



大数据技术及其防汛应用

河海大学 朱跃龙

2018年4月14日 深圳

版权说明

本课件为2018（第六届）中国水利信息化技术论坛专家发言材料，仅供参会人员内部交流使用，禁止外传及作为他用！

论坛主办单位：河海大学

论坛支持单位：水利部信息中心、广东省水利厅

获取更多信息可关注微信公众号：水务论坛在线（ID：swltzx）

中国水利信息化技术论坛会务组

电话：010-6320 3233/3104

网址：www.sinowbs.com

地址：北京市西城区白广路北口水利部综合楼732

主要内容

1

大 数 据 背 景

2

大 数 据 技 术

3

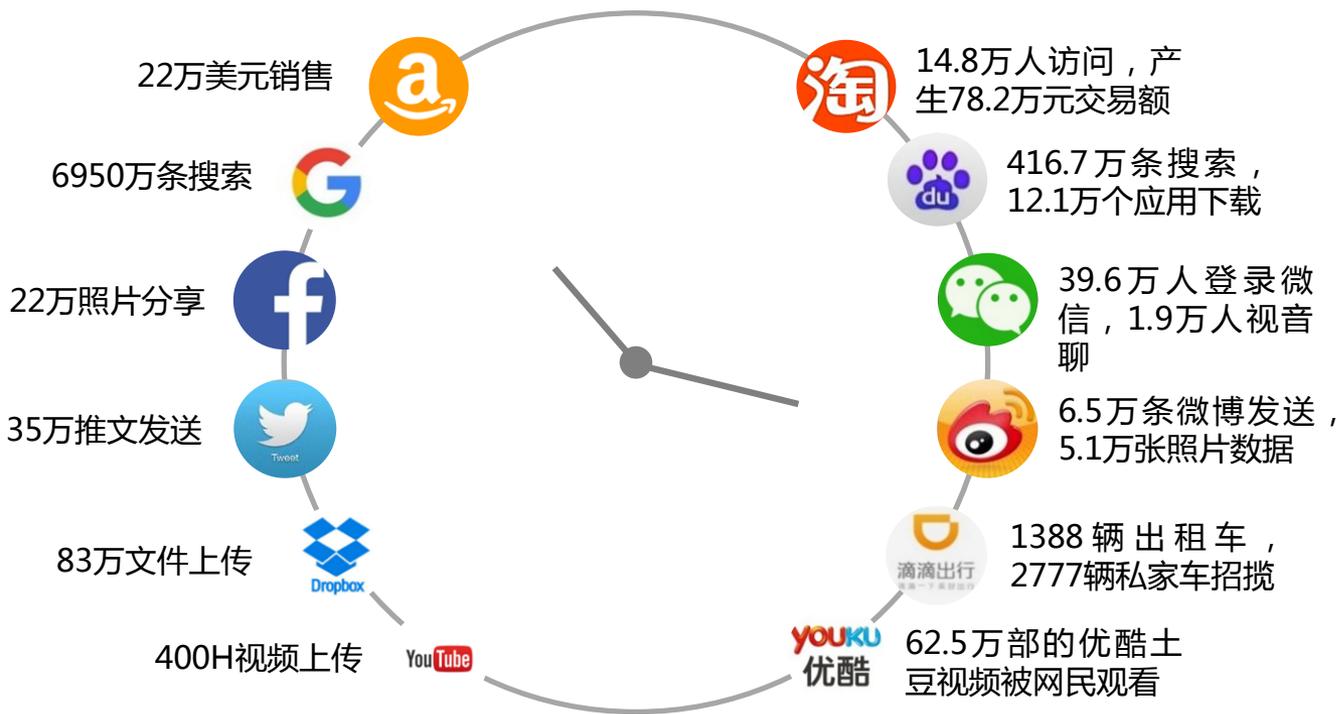
大 数 据 防 汛 应 用

第一部分

大 数 据 背 景

大数据背景

一分钟互联网发生了什么...



大数据背景

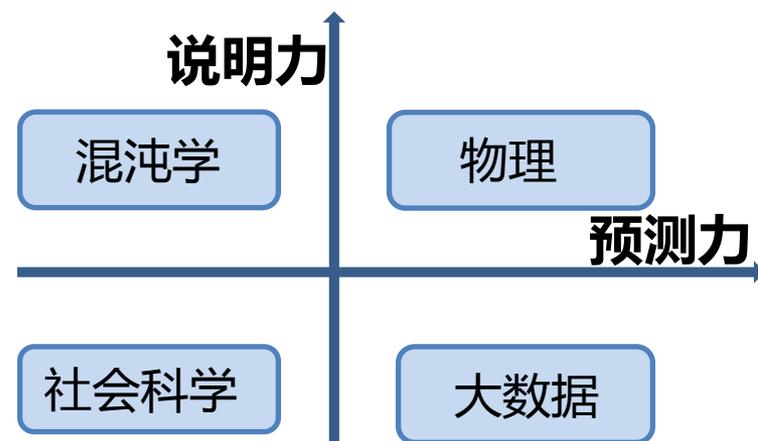
- 大数据（Big Data）最早由全球知名咨询公司麦肯锡提出
- 2008年9月，《科学》杂志发表文章“Big Data：Science in the Petabyte Era”，“大数据”这个词开始广泛传播。
- 2012年美国出台《大数据的研究和发展计划》
- 2013年英国出台《英国数据能力发展战略规划》
- 2013年澳大利亚发布了《公共服务大数据战略》
- 2013年日本制定《创建最尖端IT国家宣言》
- 2015年8月19日国务院发布《关于促进大数据发展的行动纲要》
- 2017年7月8日国务院发布《新一代人工智能发展规划》

大数据背景

计算机图灵奖获得者Jim Gray：提出了一种新的数据探索型研究方式，称之为科学研究的“第4种范式”（The Fourth Paradigm）

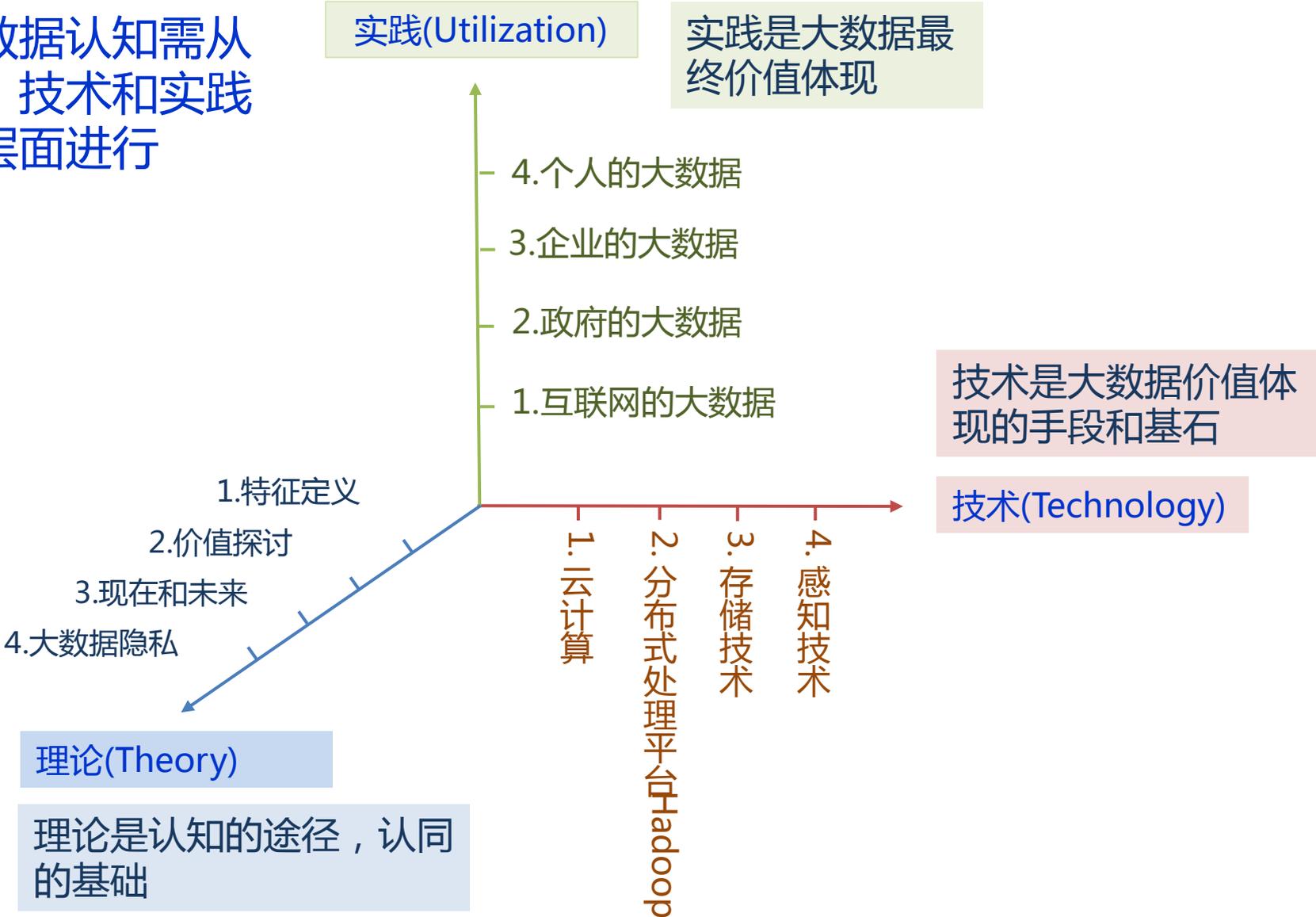


- 科学的两个主要目标：解释（说明）与预测。
- 大数据方法解释能力弱（Google翻译不懂语法）
- 大数据方法预测能力强（特别适合做情报分析）
- 大数据技术在智能性问题的出色结果（如IBM沃森、Google的AlphaGo）令人“难以理解”，正是大数据方法的特征。



大数据背景

对大数据认知需从理论、技术和实践三个层面进行



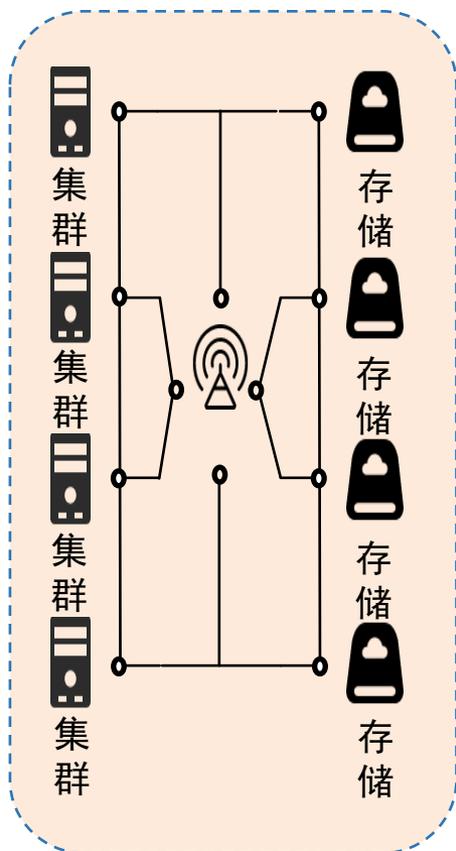
第二部分

大 数 据 技 术

大数据技术

大数据系统的层次结构 / Layered Architecture

基础设施层



计算层



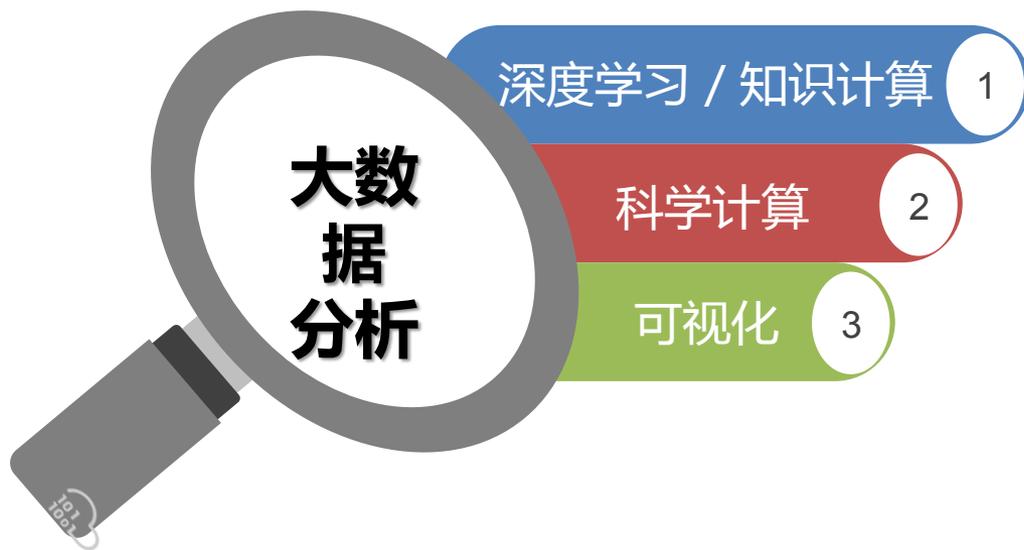
应用层



大数据技术

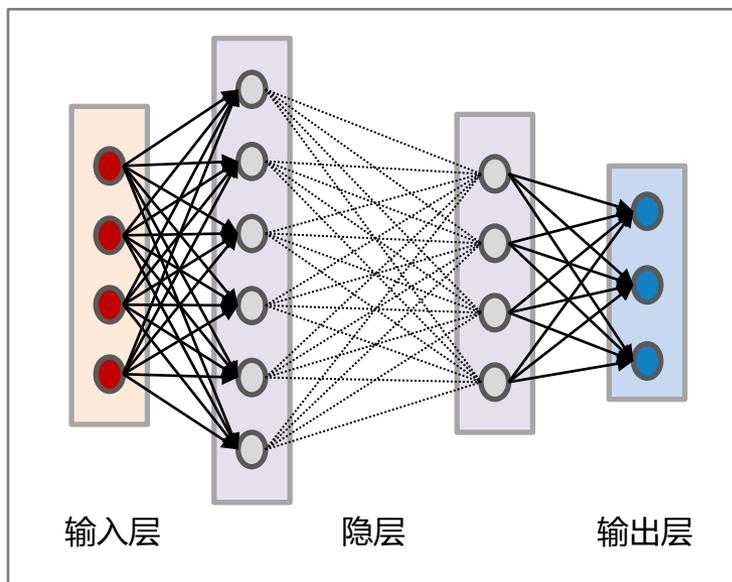
大数据分析 / Analysis

要挖掘大数据中所蕴涵演变规律和有价值的信息，就必然要对大数据进行内容上的分析与计算，即大数据分析。



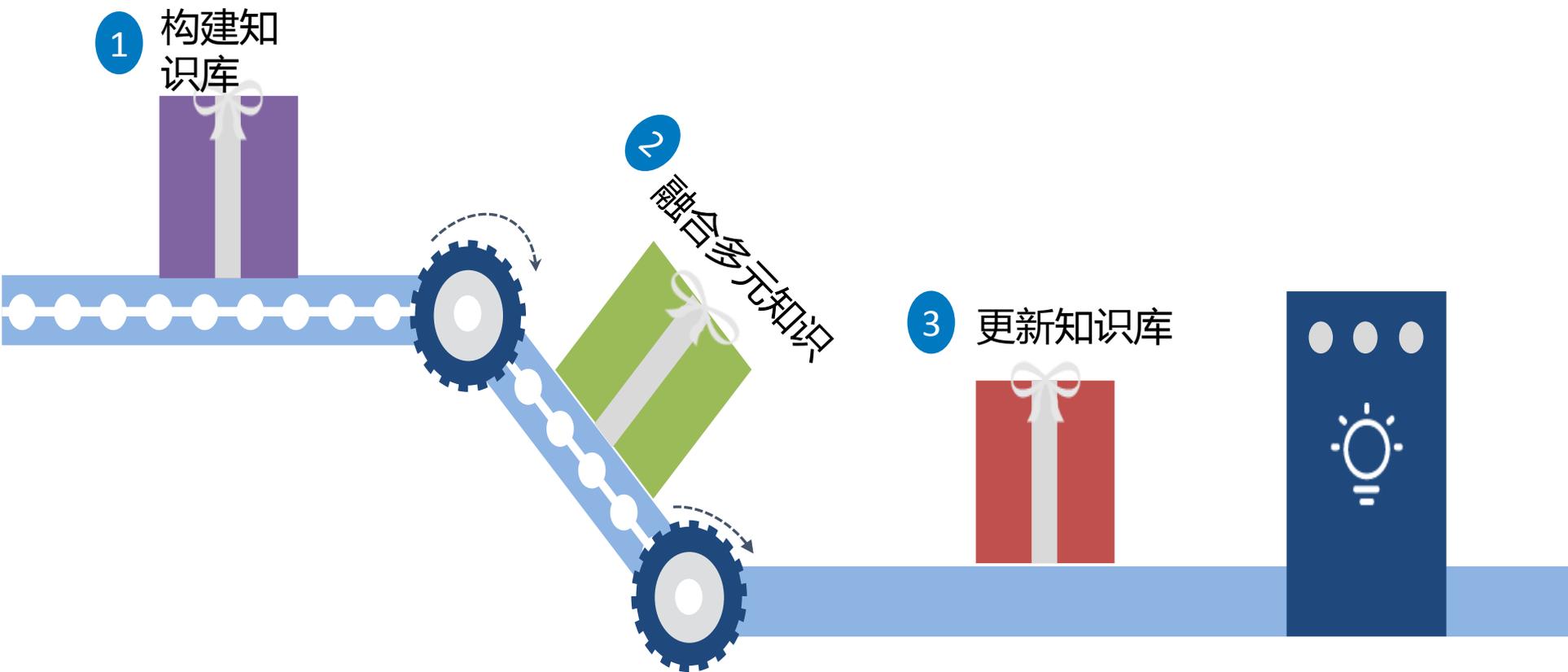
深度学习 / Deep learning

深度学习是一种基于机器学习、数据挖掘技术以及神经网络理论，分析大数据潜在价值的过程，利用层次化的架构学习出对象在不同层次上的表达，从而解决复杂抽象的问题



知识计算 / Knowledge Computing

基于大数据的知识计算是大数据分析的基础。目前，世界各国各个组织建立的知识库多达 50 余种，相关的应用系统更是达到了上百种。



科学计算 / Scientific Computing

由于大数据存在复杂、高维、多变等特性，如何从真实、凌乱、无模式和复杂的大数据中挖掘出人类感兴趣的知识，迫切需要更深刻的机器学习理论进行指导，使用系统科学、人工智能、数据挖掘等科学计算，从而解决政治、经济、文化等领域复杂性社会问题的一种理论和方法论体系

大数据聚类

聚类所面临的不仅是数据量越来越大的问题，更重要的还是数据的高维问题。

大数据关联分析

解决大数据的关联分析主要有两种途径：
并行和增量。

大数据并行算法

把传统机器学习算法运用到大数据环境中，一个典型策略是对现有学习算法并行化。



大数据分治策略

数据分治与并行处理策略是大数据处理的基本策略。

大数据特征选择

大数据存在复杂、高维、多变等特性，采用降维和特征选择技术以降低大数据处理难度。

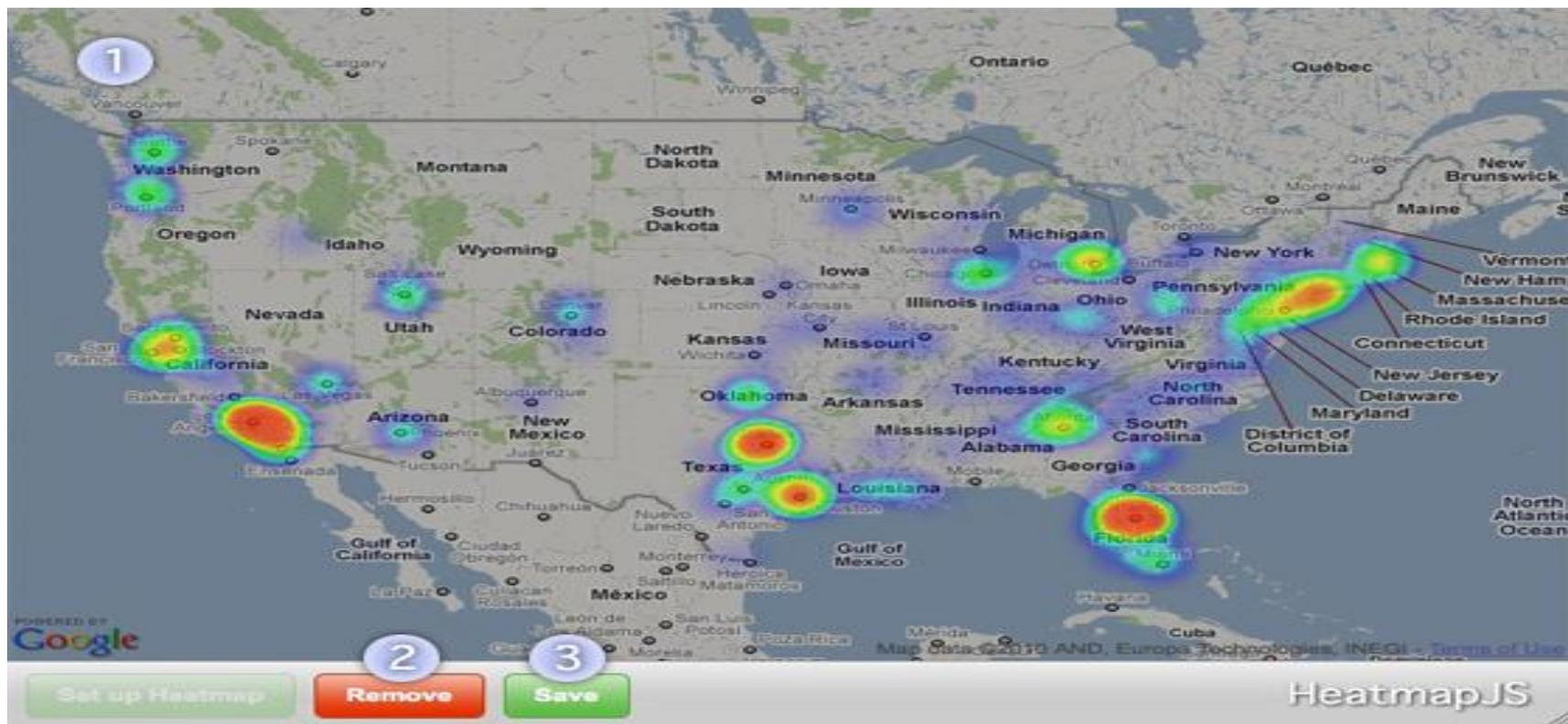
大数据分类

- 支持向量机分类
- 决策树分类
- 神经网络与极端学习机

大数据技术

计算结果展示

热图 (Heatmap) 是一项数据展示技术，将变量值用不同的颜色或高亮形式描绘出来。可以非常直观的呈现一些原本不易理解或表达的数据，比如密度、频率、温度等。



大数据技术

大数据分析展示系统

预警监控平台

全部系统

互联网出口监管平台

序号	报警类型	报警总数	标星报警
1	邮件附件	121	21
2	文件传输	0	0
3	邮件内容	0	0
4	搜索可疑	0	0
5	联系人可疑	0	0
6	言论报警	0	0
7	其他可疑	0	0

门户网站监管平台



网站总数	20165 条
检查网站	201 条
链接数	2016 条
疑似涉密网页	165 条
涉密网页	267 条
疑似涉密条目	2360 条
待处理条目	1267 条
涉密信息	3260 条
正在检查网站数	17 条
正在检查网站名	站能网



违规外联监管平台

2017年违规外联报警总数: 6

最近3条报警

设备名称	IP地址	发生时间	单位名称	责任人
设备名称名称	http://hua...	12-01 07:00	单位名称	李某某
设备名称名称	http://hua...	12-01 07:00	单位名称	李某某
设备名称名称	http://hua...	12-01 07:00	单位名称	李某某

数据量汇总

互联网出口 监管平台:	门户网站 监管平台:	违规外联 监管平台:	敏感信息 监控平台:	保密在线 检查系统:
202156	202156	202156	3012562	202156

违规事件数量



保密安全态势

终端态势 网络态势 互联网出口 门户网站 敏感信息

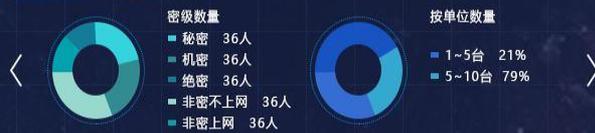
攻击类型	攻击目标	时间	攻击来源	攻击IP	危险等级	攻击端口
攻击类型	攻击目标	12-01 07:00	http://hua...	56235954	中等	TCP
攻击类型	攻击目标	12-01 07:00	http://hua...	56235954	中等	TCP
攻击类型	攻击目标	12-01 07:00	http://hua...	56235954	中等	TCP
攻击类型	攻击目标	12-01 07:00	http://hua...	56235954	中等	TCP

保密业务管理系统

涉密人员



计算机及网络



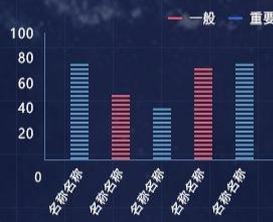
敏感信息监控系统



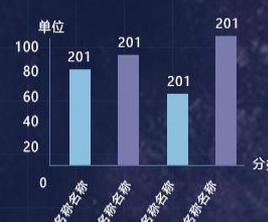
保密自查自评系统



保密实训平台



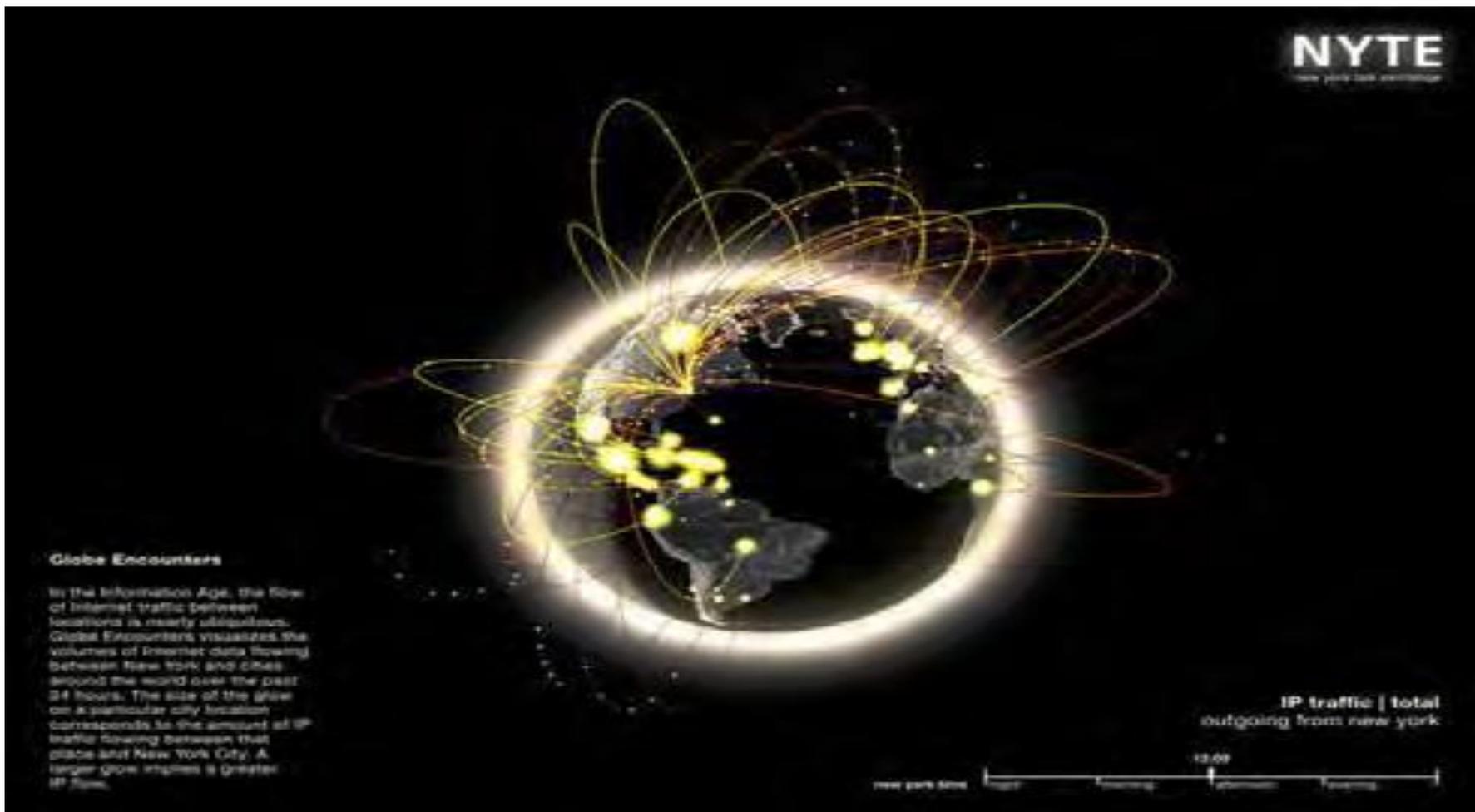
保密在线检查系统



大数据技术

计算结果展示

空间信息流 (Spatial information flow) 是展示信息空间状态的一种可视化技术。



第三部分

大数据防汛应用

大数据防汛应用



流域大洪水



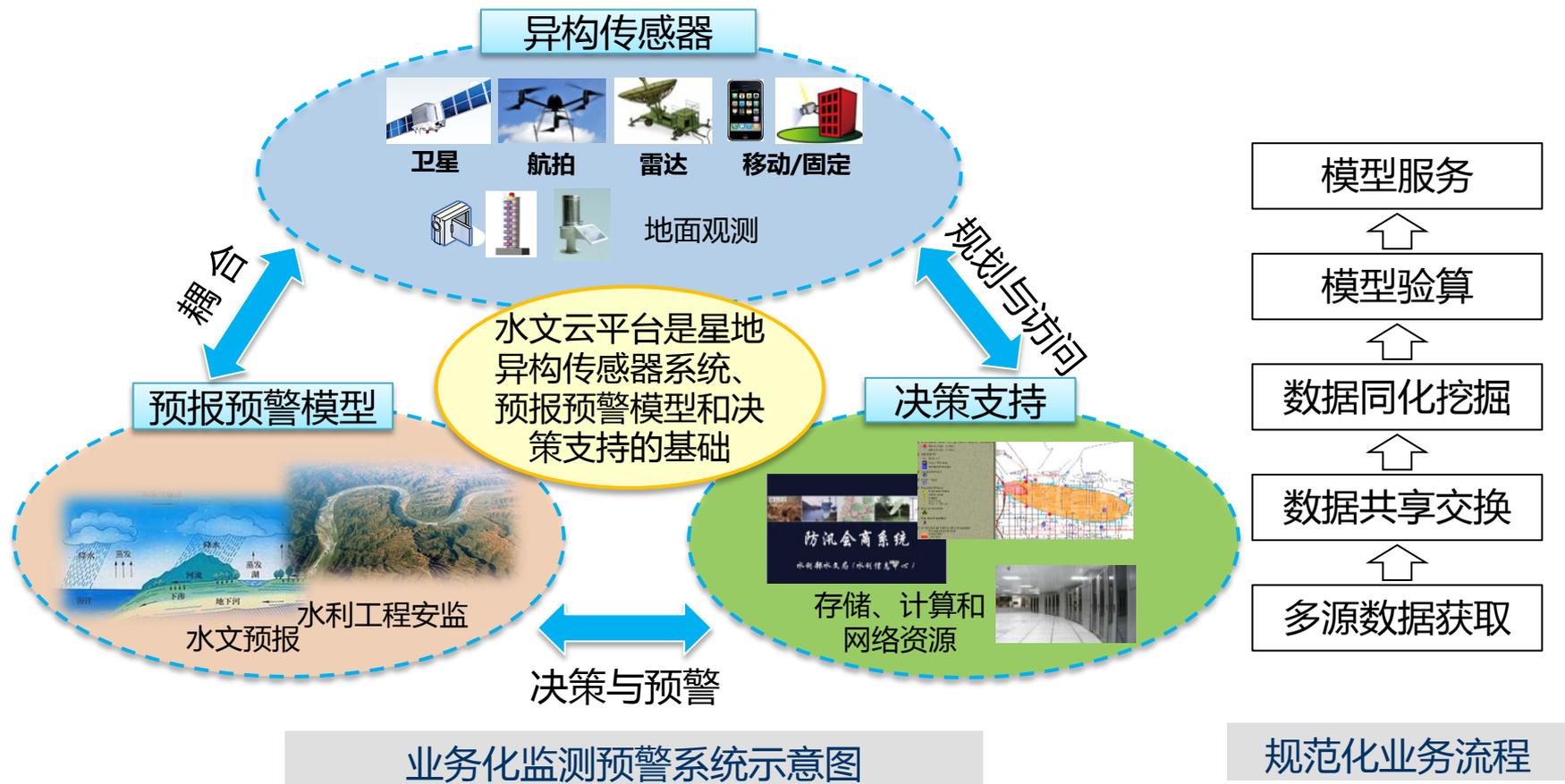
城市大暴雨



山洪泥石流

大数据防汛应用（看情势）

基于云平台的星地一体水利监测业务化系统，能够有效提高洪涝灾害、突发水事件监测预警时效性和精度



大数据防汛应用（看情势）

- 国际上对地观测计划已有一百多颗卫星可用于水利建设与管理。
- 应用范围包括：水文要素监测、大坝变形监测、水利工程立体测绘、多光谱地面过程反演、陆地水储量监测等。



大数据防汛应用（看情势）

根据国家民用空间基础设施中长期发展规划（2015-2025年），我国卫星遥感空间系统包括：

陆地观测卫星系列：高分辨率光学观测星座、中分辨率光学观测星座、合成孔径雷达（SAR）观测星座、地球物理场探测卫星。

海洋观测卫星系列：海洋水色卫星星座、海洋动力卫星星座、海洋监视监测卫星。

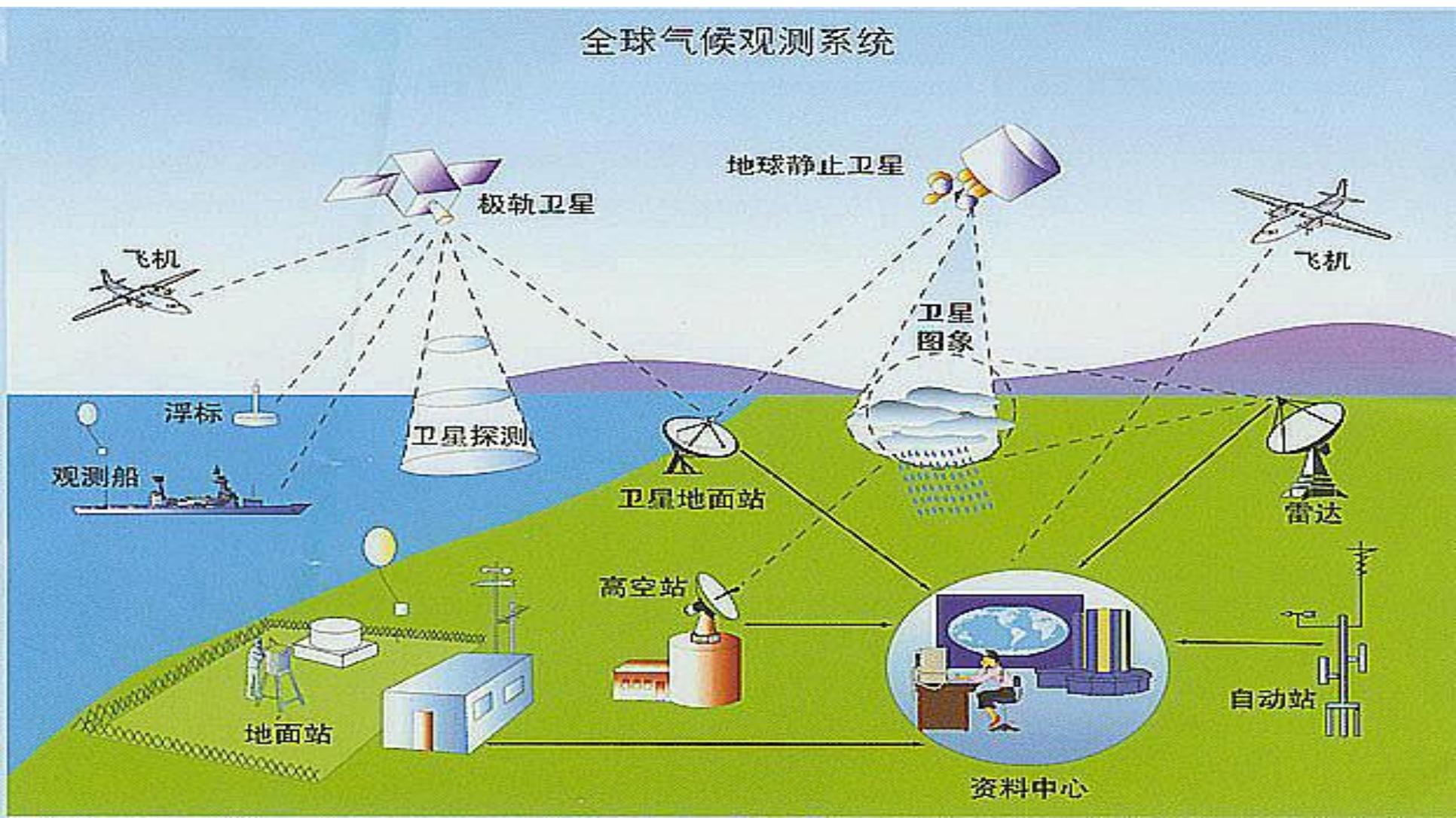
大气观测卫星系列：天气观测卫星星座、气候观测卫星星座、大气成分探测卫星。

目前，可应用于水利行业的卫星有：GF-1/2卫星、HJ-1A/B卫星、ZY-3卫星、ZY-102C卫星、SAR系列卫星、以及MODIS卫星等。

卫星型号	重要参数		
HJ-1A/B	30米分辨率，	幅宽700公里，	重访周期4天
ZY-102C	2.36米分辨率，	幅宽54公里，	重访周期3天
ZY-3	2.1米分辨率，	幅宽51公里，	重访周期5天
GF-1	2米分辨率，	幅宽800公里，	重访周期4天
GF-2	0.8米分辨率，	幅宽45公里，	重访周期5天

大数据防汛应用（看情势）

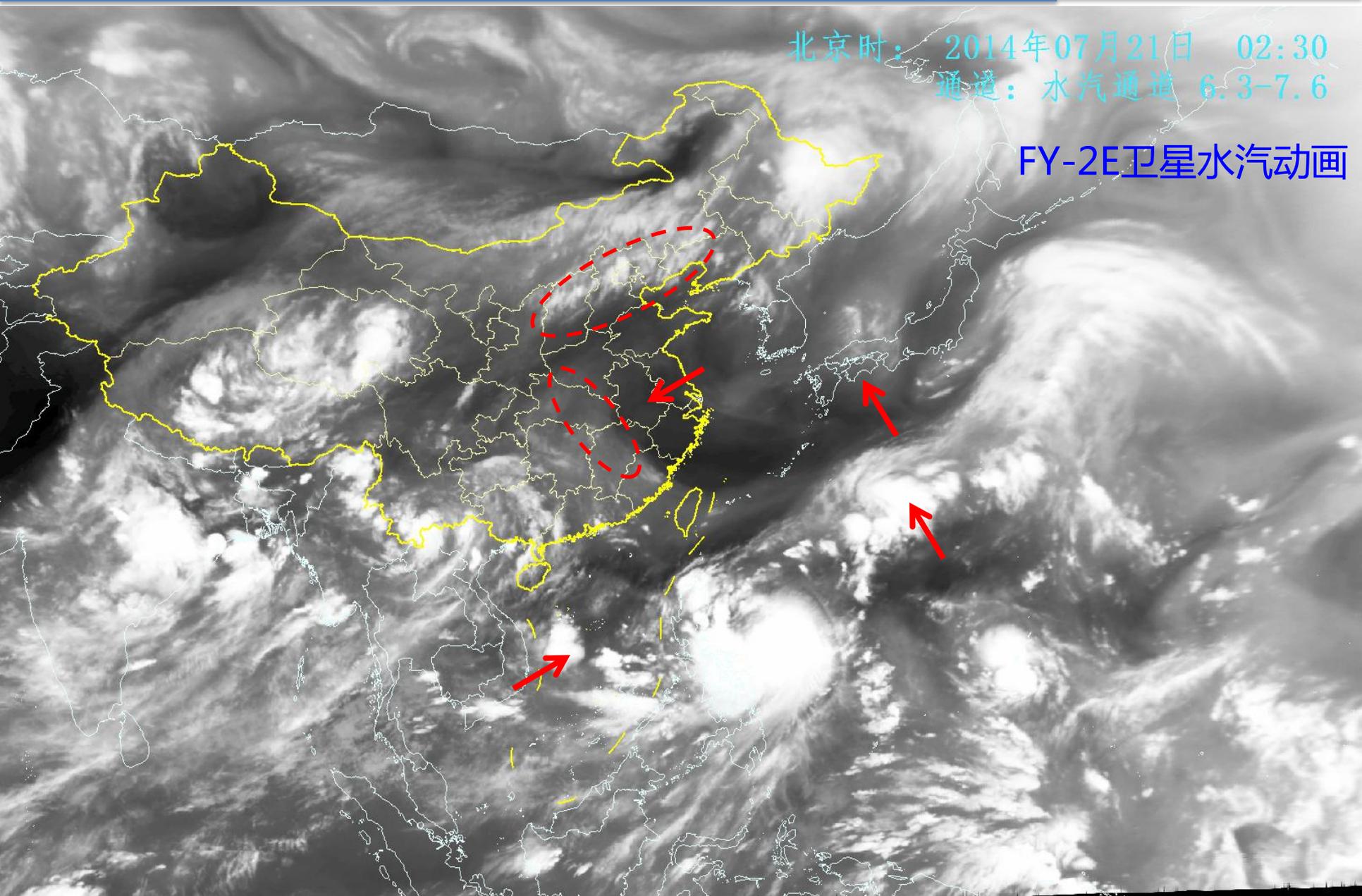
初步形成了天空地观测与数据采集体系



大数据防汛应用（看情势）

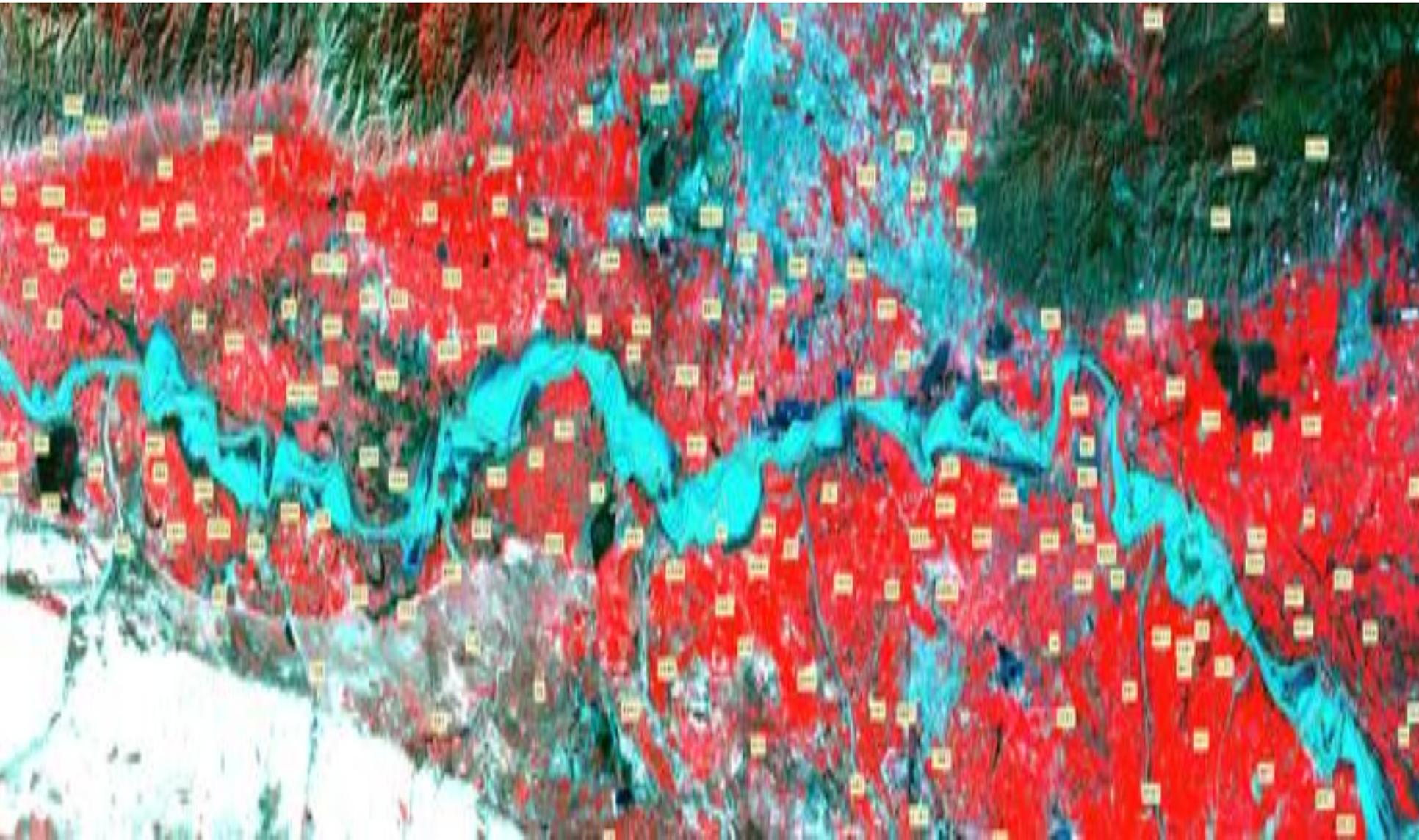
北京时：2014年07月21日 02:30
通道：水汽通道 6.3-7.6

FY-2E卫星水汽动画



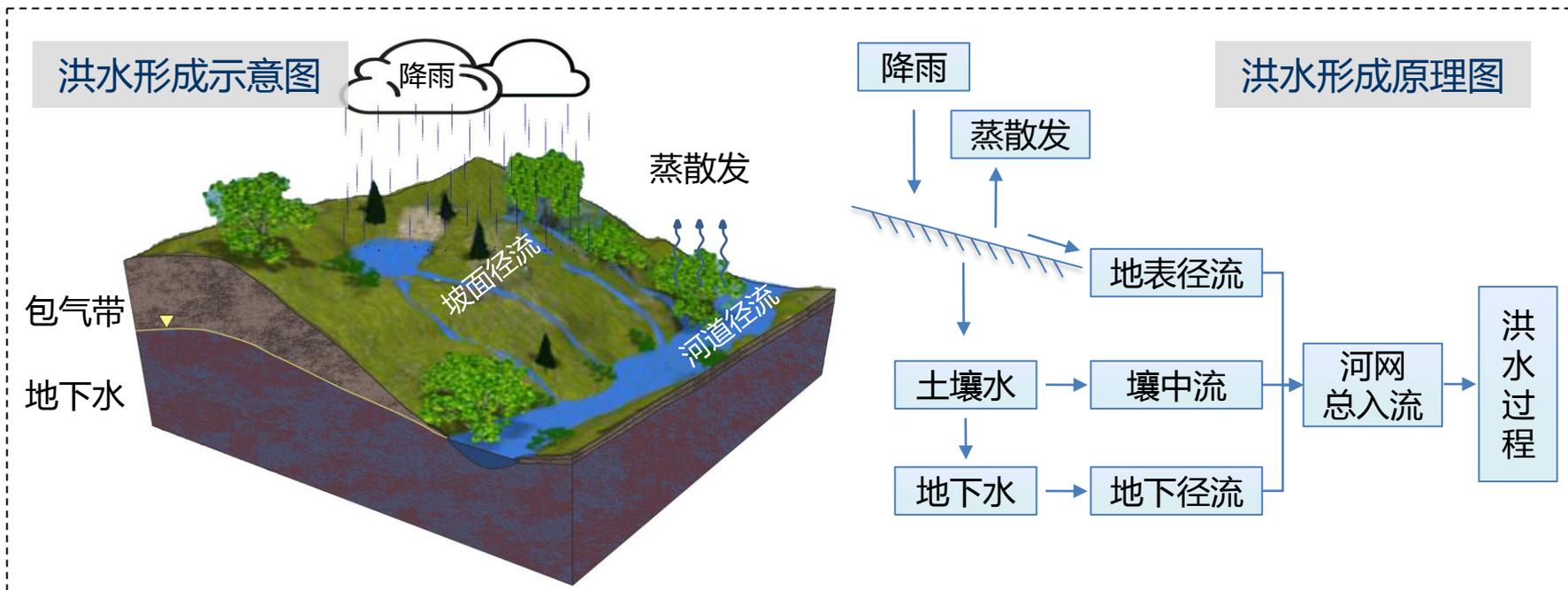
大数据防汛应用（看情势）

洪水期间掌握堤防偎水和漫滩淹没情况，城市还可结合地面监控进行观察



大数据水利应用（问规律）

洪水预报是防洪减灾最重要的非工程措施，高精度预报是建立在暴雨洪水形成的水文多要素物理过程精细刻划的基础上。



精细刻划的核心是水文、水利和气象模式的结合，其难点在于需要跨越不同模式之间的尺度差异，即将气象要素尽可能准确地降尺度到不同河道上。此外，为了确保降雨和径流预报的准确性，需要同化大量雷达降水等气象资料和其他实时观测数据等。在确保业务预报时效性的前提下，需要有效解决大数据和大计算量下的一系列问题。

大数据水利应用（问规律）

National Water Model

美国国家洪水预报系统



大数据水利应用（问规律）

Flow Continuum Model – a national stream network, atmosphere to oceans, coast to coast

一个全国连续的河网模型，连接大气、海洋、海岸

Blanco River at Wimberley

Current: 6600 basins and 3600 forecast points

Two basins and one forecast point

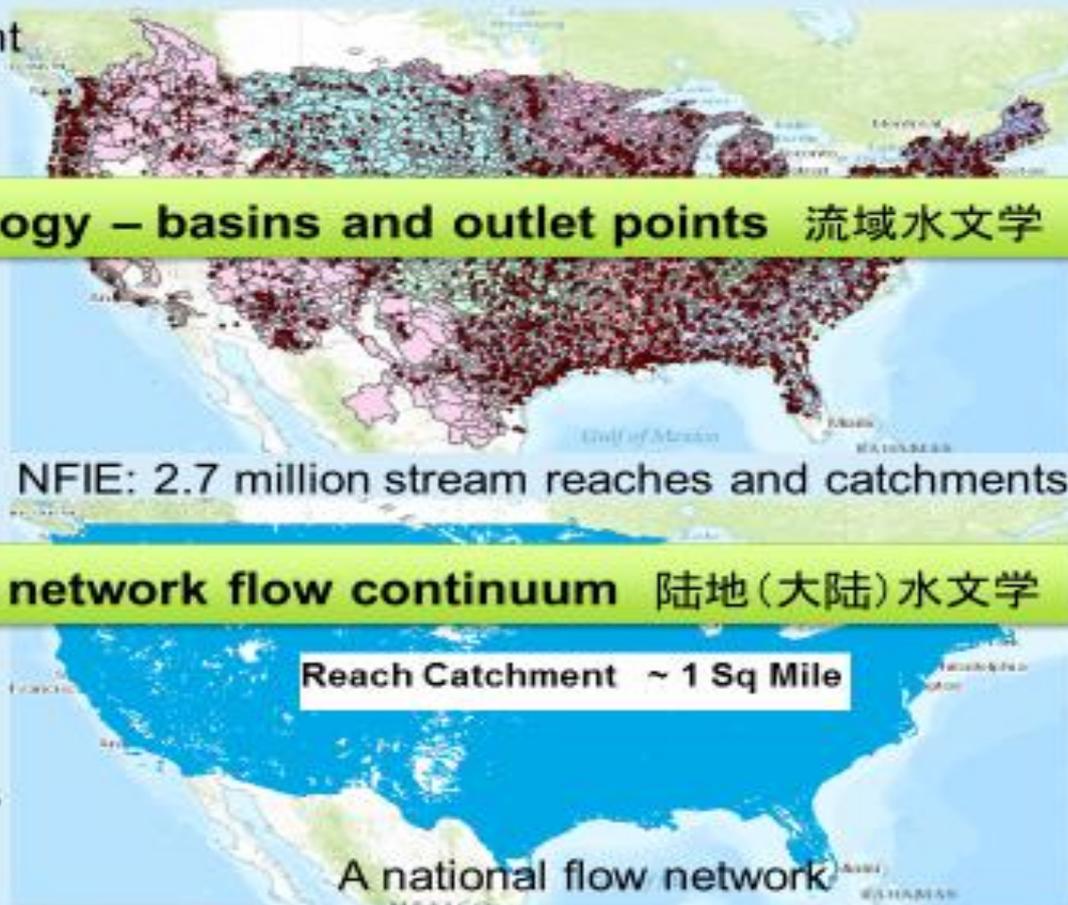
Watershed Hydrology – basins and outlet points 流域水文学

becomes



Continental Hydrology – network flow continuum 陆地(大陆)水文学

130 Catchments and Flowlines uniquely labelled

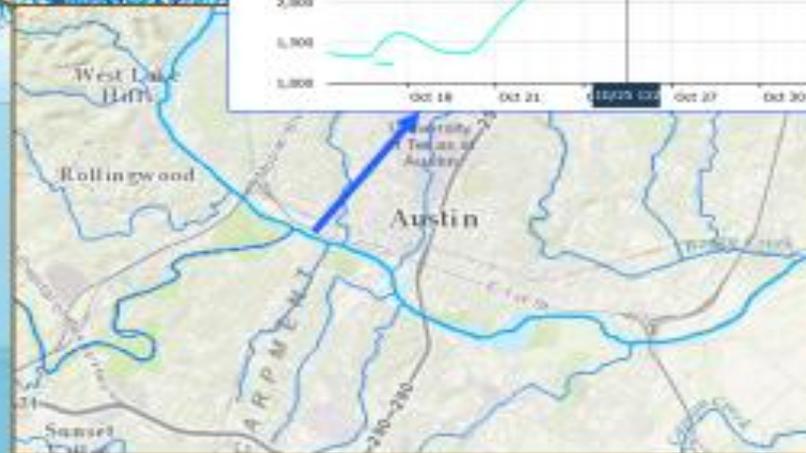
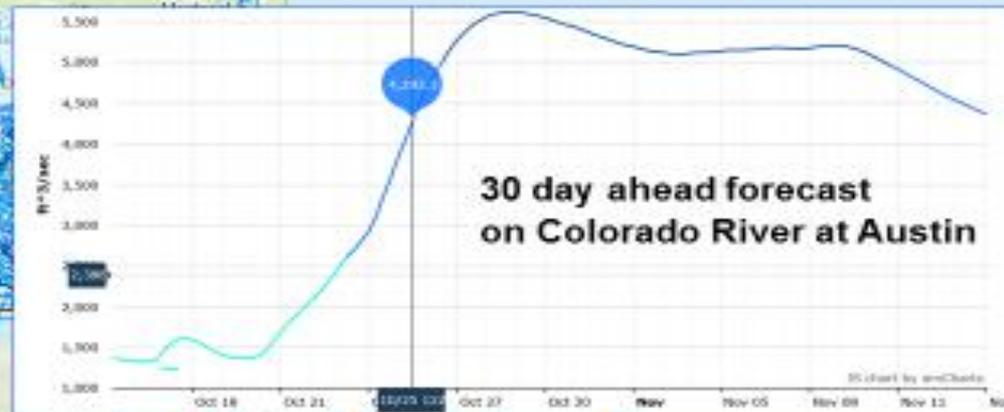


大数据水利应用（问规律）

National Water Model Forecasts

美国国家洪水预报 (对267万河段进行预报)

<http://water.noaa.gov/map>



**Operational as of
16 August 2016**

Forecasts available for 2.7 million stream reaches in continental US

大数据水利应用（问规律）

Stampede 超级计算机



1200万加仑冷却塔



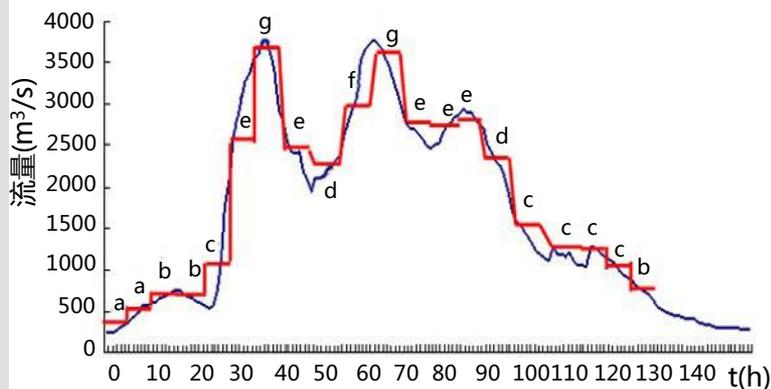
50万台处理器并行计算



大数据水利应用（问规律）

典型洪水特征模式挖掘 针对暴雨-历史洪水相似问题，进行暴雨洪水响应语义概化挖掘，构建暴雨洪水特征模式库，为实时洪水预报提供智能化校正依据

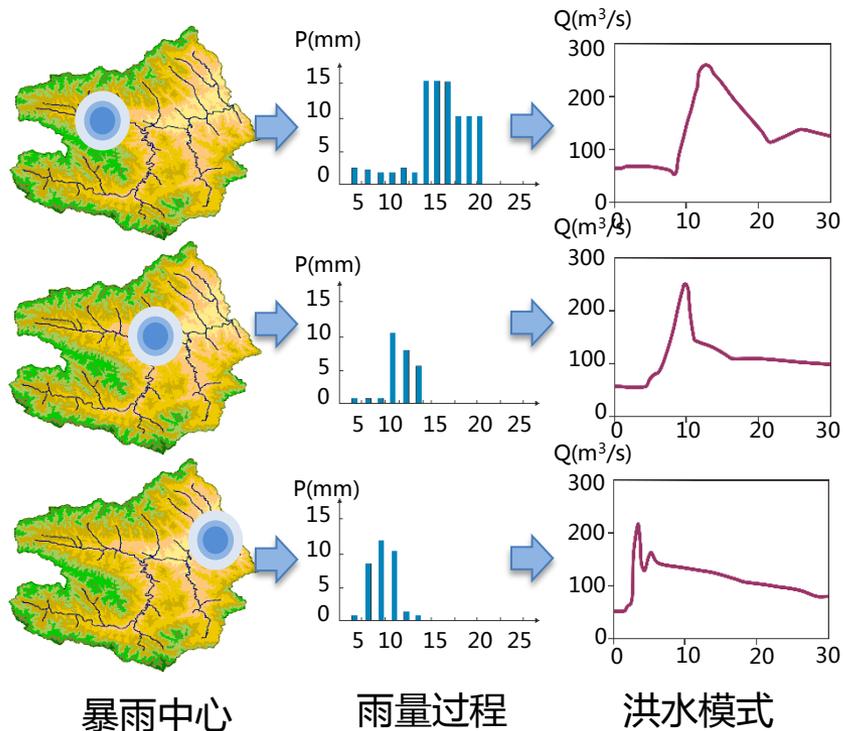
洪水语义概化



洪水模式库

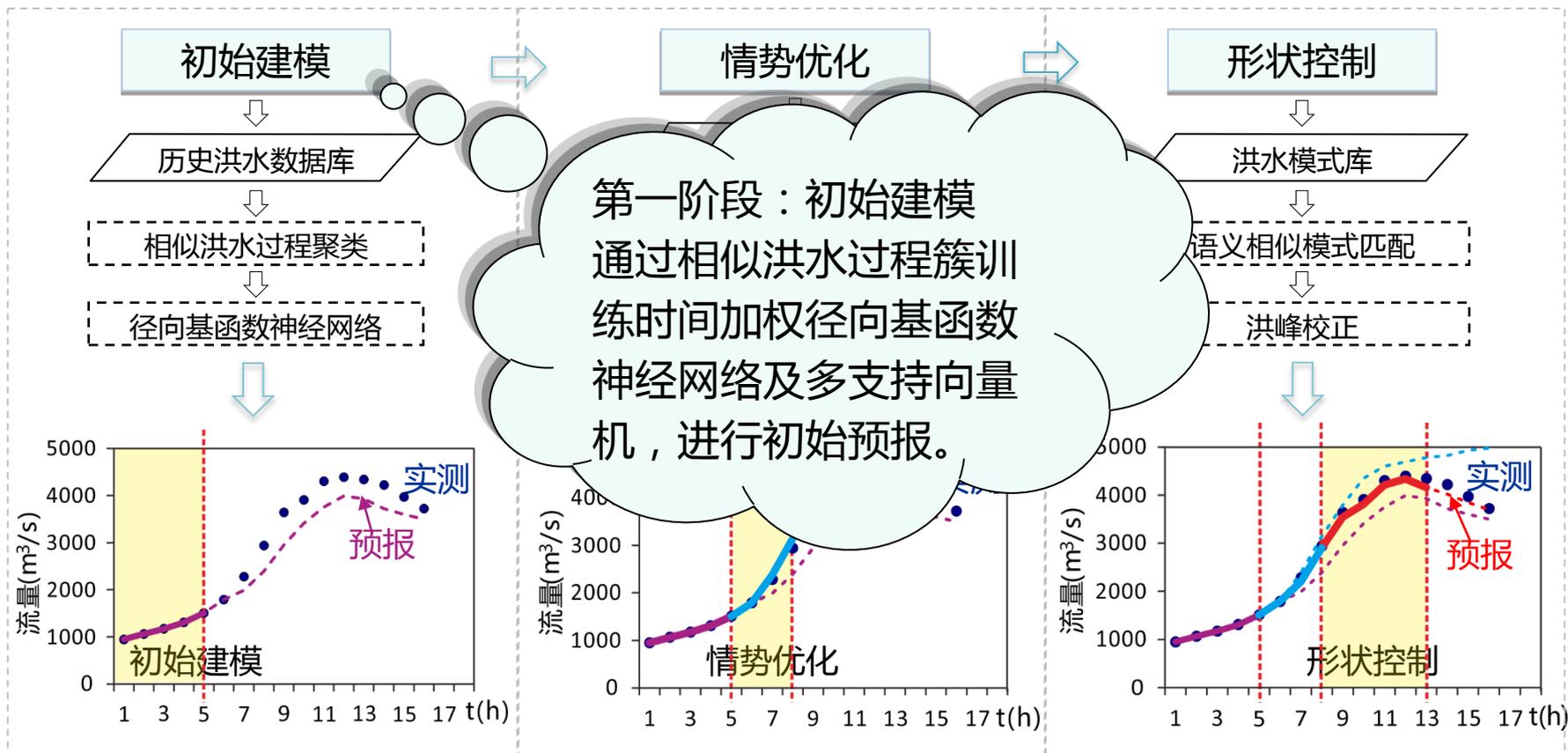
天气形势(前期降水、气温、天气环流...)
降雨时空分布(C_v, C_s , 暴雨极值(1,3,5,7天)...)
洪水过程(洪峰、峰现时间、洪量...)
洪水情势(行洪区域、区域间洪水...)
下垫面地理特征(地形指数、植被指数...)

暴雨洪水时空模式



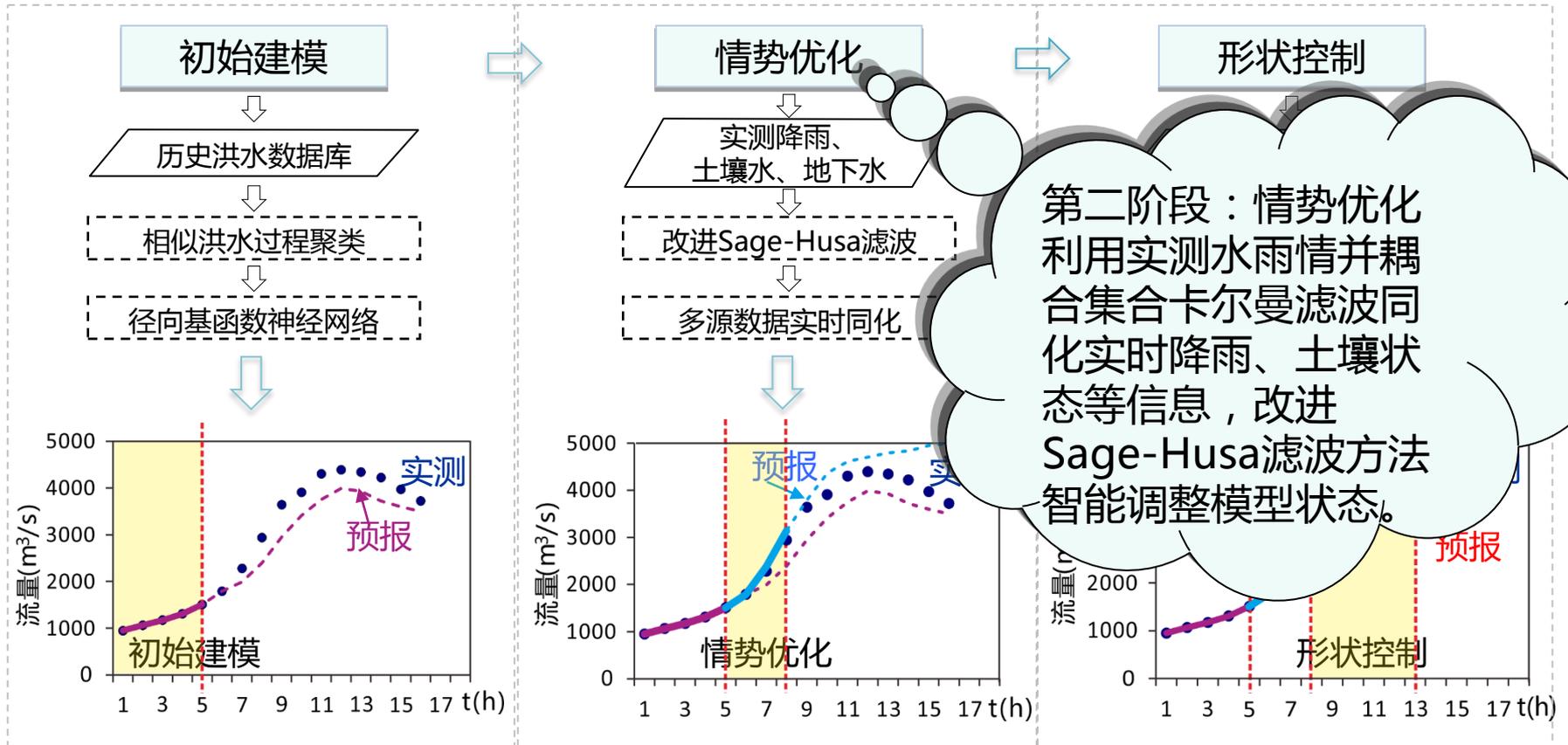
大数据水利应用（问规律）

智能预报



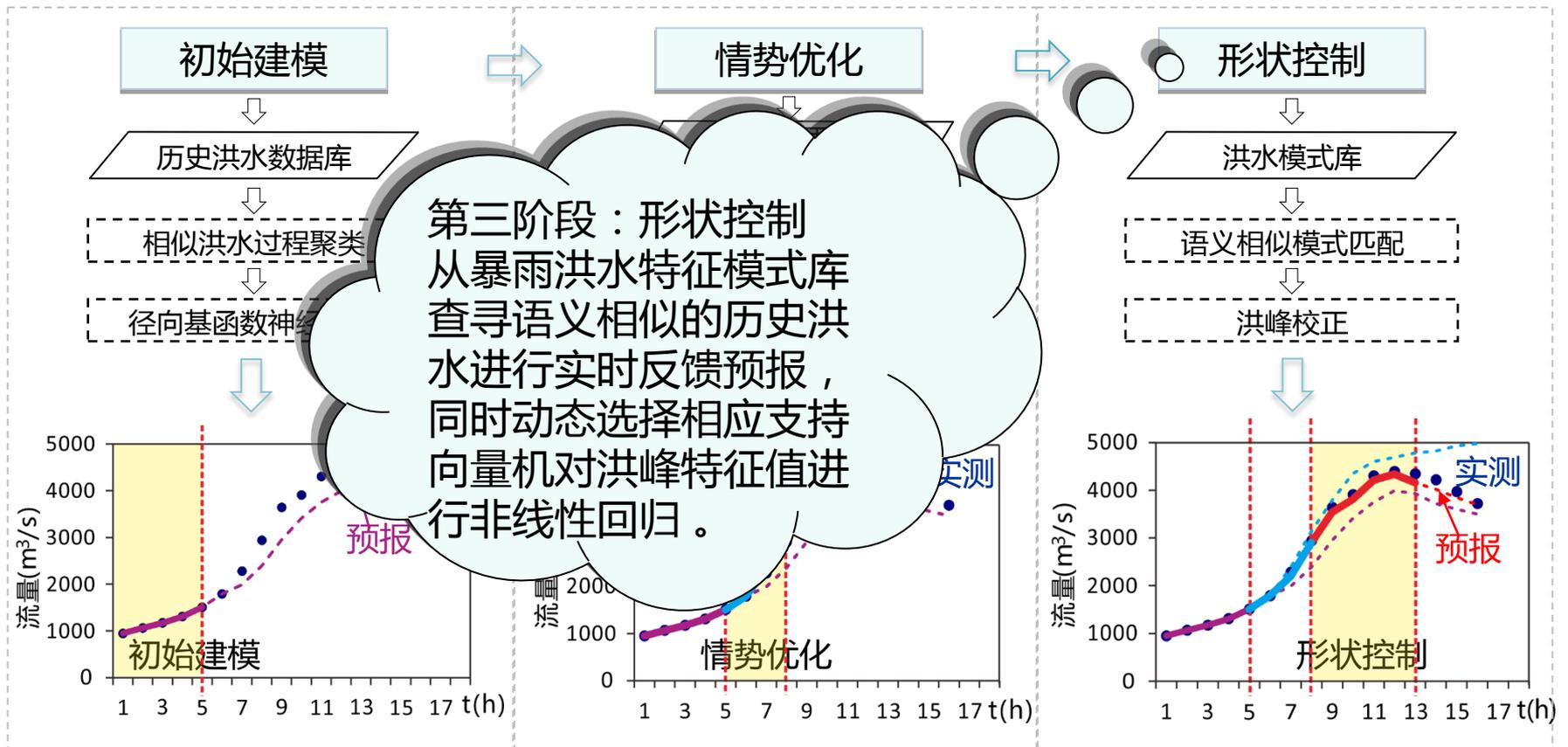
大数据水利应用（问规律）

智能预报



大数据水利应用（问规律）

智能预报



大数据水利应用（问规律）

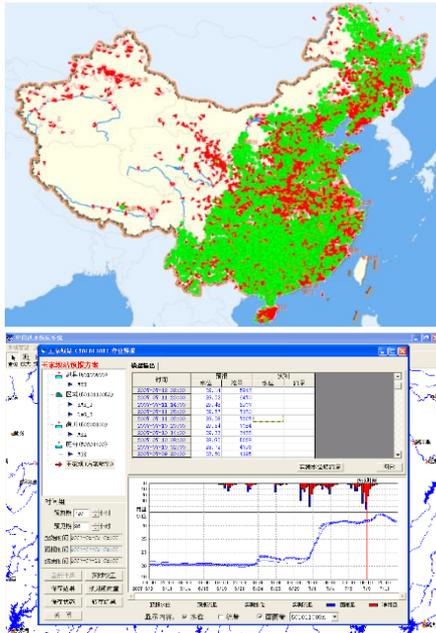
智能水文预报模型分类

序号	模型名称	典型算法	优点	典型应用
1	基于粗糙集模型	逻辑模糊算法(Fuzzy logic)等	将不确定性作为变量引入建模	水文预报过程中有水库操作的建模
2	基于进化计算的模型	蚁群优化算法(ACO),遗传算法(GA)等	对问题域直接建模；算法自适应性强	基于时序数据的水文预报建模
3	基于组合小波变化的模型	离散与连续小波变换(DWT and CWT)等	多分辨率，善于处理非平稳信号	与人工智能模型联动的水文预报建模
4	基于人工智能的模型	支持向量机(SVM)	处理噪声能力强，可解释性强	中长期水文预报建模
5		人工神经网络(ANN)	非线性建模能力强	短时水文预报建模
6	卷积神经网络(CNN)模型	卷积神经网络(CNN)	支持多特征，参数可共享，因而能处理高维数据	短时水文预报建模
7	长短时记忆网络(LSTM)模型	长短时记忆网络(LSTM)	适于处理时序信息，记忆机制，用于解决梯度消失问题	水文知识库建模

大数据水利应用（评灾情）

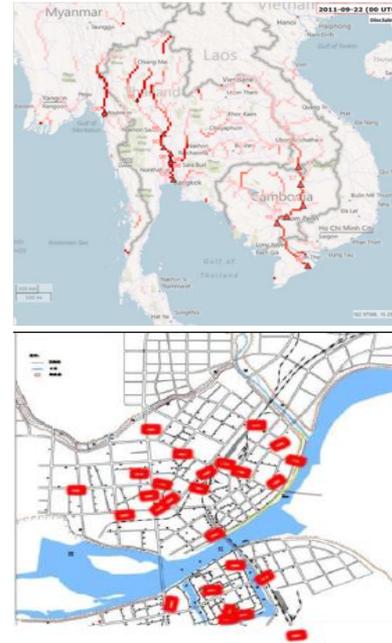
发展基于影响的预报业务：预报水灾害对江河、湖库、城市、交通、农业等目标可能的影响，实现从水文要素预报向灾害影响预报的转变；从断面、过程预报，向点、线、面多维预报转变。

现状



基于要素的预报
(水位、流量、洪水量级等)

未来

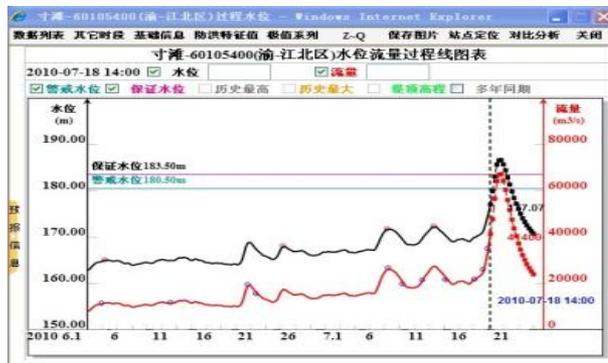


基于影响的预报
(预报数据与DEM、RS、GIS数据融合)

大数据水利应用（评灾情）

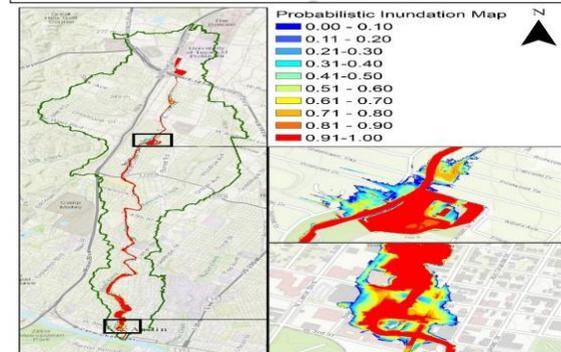
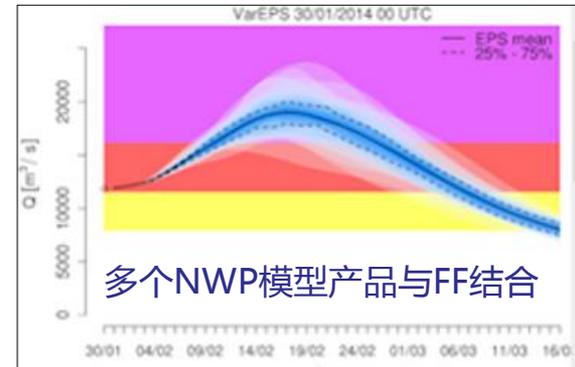
发展基于风险的预警业务：预报水灾害发生的概率及水库群、工农业、交通等目标的风险，实现将水量-流量过程转换成淹没-破坏风险信息。从过去的基于阈值预警，向针对风险信息响应预警转变。

现状



基于阈值的预警

未来



基于风险的预警

大数据的发展

合理使用大数据

大数据时代的10个重大变革

目标驱动型 → 数据驱动 (决策方式)

数据是资源 → 数据是资产 (数据的属性)

基于知识的方法 → 基于数据的方法 (方法论)

统计学 → 数据科学 (数据分析)

复杂算法 → 简单算法 (计算智能)

以战略为中心 → 以数据为中心 (竞合关系)

业务数据化 → 业务数据化 (数据管理)

不接受 → 接受 (数据复杂性)

第三范式 → 第四范式 (研究范式)

小众参与 → 大规模协同 (数据处理模式)

敬请雅正 欢迎交流
谢谢！