DOI:10.3880/j.issn.1006-7647.2016.03.001

水文学与“大数据”

芮孝芳

(河海大学水文水资源学院,江苏 南京 210098)

摘要:论述了“大数据”方法的特点和优势,即用密集的数据替代数学公式来精细地表达水文现象的时空变化或微分方程的解,讨论了水文学需要“大数据”方法的理由,以及获取水文大数据的技术支撑。认为“大数据”方法可能引发科学思维方法的创新,可能成为解决水文学一些困难问题的途径之一。

关键词:大数据;水文学;还原论;空间分布;无人机遥感

中图分类号:P33 **文献标志码**:A **文章编号**:1006-7647(2016)03-0001-04

Hydrology and big data//RUI Xiaofang(*College of Hydrology and Water Resources*, *Hohai University*, *Nanjing* 210098, *China*)

Abstract: The characteristics and advantages of the big data method are discussed in this paper. In the method, mathematical formulas are replaced by intensive data in order to precisely describe the temporal and spatial variation of hydrological phenomena or the solution of a differential equation. The reasons why hydrology needs big data and technical support for obtaining hydrological big data are also discussed. The big data method may inspire the innovation of scientific thinking and become a way of solving complex problems in hydrology.

Key words: big data; hydrology; reductionism; spatial distribution; UAV remote sensing

现在,“大数据”这一名词很热。笔者初次听到“大数据”有点不以为然,因为脑海中传统的科学思维方式与此有些不相容,认为“拼数据”不算理论研究,靠数据解决问题是笨办法。一个偶然机会接触到利用“大数据”构建精细流域水文模型的研究,笔者对“大数据”的态度才从“不以为然”变得“有点感觉”。现阶段资讯十分发达,上网检索、查询、学习方便快捷。通过了解别人的见解,嚼一嚼专家们嚼过的“馍”,再结合笔者在水文学领域“摸爬滚打”半个多世纪经常思考的几个水文学难题,觉得“大数据”的出现也许不可小视,对“大数据”的好奇和探索欲望油然而生。

诚然,“大数据”会不会像科学试验和逻辑推理那样对科学发展产生划时代的作用,现在还无法回答,笔者对“大数据”的认识和理解还处于启蒙阶段,对如何应用“大数据”解决具体问题更是心中无数,不知深浅。此时此刻,深感无能力也无资格对“大数据”进行评论,只试图结合自己的专业,从水文学的历史和现状出发,初步表达由“大数据”引起的笔者对水文学的新思考。为什么叫“新思考”?因为在此之前笔者也从其他角度,如动力学角度、系统论角度、统计物理角度等进行过思考。拙作定名为“水文学与‘大数据’”,旨在表明水文学要主动引进“大数据”的思想和方法。如果定名为“大数据与水文学”就变成“大数据”要向水文学渗透了,这不符合事实。但如果定名为“大数据水文学”,显然还未到时候,这样冒进显得不切实际。本文主要讨论 4 个问题:“大数据”时代已经来临;传统科学思维方式走出困境的一种选择;水文学需要“大数据”;水文“大数据”的获取与应用。写作本文只是为了抛砖引玉!

1 “大数据”时代已经来临

数据是通过观察、试验、计算等得出的结果。它是科学的量化,是知识的重要来源之一。数据和符号是人类认识世界,了解自然不可缺少的两个方面。数据用于量化,符号表达逻辑关系。没有数据,思维将模糊不清;没有符号,逻辑将杂乱无章,令人费解。正因为发明了数据和符号,思维逻辑才会层次分明,对世界的认识和解释才能不断清晰。

基金项目:国家自然科学重点基金(41430855)

作者简介:芮孝芳(1939—),男,教授,主要从事水文学及水资源研究。E-mail:jiangguol@hotmail.com

“大数据”的概念最早由美国未来学家 Toffler 于 1980 年提出[1]。如果望文生义地理解,“大数据”就是大量数据甚至海量数据之意,但“大数据”绝非仅仅是数据大,其深刻的含义决非如此直白。下面罗列两位有代表性的学者对“大数据”的论述。

美国著名经济学家、“长尾理论”的发明及阐述者、一个喜欢从数字中发现趋势的学者 Anderson 2008 年认为:“数量庞大的数据会使人们不再需要理论,甚至不再需要科学方法。”又说:“大数据用事实宣告,知其然,而不必知其所以然。只要知道该怎么做就行。这既是电脑不如人脑的地方,也是电脑超越人脑的地方。”他甚至大胆地说:“数据爆炸使得科学的研究方法都落伍了。”“大数据可能是许多行为科学理论乃至科学理论的终结者。”[2]

英国牛津大学教授 Schonberger 与 Cukier 两人合作,于 2012 年出版了《大数据时代》一书[2],指出:“大数据是指不用随机分析法(抽样调查)这样的捷径,而是采用全数据的分析方法。”认为:“抽样研究的缺陷是试图以少量的数据获取最大的信息。假如人们能够获得海量级的数据,抽样就自然失去它的意义,人们的抉择必然是放弃随机分析这一捷径,而采用所有数据。”

显然,上述 Anderson 对“大数据”的认识虽深刻,但有点激进。Schonberger 的认识虽近实用,但欠缺新意。

如此看来,“大数据”决非一个泛词,也许是一个有深刻科学内涵的学术名词。它不是一堆杂乱无章的数据,而是能说明复杂问题的密集数据集合,能表达自然现象或社会现象的精确存在。“大数据”与要求用数学方程和数学函数表达世界的传统思维方式不同,是一种不抛弃任何一个来自现实世界的可靠数据,凭借密集数据说话,达到准确描绘复杂世界的思维方式。世界的本质就是数据,有了“大数据”就能发现世界上的规律,没有“大数据”,对于复杂的世界可能是难以精确认识的。与“小数据”比较,“大数据”的简单算法比“小数据”的复杂算法更有效。将“小数据”搞精确,不一定比夹杂着一些不准确数据的“大数据”更能有效地描述世界;因此,“大数据”有可能促使科学理论突破。通过“大数据”的变换、整合和计算,人类可以发现新知识,创造新价值。

“大数据”时代,人类思维方式可能向更高层次发展,人类认识处理复杂问题的能力将可能由此得到提升。

2 传统科学思维方式走出困境的一种选择

因果论认为,自然界发生的各种现象都是有其原因的。刮风、下雨、电闪雷鸣、洪水、枯水、产流、汇流、结冰、融冰、气候变化、海平面上升等与人类生存与发展息息相关的水文、气象现象为什么会发生?在什么情况下发生?发生在什么地点?什么时间发生?为什么有时空变化?从根本上说都是有其原因的。揭示自然界各种现象的发生原因和变化规律从而进行预测、预报是自然科学研究的目的和任务。

欧洲文艺复兴距今已有 300 多年,虽然它只是人类历史长河中的一瞬间,但取得的科学成就却是惊人的,而且世界科学中心一直在西方[3],15—16 世纪在意大利,17 世纪在英国,18 世纪在法国,19 世纪在德国,20 世纪至今在美国,为什么一直没有出现在东方?爱因斯坦曾经的一段叙述也许能回答这个问题[4],他说:“西方的科学发展是以两个伟大的成就为基础的”,其一是“古希腊哲学家发明的形式逻辑体系”;其二是“文艺复兴时期发明的寻找因果关系的试验系统”。前者为“理论推导”奠定基础,后者为“科学实验”奠定基础。“科学实验”又称为“主动实验”。经过 300 多年的实践,这两者相辅相成,成为基本的科学思维方式。按照这种传统的科学思维方式进行科学研究的基本程序是通过揭示事物或现象发生的机理、时空变化特点,寻找或分析影响因子,基于物理和数学理论,经逻辑推理建立描述事物或现象时空变化所遵循的微分方程式,再解微分方程式,得出定量表达的事物或现象的时空变化规律,用于预测、预报其未来。对这种传统的科学思维方式习以为常,就会造成一种印象:找出描述事物或现象时空变化所遵循的微分方程或数学函数才是科学研究的最高境界,若未达到此境界,就不算取得完全成功,甚至坚信这种最高境界总是可以达到的,今天达不到,也许日后某一天就能达到。

上述爱因斯坦所归纳的传统科学思维方式,在哲学上属于“还原论”思维方式。“还原论”认为[5],通过将事物或现象分解成越来越细的组成部分就能最终对其发生、发展的原因进行彻底的理解。“还原论”是一种化整体为部分、化复杂为简单的思维方式。但是,自然界有些现象是极其复杂的,随着科学技术的发展,这种极其复杂性越来越凸显,“还原论”思维遇到挑战,行不通的情况屡屡发生。束缚在传统的科学思维方式里,路子似乎越来越窄,几乎到了“山穷水尽疑无路”的地步,面对极其复杂的自然现象,似乎越来越感到束手无策。在水文学中,现行中长期预报方法和试图通过布设雨量站网精准掌握雨量空间分布之所以陷入困境,也许就是这个原因。这就提出了一个令人深思的问题:难道世界上事物的发展,所有现象的时空变化都能一一找出所

有的影响因子并导出数学中已经存在的或可预言的微分方程式吗？难道所有已经导出的微分方程都能用现有的数学理论和方法求解吗？

自然界发生的各种现象无不直接或间接地作用于人的感官，然后传递给大脑，而且随着科学技术的发展，还将不断发明出各种仪器仪表帮助人类对各种自然现象进行观测，观测的精细、精确程度也越来越高。人们不禁要问：这些“被动实验”所观测得到的结果到底是什么？难道仅仅是一些用来验证由传统科学思维方式所获得的那些现象的时空变化规律的数据吗？否！如果从因果论出发进行逆向思维，即不是着眼于由“因”求“果”，而是试图由“果”寻“因”，那么就会产生另外一种理解：人们通过观测所得到的自然现象的时空变化数据，实际上就是描述其时空变化规律的微分方程在一定初始条件和边界条件约束下的一个解。也就是说，虽然根据迄今为止人们积累的知识还不能导出这个微分方程式，也不知道受控制的初始条件和边界条件是什么，甚至永远无法导出这个微分方程及相应的初始条件和边界条件，但它的解却被人类真真切切地捕捉到了，观测到了。因此，人类是否应当改变一下思维方式，即从观测到的现象的时空变化数据（描述该现象时空变化所遵循的微分方程在一定初始条件和边界条件下的一个个解）入手，达到揭示规律、预测预报未来的目的。这种逆向思维方式是传统思维方式的补充，也是“还原论”思维的一种新模式。

笔者冒昧地认为，“大数据”也许是实现上述逆向思维，由“果”寻“因”的有效途径之一。获取密集“大数据”技术虽不是建立微分方程的方法，但应该是一种寻求微分方程在一定初始条件和边界条件约束下的解的方法。建立微分方程原本不是目的，而是手段，求得解才是目的。因此，“大数据”将可能成为处理传统科学思维方式难以处理，甚至无法处理的复杂自然现象，让传统科学思维方式走出困境的一种选择。

因果论是人们观察事物和现象的根本哲学观点，根据这个观点，可具体产生两种科学思维：一是“还原论”思维；二是“还原论”逆向思维，也可称为“逆还原论”思维。前者由“因”求“果”，后者由“果”寻“因”。若将由“因”求“果”比喻为“事前诸葛亮”式思维，则由“果”寻“因”可喻为“事后诸葛亮”式思维。两者均受形式逻辑指导，但前者以科学实验即“主动实验”为基础，而后者以科学观测即“被动实验”为基础。

3　水文学需要“大数据”

水文学的研究对象是水文现象。水文现象是降水、蒸发等气象现象与地形、地貌、土壤、植被、地质、水文地质等下垫面条件共同作用或相互作用的产物。气象现象通过其时空变化影响水文现象。下垫面条件对水文现象的影响则体现为对气象现象的作用。容易理解，如果不存在下垫面，那么水文现象将完全等同于气象现象。下垫面条件对气象现象的作用可概括为“损失”作用和“调蓄”作用[6]。“损失”作用表现为“筛子”、“门槛”和“蒸散发”作用。“调蓄”作用表现为“推移”和“坦化”作用。现在，水文学虽有了长足进步，但制约其进一步发展的瓶颈也日益凸显，尤其是如何真实地描述气象现象时空分布和下垫面条件空间分布以及如何考虑其对水文现象的影响，是长期以来没有得到解决、困扰水文学发展的问题[7]。

时至今日，水文学家认识和掌握气象现象和下垫面条件时空变化的基本手段几乎没有根本性的变化，仍是布设站网、定点定时或连续观测。这种观测手段必然难以真实描述气象现象和下垫面条件的时空变化。20 世纪 70—80 年代，曾经有人试图通过“站网优化”来解决这个问题，但这种所谓的“站网优化”却是一个没有唯一解或者几乎无法实现的命题。水文学家长期以来有一种幻想，认为随着科学技术的发展总能找到一种数学函数来精确描述气象现象和下垫面条件的时空分布或者总能找到一种高精度的插值方法来推求无观测站点处的气象过程和下垫面条件。但这种幻想至今仍是幻想，而且人们越来越意识到实现无望，以致与 30 多年前比较这类研究的热度大幅降温。

数字高程模型（DEM）的出现，使水文学家从上述幻想中觉醒，并看到了新的希望。地形高程的空间变化复杂多变，最早是根据空间上有限测点高程通过内插绘制的地形等高线来表示其空间变化，后来发展的地统计法，又试图通过回归分析得到的趋势面数学表达式来描写地形的空间变化。20 世纪 60 年代，美国麻省理工学院教授 Miller 将计算机与测量技术结合起来创建了 DEM[8]。DEM 以一组有序数值的阵列形式，在计算机软件支持下，达到了精细、精确表达地面高程空间变化的目的。用于构建 DEM 的数据可以十分密集而庞大，它不像地形等高线和趋势面分析那样通过抽取样本做分析，而是利用全部密集的海量数据做分析，可见，DEM 描述地面高程空间变化的真实性是传统的地形等高线或趋势面数学表达式难以比拟的。DEM 的成功使人们相信，欲真实描述现象时空分布变化，不一定通过数学函数，也不必苛求内插方法，而是可以通过密集的具有阵列形式的大数据来刻画的。这样的表达方式将更直观、更真实、更精彩[9]。

水文学家由此得到的启示是:人们观测到的气象现象时空变化和下垫面条件空间分布,是其所遵循的微分方程在一定初始条件和下垫面条件约束下的一个解,对其进行精细、真实的表达将寄希望于密集的“大数据”。用“大数据”破解水文学中一些难题是水文学未来发展方向之一。

4 水文“大数据”的获取与应用

水文“大数据”将如何获取?在当代科学技术条件下将寄希望于遥感技术。遥感是一门对地观测技术,物理基础扎实牢固,能实施大面积密集观测,具有测量要素综合、时效性强、获取信息受限条件少等特点,只是目前所能达到的精度还有待提高。

按搭载传感器的平台,可将遥感分为航空遥感和卫星遥感两类。卫星遥感因卫星位置很高,因而分辨率较低。对于航空遥感,飞机有导航系统,价格昂贵,也不可能一直在飞行。现在正在发展的新型遥感——无人机遥感[10],能看到更小的区域,任何时间、任何地点皆能飞行,可以做到一个小流域就可投放一个无人机,只需在云层下面飞行,成本低,因而受到青睐。目前无人机遥感的分辨率可以达到10cm,这是卫星遥感难以达到的。

应用“大数据”解决水文学的一些难题将不同于传统的“数据挖掘技术”,也不是应用现有的数值计算方法就万事大吉,这需要水文学的创新,更需要计算机算法创新。笔者近期将要发表的《随机产汇流理论》和《单元嵌套网格产汇流理论》两篇论文将尝试初步回答这一问题。

5 结 语

“大数据”的本质特征是凭数据说话,用“大数据”认识世界。这与传统方法通过微分方程和数学函数认识世界相比,更具有普适性,将可能成为处理复杂问题的途径之一,由此有可能引发数学革命和算法革新,“大数据”产生的新的数学和算法问题可能成为数学和计算机科学的重要内容之一。“大数据”提供的产生新知识和科学新发现的模式有可能使传统科学思维方式走出困境。

水文现象是最复杂的自然现象之一。试图依赖现有的物理定律和数学方程来认识、预测预报如此复杂的自然现象,事实证明困难重重。“大数据”也许是破解水文学中一些难题的途径之一。人们已经发现“大数据”在分析水文现象时空分布,揭示水文特征与地貌特征关系、构建真正的分布式流域水文模型或者精细流域水文模型等方面起着重要的作用。建立微分方程是物理学家的任务、解微分方程是数学家的任务的观点只能代表传统科学思维方式,并不代表未来。未来也许是“大数据”的未来。“大数据”将成为推动水文学发展的重要驱动力。笔者认为水文学的未来将会是:对于根据普适性物理定律、本构关系就能解决的问题,取得数据只是为了验证,“小数据”就足矣;对于必须借助密集“大数据”才能精确描述的现象,“大数据”必然是一种新的科研方法;对于那些时空无限延伸的问题,如水文现象年际演变问题,由于永远不可能掌握其全部数据,按样本推断总体的科学思维方式仍是必须的。笔者希望有人来创立和发展“大数据水文学”。

当然,对“大数据”也有不同声音,例如美国亚利桑那大学 Fricke 教授就认为“被称为第四种范式的数据驱动的科学其实是一个不伦不类的怪物”“科学需要更多的理论,而不是更多的数据”[11]。有不同的见解是正常的,善于听取不同的声音,对自己的研究工作百利而无一害。

参考文献:

[1] TOFFLER A. 第三次浪潮[M]. 黄明坚,译. 北京:中信出版社,2006.

[2] SCHONBERGER V M, CUKIER K. 大数据时代[M]. 盛杨燕,周涛,译. 杭州:浙江人民出版社,2013.

[3] 李铁林. 世界科学中心的转移与一流大学的崛起[D]. 湖南师范大学,2009.

[4] TOBY H. 近代科学为什么诞生在西方?[M]. 2 版. 北京:北京大学出版社,2010.

[5] 刘明海. 还原论研究[M]. 北京:中国社会科学出版社,2012.

[6] 芮孝芳. 水文学原理[M]. 北京:高等教育出版社,2013.

[7] 芮孝芳. 水文学前沿科学问题之我见[J]. 水利水电科技进展,2015,35(5):95-102. (RUI Xiaofang. Discussion of some frontier problems in hydrology[J]. Advances in Science and Technology of Water Resources. 2015,35(5):95-102. (in Chinese))

[8] 李志林,朱庆. 数字高程模型[M]. 武汉:武汉测绘科学科技大学出版社,2000.

[9] 宫兴龙. 流域地貌特征及流域地貌单位线研究[D]. 南京:河海大学,2009.

[10] 李德仁,李明. 无人机遥感系统的研究进展与应用前景[J]. 武汉大学学报(信息科学版),2014,39(5):505-513. (LI Deren, LI Ming. Research advance and application prospect of unmanned aerial vehicle remote sensing system[J]. Geomatics and Information Science of Wuhan University,2014,39(5):505-513. (in Chinese))

[11] FRICK M. Big data and its epistemology[J]. Journal of the Association for Information Science and Technology,2015,66(4):651-661.

(收稿日期:2016-01-11 编辑:骆超)