | 201,060 |
| FILE_BYTES_WRITTEN | 153,977,137 | 0 | 153,977,137 |
| Map input records | 7,349,996,740 | 0 | 7,349,996,740 |
| Physical memory (bytes) snapshot | 332,178,763,776 | 0 | 332,178,763,776 |
| Spilled Records | 0 | 0 | 0 |
| Total committed heap usage (bytes) | 858,585,563,136 | 0 | 858,585,563,136 |

⑥计算平台-大数据集群性能



A: 单作业VS多作业

负载均衡测试

| 序号 | 场景数量 | 作业数 | 节点数 | 任务槽/节点 | 任务数/作业 |
|----|------|-----|-----|--------|--------|
| 1 | 100 | 1 | 1 | 1 | 100 |
| | 200 | 1 | 2 | 1 | 200 |
| | 300 | 1 | 3 | 1 | 300 |
| | 400 | 1 | 4 | 1 | 400 |
| | 500 | 1 | 5 | 1 | 500 |
| 2 | 100 | 100 | 1 | 1 | 1 |
| | 200 | 100 | 2 | 1 | 2 |
| | 300 | 100 | 3 | 1 | 3 |
| | 400 | 100 | 4 | 1 | 4 |
| | 500 | 100 | 5 | 1 | 5 |

B: 动态增/删计算节点

可扩展测试

C: 主机硬件故障

容错性测试

D: 任务槽个数对计算性能影响

E: JVM开启对计算性能影响

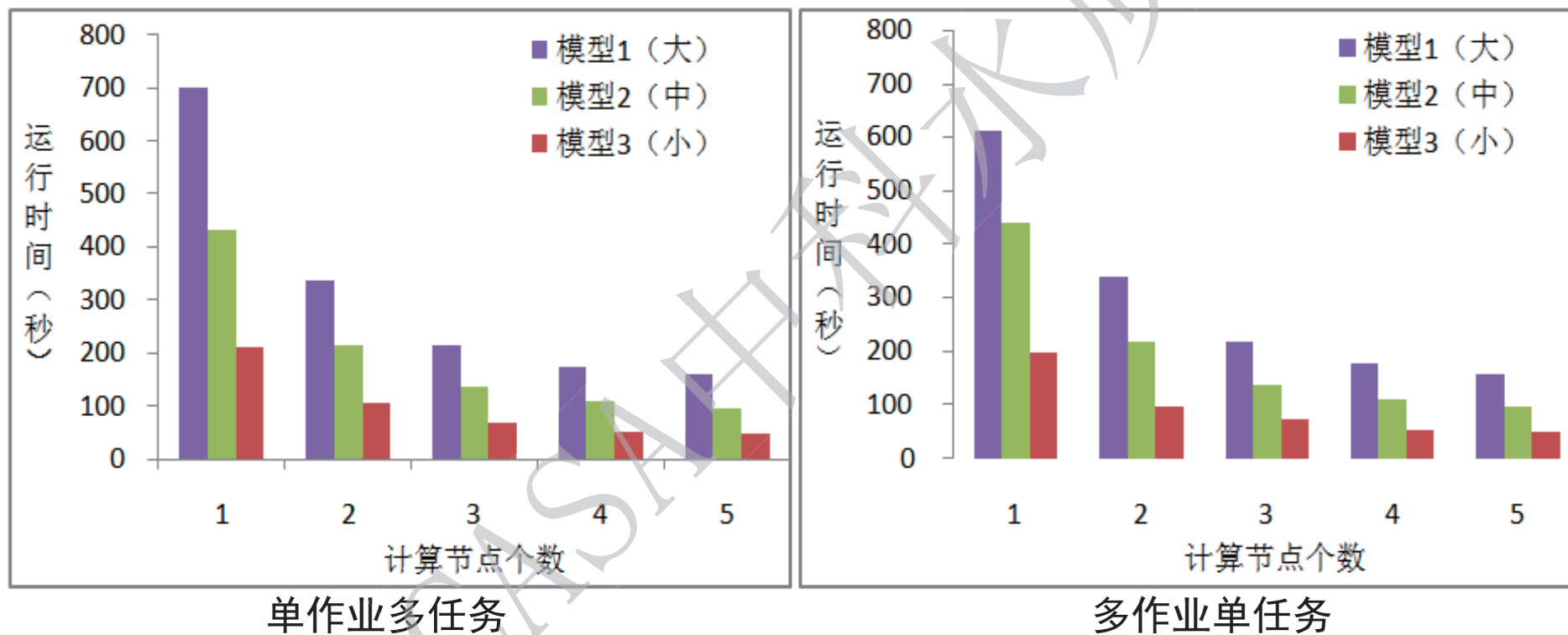
$$\text{加速比} = \frac{n * \text{单模型运行时间}}{\text{并行时间}}$$

$$\text{效率} = \frac{\text{加速比}}{\text{计算节点个数}}$$

⑥计算平台-大数据集群性能



运行时间



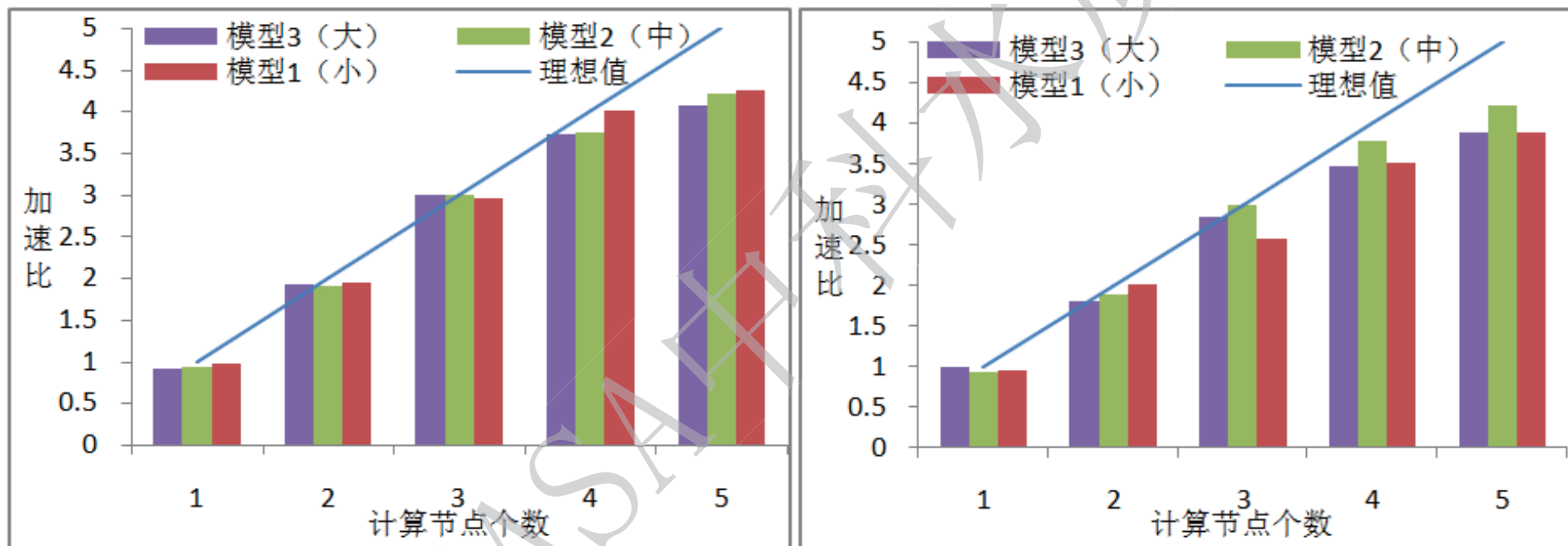
运行时间同时减少

随着计算节点增加，单作业多任务运行速度快于多作业单任务

⑥计算平台-大数据集群性能



加速比



单作业多任务

多作业单任务

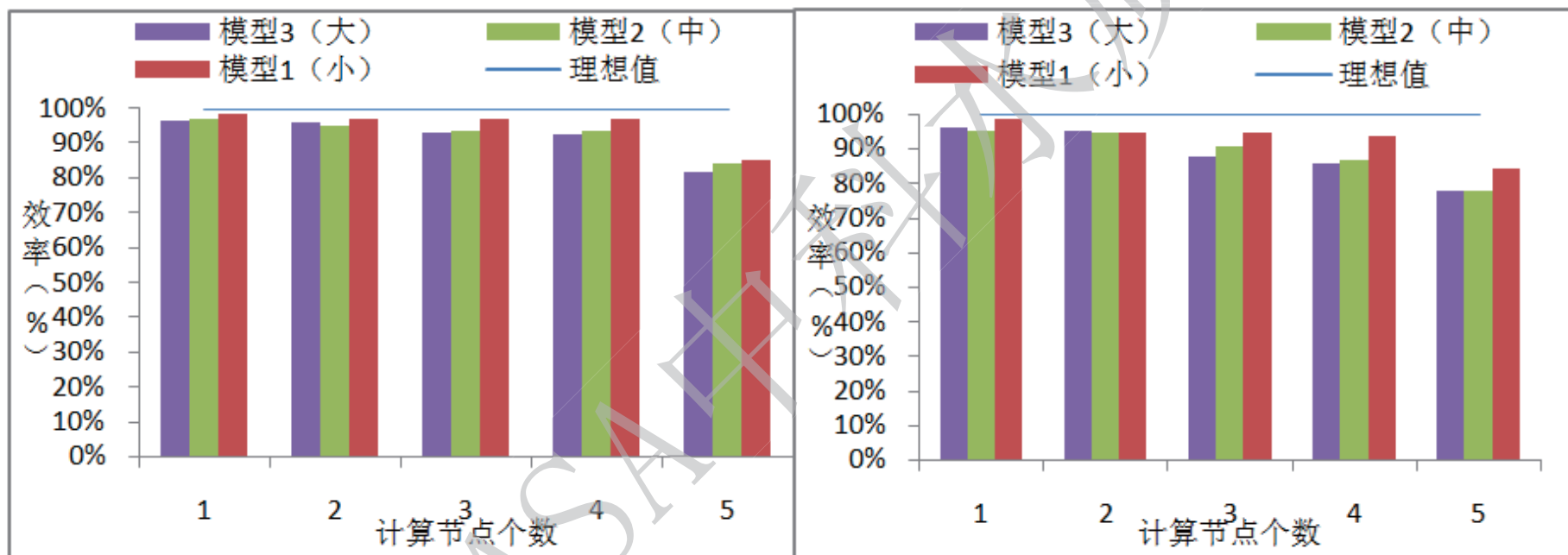
加速比同时提高

随着计算节点增加, 单作业多任务加速比大于多作业单任务

⑥计算平台-大数据集群性能



计算效率



单作业多任务

多作业单任务

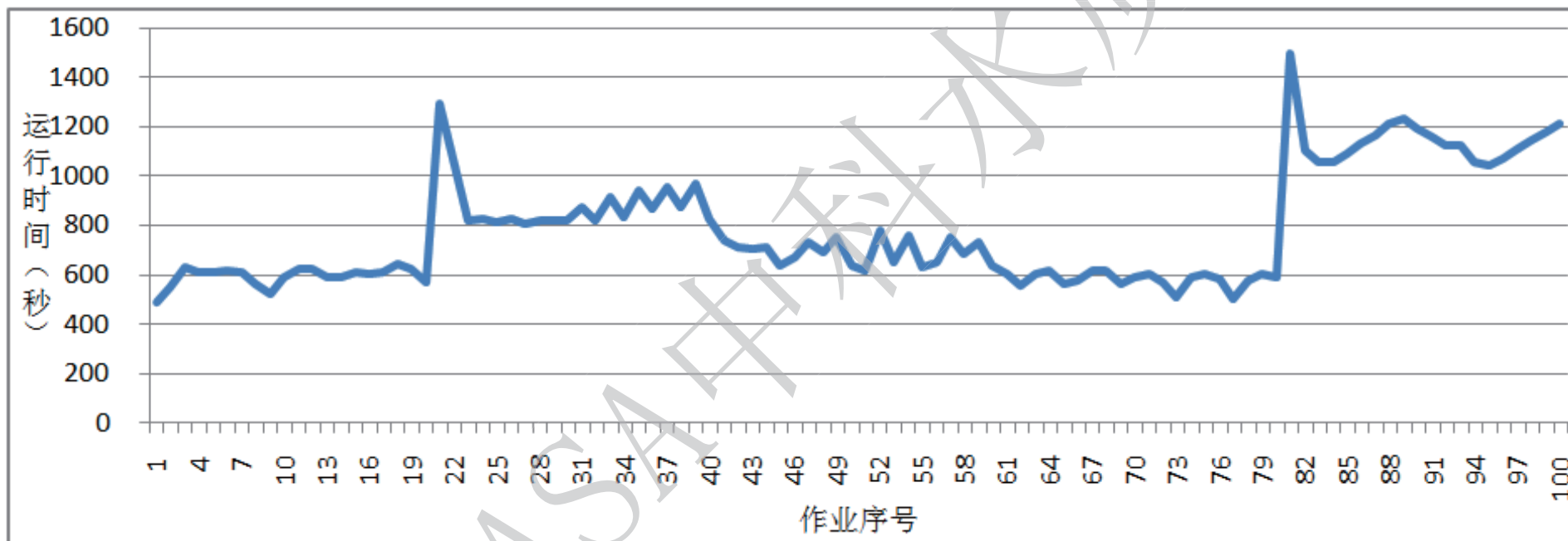
计算效率同时缓慢降低

随着计算节点增加，单作业多任务计算效率降低幅度小于多作业单任务

⑥计算平台-大数据集群性能



增/删计算节点



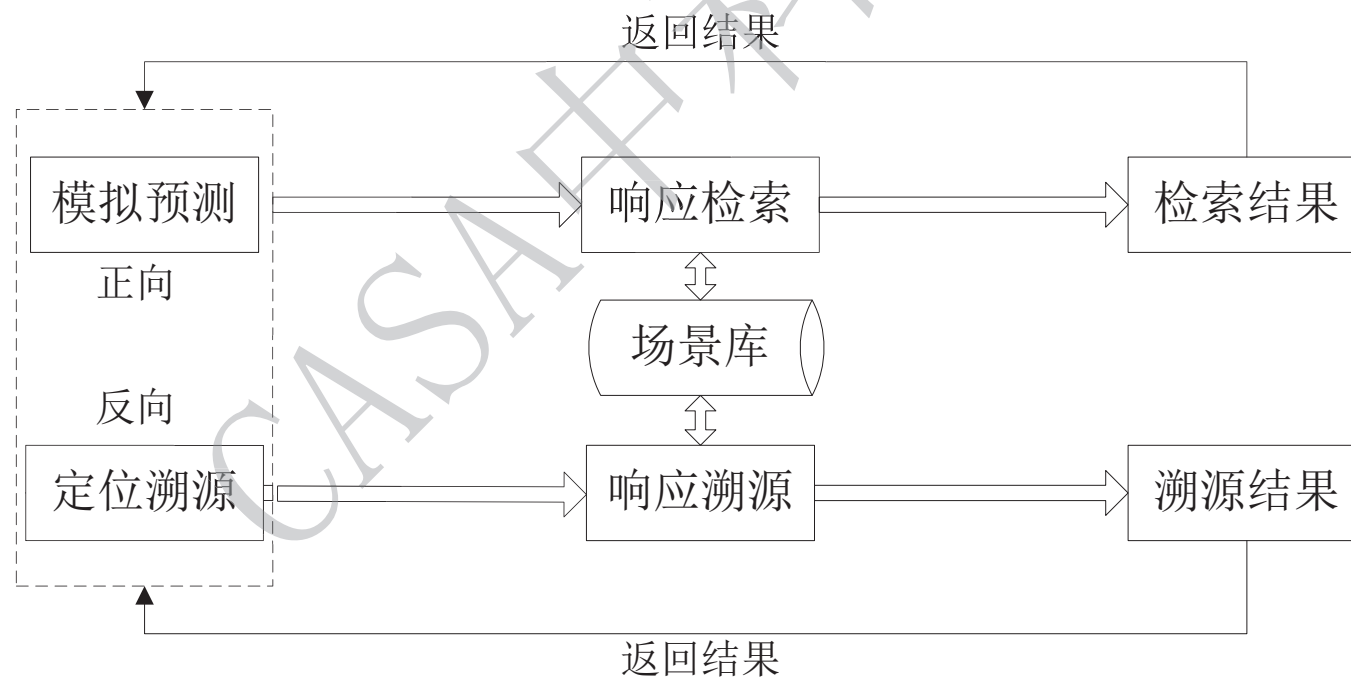
| 作业编号 | 运行节点 | 关闭节点 |
|--------|------|------|
| 1-19 | 5节点 | 0 |
| 20-39 | 3节点 | 2 |
| 40-59 | 4节点 | 1 |
| 60-79 | 5节点 | 0 |
| 80-100 | 2节点 | 3 |

证明了集群的可扩展性和故障修复

⑥计算平台-大数据集群性能



表明集群可以显著加快运算过程，提供良好的横向扩展、负载均衡、故障修复和容错的模型执行环境，因而是解决水环境模拟大规模计算需求问题的一种理想方法。





⑥ 计算平台-大数据集群性能

分布式计算

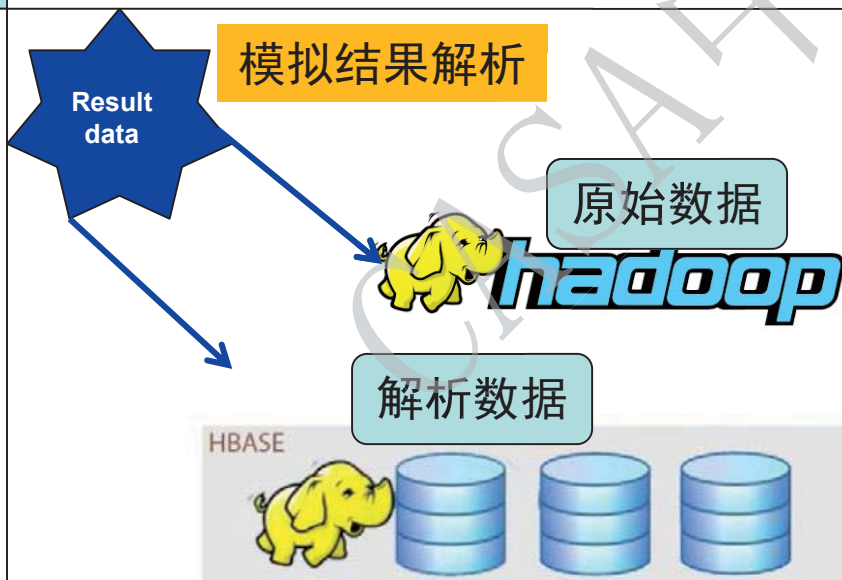


- 以南京八卦洲区域Delft3d模型为例
- 创建**10000**次模拟情景进行运算
- 耗时**187小时**，约8.7天成功计算完成
- 优点：过程**简单**；可**扩展**；可**容错**

Job Setup: **Successful**
 Status: Succeeded
 Started at: Thu Oct 19 11:28:30 CST 2017
 Finished at: Fri Oct 27 06:23:42 CST 2017
 Finished in: 186hrs, 55mins, 11sec
 Job Cleanup: **Successful**

| Kind | % Complete | Num Tasks | Pending | Running | Complete | Killed | Failed/Killed Task Attempts |
|--------|------------|-----------|---------|---------|----------|--------|-----------------------------|
| map | 100.00% | 10000 | 0 | 0 | 10000 | 0 | 0 / 0 |
| reduce | 100.00% | 0 | 0 | 0 | 0 | 0 | 0 / 0 |

海量存储

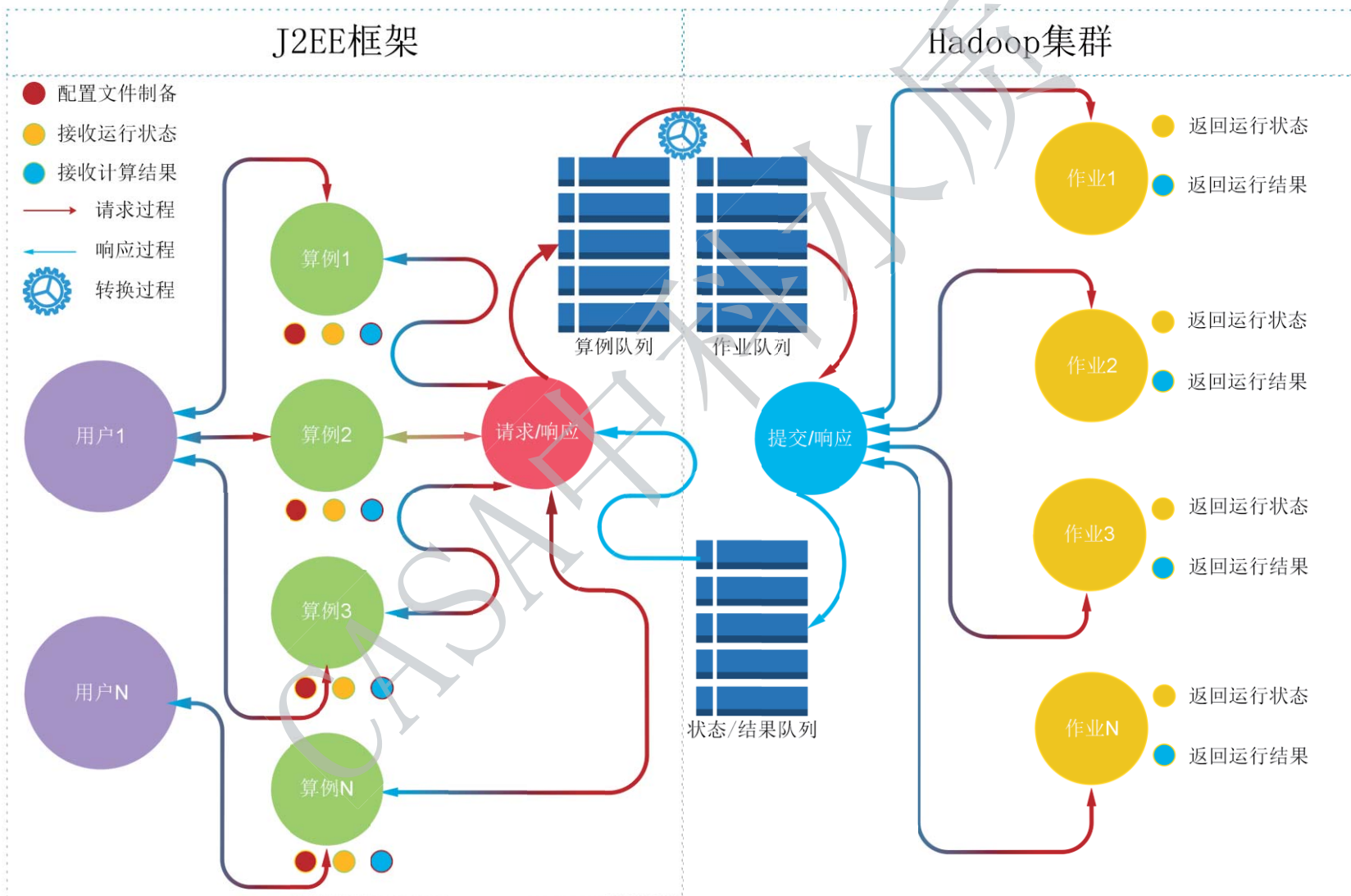


- 原始结果存入Hadoop HDFS文件系统
- 解析结果存入Hbase列式数据库
- 耗时**165小时**，约6.8天写入**73.5亿条**
- 优点：高速随机读写；大容量

| | | | |
|------------------------------------|-----------------|---|-----------------|
| ROWS | 7,349,996,740 | 0 | 7,349,996,740 |
| HDFS_BYTES_READ | 201,060 | 0 | 201,060 |
| FILE_BYTES_WRITTEN | 153,977,137 | 0 | 153,977,137 |
| Map input records | 7,349,996,740 | 0 | 7,349,996,740 |
| Physical memory (bytes) snapshot | 332,178,763,776 | 0 | 332,178,763,776 |
| Spilled Records | 0 | 0 | 0 |
| Total committed heap usage (bytes) | 858,585,563,136 | 0 | 858,585,563,136 |



⑥ 计算平台-服务交付





专注于水环境模拟

马金锋

中国科学院生态环境研究中心

联系方式-手机/微信:13811789757