

对流域水文模型的再认识

芮孝芳

(河海大学水文水资源学院,江苏 南京 210098)

摘要:讨论了流域水文模型的本质,指出模型只是原型的近似,不可能反映原型的一切特征,尤其像流域降雨径流形成这样复杂的现象。诠释了“分解”与“集成”的科学思维在流域水文模型建模中的作用。认为只有产汇流理论和集成基本水文现象为流域水文模型的方法有所突破,才会有流域水文模型的实质性进步。分析了流域水文模型反问题不适定的原因,提出了直接最优化方法用于解决流域水文模型反问题时应注意的事项,以及用“大数据”方法解决流域水文模型反问题的设想和初步思路。认为随着水文观测手段的进一步精准,“大数据”方法在解决流域水文模型反问题和正问题中均有美好的应用前景。

关键词:流域水文模型;产汇流理论;分解;集成;反问题;不适定问题;直接最优化方法;大数据
中图分类号:P33 **文献标志码:**A **文章编号:**1006-7647(2018)02-0001-07

More discussion of watershed hydrological model//RUI Xiaofang(*College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China*)

Abstract: The nature of watershed hydrological model is discussed. Models are always approximations for prototypes and a model cannot reflect all the characteristics of a prototype, especially for the complicated phenomena such as the formation of rainfall-runoff of a watershed. The role of scientific thinking for decomposition and assembly in the establishment of watershed hydrological models is explained. Substantial progress of watershed hydrological models can be made only if the breakthrough of both the theories for runoff yield and concentration and the assembly methods from basic hydrological phenomenon to watershed hydrological models are realized. The reasons of ill-posed problems in the inverse problems of watershed hydrological models are analyzed. Some shortcomings of the solution for inverse problems of watershed hydrological models by means of direct optimization methods are pointed out. Preliminary ideas on the solution for the inverse problems of watershed hydrological models through big data are proposed. With more accurate and refined hydrological observation, big data method has bright prospects in the solution for both direct and inverse problems of watershed hydrological models.

Key words: watershed hydrological model; theories of runoff yield and concentration; decomposition; assembly; inverse problem; ill-posed problem; direct optimization method; big data

半个多世纪以来,流域水文模型就像幽灵一样在世界游荡着,忽近忽远,忽明忽暗。时至今日,水文学家对流域水文模型的认识虽已从盲目乐观走向理性对待,但在一些水文学家心目中又正面临着“弃之可惜,食之无味”的两难选择。一个个新冠名的流域水文模型仍在不断面世,从国外引进流域水文模型的热度仍不见消退,著书立说者并不少见,作为商品入市者也不在少数。不过冷静后仔细剖析,不免感到这仅仅是表面的热闹。在现有的数以百计的流域水文模型中,能具有水文学新意的可能不足10%,尤其在洪水预报实践中,应用流域水文模型进行预报能够取得令人满意预报结果的更是罕见。为了客观地认识流域水文模型的潜在价值和面临的困难,寻求其健康发展之路和通向应用的桥梁,笔者近年来曾撰写《论流域水文模型》^[1]一文,就流域水文模型的结构和参数、集总式模型和分布式模型、模型的解算、不敏感参数和异参同效、模型的验证和比较等问题展开了讨论,但仍感余兴未尽,对流域水文模型的继续思考从未停止。现将笔者近期对流域水文模型的新思考写成本文,旨在与广大读者切磋,分享研究之乐,聆听各方指教。本文涉及5方面内容:一是模型之本质,二是“分解”与“集成”,三是正问题和反问题,四是最优化方法解反问题,五是“大数据”方法解反问题。

1 模型之本质

自然科学研究的目的是发现和理解自然规律。在复杂的自然面前,有人坚持世界是可认识的,世界最终将被知之;另外一些则觉得完全认识世界极其困难,世界是不可知的。站在不同的角度或尺度观察或理解同一问题,既可能得出不同的结论也可能得出相同的结论,这也是常有的事。科学家们就是这样从各自的领域或所处的角度对世界进行着各自的阐述。令人可喜的是,正是这样的不同阐述,才使得人类对自然的认识不断丰富和完善,不断从错误走向正确。生活在17—18世纪的德国哲学家、物理学家、数学家莱布尼茨曾指出“自然从来不飞跃”,认为“事物永远要经过程度上和部分上的中间阶段才能从小到大或从大到小”,即“连续律”是宇宙间的一条根本规律^[2]。据此,可以将自然现象的时空变化视为数学上的连续函数。人们对自然现象的全部认识就相当于要理解这个连续函数上的每一个点,但这几乎是不可能做到的。因此,人们只能通过采样,也就是离散的方式来观察并理解连续函数上的一部分点,从而形成对连续函数在一定条件下表现出的性质的理解。采样总是要分时间和空间步长的,总是在一定范围内一定时间点采样的。采样的步长有长有短,采样的范围有大有小,采样范围所处位置也有不同(图1),采样间隔时间有长有短,采样的技术和精度也在演变中,对同一现象的表现就有可能得出相同或不同的结论。

模型就是对自然规律进行采样分析的一种工具。人类对自然的认识总是从“典型”开始的,这相当于采样。当认识的“典型”达到一定数量,就可进行归纳、总结。当采样足够多,就可分析描述出其中某一段的近似函数,这就形成了模型。所以,模型只相当于连续函数上某一段,这个某一段就是模型的适用范围,实际上也就是理论的适用范围,超出这个范围,模型或理论就不适用了。人们建立流域水文模型时所采用的基本假定就是它的适用范围。任何模型或理论都有其存在的前提,离开这个前提很可能就是谬误。不讲前提,模型或理论很可能是胡言乱语。按照极限概念,如果在连续变化的时间和空间上采样点多到无穷,那么人们得到的观察和理解就可能逼近连续函数的真实情况。不过,由于自然的无限宽广,凭人类有限的认识和理解,至今仍无法找到作为连续函数的自然现象的边界。这样,站在不同的角度、采用不同的分析尺度和手段就有可能得到相同或不同的结果,而真实的规律也许仍游荡在人们的认识和理解之外。然而,人们完全没有必

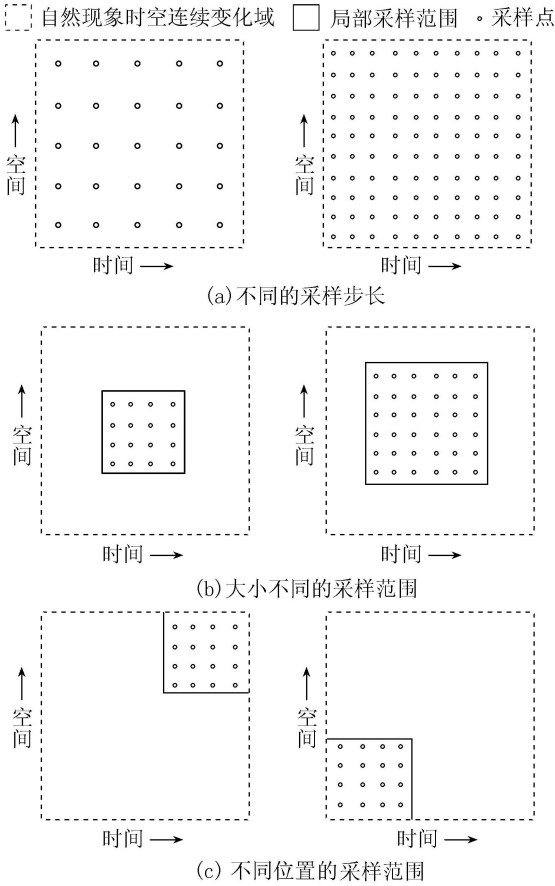


图1 自然现象时空连续变化范围及采样点分布

要为此感到悲哀,陷入苦恼之中。正因为存在着自然变化无限与人类认识有限的对立,世界才变得无比精彩,魅力无穷,使人们在清楚与模糊的交替中享受着科学研究的乐趣。模型为人们提供了在一定范围内,一定程度上认识自然规律的可能,这就是它的潜在价值。

流域水文模型的出现是水文学家对流域降雨径流形成机理和变化规律认识的一个飞跃,但模型毕竟只是实际存在的原型的一种近似或类似,与原型相比必然存在差别。原型一般有很多层次和特征,十分复杂,因此,真正能反映原型一切特征的只能是原型本身。这就是流域水文模型不可能将流域降雨径流形成机理和变化规律的一切方面都反映出来的原因。为着一定目的而建立的流域水文模型,只要能表达出人们最需要了解的那些特征就可以了。

不论是否能将流域水文模型表达成解析形式,它实际上可以看作是一个数学物理方程的定解问题,模型结构就是泛定方程,模型参数就是这个泛定方程的系数,模型的状态变量就是所包含的未知函数,模型的激励、初始和边界状态就是定解条件。这是从物理与数学的角度对流域水文模型的理解。也可以用系统论来理解模型。描写系统输入与输出之间关系的结构就是模型。模型与输入、输出一起就

成为系统。模型是内因,是事物或现象的本质;输入是外因,是模型的驱动力。外因通过内因起作用,就产生了模型的输出。

2 “分解”与“集成”

“分解”和“集成”也许是认识并揭示复杂现象形成机理和时空变化规律的一种有效研究方法。所谓分解就是“拆零”,就是将复杂问题尽可能拆成一些细小的独立的“碎件”,以便于观察这些“碎件”的发生和变化,通过科学实验和理论分析将其形成机理和时空变化规律揭示清楚。所谓集成就是“组装”,就是将那些已经被揭示了的机理和时空变化规律的“碎件”按相互之间的物理联系重新集成原先的整体,以描述原先复杂现象的形成机理和时空变化规律。分解是解耦,是为了便于寻找根源,集成是耦合,是为了回归原来的复杂问题。分解与集成就构成了认识并揭示复杂现象的完整过程。

流域降雨径流的形成是一种复杂的水文现象,它包括产流和汇流两个阶段,其中产流涉及截留、填洼、下渗、蒸散发等基本水文现象;汇流则涉及坡面水汇流、地下水汇流、洪水波运动等基本水文现象。将复杂的流域降雨径流形成过程拆分成一系列基本水文现象就是分解,而将这些基本水文现象进行组装,以模拟流域降雨径流的形成就是集成。探讨基本水文现象形成机理和时空变化规律是产汇流理论^[3-4]研究的任务;将基本水文现象形成机理和时空变化规律进行集成,以模拟流域降雨径流形成是流域水文模型研究的任务。

流域水文模型之所以能集成基本水文现象的形成机理和时空变化规律,达到模拟流域降雨径流形成的目的,靠的是模型结构和参数,即流域水文模型的结构和参数决定了它能否合理、正确地模拟复杂的流域降雨径流形成问题。

流域水文模型的结构有集总式、半分布式和分布式之分^[1]。集总式流域水文模型适用于降雨、下垫面条件、初始状态等空间分布均呈均匀分布的情况。在三者中至少有一呈空间分布不均匀的情况下,虽然可以作均化处理但必定会损失物理意义。半分布式流域水文模型适用于降雨空间分布均匀、下垫面条件和初始状态空间分布不均匀,或下垫面条件和初始状态空间分布均匀、降雨空间分布不均匀的情况。将半分布式流域水文模型用于降雨、下垫面条件和初始状态空间分布均不均匀的情况,虽然可以通过将流域划分为若干个不嵌套的子流域来实现,但仅仅是一种近似处理。分布式流域水文模型适用于降雨、下垫面条件和初始状态空间分布均不均匀的

情况。分布式流域水文模型的构建需要有新的理论和技术来支撑,文献^[5,6]就是这方面的一些尝试。

流域水文模型的结构又有线性、一类非线性和非线性之分。凡依靠倍比原理和叠加原理将时空分布净雨汇集成流域出口断面流量过程的就是线性流域水文模型,其数学表达式为^[4]

$$Q(t) = \int_0^t u(t - \tau) I(\tau) d\tau \quad (1)$$

式中: $I(t)$ 、 $Q(t)$ 分别为流域上分布均匀的净雨过程和相应的流域出口断面流量过程; $u(t)$ 为流域瞬时单位线。

凡倍比原理不适用,但仍可依靠叠加原理将时空分布净雨汇集成流域出口断面流量过程的称为一类非线性流域水文模型,其数学表达式为^[4]

$$Q(t) = \int_0^t I(t - \tau) u' [I(t - \tau), \tau] d\tau \quad (2)$$

式中: $u'(I(t), t)$ 为变动瞬时单位线。

凡既不依靠倍比原理,也不依靠叠加原理将时空分布净雨汇集成流域出口断面流量过程的称为非线性流域水文模型,其数学表达式为^[4]

$$\begin{aligned} Q(t) = & \int_0^t u_1(t, \tau) I(\tau) d\tau + \\ & \int_0^t \int_0^t u_2(t; \tau_1, \tau_2) I(\tau_1) I(\tau_2) d\tau_1 d\tau_2 + \\ & \int_0^t \int_0^t \int_0^t u_3(t; \tau_1, \tau_2, \tau_3) I(\tau_1) I(\tau_2) I(\tau_3) d\tau_1 d\tau_2 d\tau_3 + \dots \end{aligned} \quad (3)$$

式中: u_1 、 u_2 、 u_3 分别为一阶核函数、二阶核函数、三阶核函数。

实践表明,对于大多数较大流域的较大洪水,采用线性流域水文模型一般能得到令人满意的模拟精度^[5]。

分解是为了认识并揭示基本水文模型的形成机理和时空变化规律,集合是为了研制适宜的流域水文模型,两者都十分重要。但现实是大多数流域水文模型都罕见这两者之踪影,好像流域水文模型是一个拍脑袋就可得到的东西,或者只要从国外“搬运”就行了。这是不正常的,不仅有碍水文学的发展,而且有可能将水文学研究引向歧途。

3 正问题和反问题

任何一种现象都是按一定的自然顺序,如时间顺序、空间顺序、因果顺序等发生和变化着。按自然顺序提出的问题称为正问题,而逆自然顺序提出的问题则称为反问题,反问题的求解称为“反演”。在认识并揭示基本水文现象形成机理和时空变化规律时,典型的正问题是“由因推果”,而典型的反问题

是“倒果求因”。在研制适宜的流域水文模型时,典型的正问题是根据已知的降雨时空分布、初始状态、模型的结构和参数等推求流域出口断面流量过程,而典型的反问题是根据降雨时空分布、流域出口断面流量过程、初始状态推求模型结构或参数,或者根据流域出口断面流量过程、模型结构和参数、初始状态等推求降雨时空分布,或者根据降雨时空分布、流域出口断面流量过程、模型结构和参数推求初始状态。若用系统论的术语来理解流域水文模型,则正问题就是由外因和内因求输出,而反问题则是由外因和输出求内因,或者由内因和输出求外因。正问题和反问题虽然只是一字之差,但求解的难度却大相径庭。对于正问题,只要问题提得正确,一般是适定的;而对于反问题,由于求解总是违背现象发生与变化的自然顺序,故一般是不适定的,而且,即使正问题适定,反问题也会不适定;即使正问题是线性的,反问题也会是非线性的。

适定与不适定的概念是法国数学家 Hadamard 于 1923 年提出的^[7-8]。凡解存在、唯一,且具有稳定性的问题是适定的;凡至少不满足这 3 个条件之一的的问题是不适定的。这里的解是广义的,既指问题的解决,也指数学方程的解和反演的待求对象。对于流域水文模型,若模型结构是错误的,或者观测资料是错误的,则反问题的解是不存在的;若收集到的信息不足,导致方程式的数目少于未知量的数目,或者约束条件不完整,或者某一结果由多种原因引起而无法具体确定,无法确定解的性态,则反问题的解虽存在但不唯一,就会产生一个如何从许多解中挑选真解的颇费周折的问题;若观测误差、模型误差、舍入误差等在模型计算过程中所产生的累积作用,使得即使原先是很小的误差也会导致与真解的严重偏离,则反问题的解是不稳定的。因此,可将导致流域水文模型反问题不适定的可能原因归纳为:模型结构或观测资料有错误;收集到的信息、掌握的现象发生原因及约束条件等不足以确定解的性态;不当的计算方法使得误差的累积导致违背物理规律的结果。

从传统的数学眼光看,反问题似乎没有研究价值,因为它往往是不适定的,而有人认为一个不适定的问题是没有意义的。但是反问题在客观世界大量存在,无论是出于经济社会发展的需求,还是出于好奇心,人们都对反问题的求解有着迫切的追求。求解反问题的重要意义在于:已知一个现象的结果或现状希望了解其原因或过去;希望通过直接观测或间接感知到的现象的表面特征了解其本质特征。在流域水文模型研究中,如果回避反问题而不去设法解决,那么其发展就可能停止,其应用精度就无法提高。

正问题需要顺向思维,也就是按现象发生与变化的自然顺序思维。反问题则需要逆向思维,也就是逆现象发生与变化的自然顺序思维。顺、逆两种科学思维在认识世界中是互为补充的,两者综合起来就可以使人们对世界的认识更加全面和深入。

4 解反问题的最优化方法

对于那些解存在但不唯一或不稳定的反问题,怎样才能找到数学上正确或物理上合理的结果呢?苏联数学家吉洪诺夫提出的“正则化方法”就是为了解决这个问题的^[8],其基本思想是:通过对解和观测数据误差的先验估计,设法将问题的解限定在某个较小的范围内,并恰当地改变问题的提法,使原先的不适定问题转换为适定问题。“正则”即正常或基本之意,“正则化方法”就是正常化方法或基本方法。

现在来分析流域水文模型反问题之所以不适定的原因。令模型的数学表达式为

$$y = a + bx \tag{4}$$

式中: x 、 y 分别为模型的输入和输出; a 、 b 均为模型参数。

已有 n 组原型观测资料 (x_1, y_1) 、 (x_2, y_2) 、 $\cdots$ 、 (x_i, y_i) 、 $\cdots$ 、 (x_n, y_n) , 目的是要根据这些观测资料率定模型参数。这是一个典型的反问题。容易理解,将每组观测资料代入式(4)都可得到 1 个以参数为未知量的方程,这样便构成了由 n 个方程组成的方程组:

$$\begin{cases} y_1 = a + bx_1 \\ \vdots \\ y_n = a + bx_n \end{cases} \tag{5}$$

如果模型结构和观测资料均无误差,那么对于只包含 2 个参数的式(4),只需从式(5)中任意挑出 2 个联立求解就可确定其参数,而且无论取其中哪两个联立求解都能得到完全相同的结果,即解存在且唯一、稳定。但如果观测资料组数目小于模型参数数目,也就是信息量不充分,那么即使模型结构和观测资料均无误差,也会因信息量不足而成为不定解,也即解不唯一。这种方程式数目少于未知量数目的方程组属于病态方程之一。如果模型结构和观测资料至少其中之一存在误差,那么从式(5)中任意挑出 2 个方程联立求解就可能无解,即使有解,一般也不会相同,这是另一种病态方程。如果观测资料组数大于未知量数目,那么由于误差的干扰,所形成的方程式数目大于未知量数目的方程组成为无解的矛盾方程组。由此可见,流域水文模型反问题之所以不适定,是因为模型结构和观测资料存在误差,或者所

掌握的有关信息不充分。

再来讨论如何利用正则化思想将流域水文模型反问题由不适定转换为适定的问题。任一输入 x_i , 由原型得输出的观测值为 y_i , 由式(4)得模型输出的计算值为 $y_{c_i}(y_{c_i}=a+bx_i)$, 两者之差即模型结构误差、资料观测误差和数值计算误差所致, 令其为 Δy_i ($i=1,2,\cdots,n$):

$$\Delta y_i = y_i - y_{c_i} = y_i - (a + bx_i) \tag{6}$$

式中: Δy_i 为离差, 它可正可负, 但其平方即离差平方值却总为正值。因此, 对于 n 组观测资料可得离差平方和

$$e^2 = \sum_{i=1}^n (\Delta y_i)^2 = \sum_{i=1}^n \left[y_i - (a + bx_i) \right]^2 \tag{7}$$

e^2 可作为模型模拟精度的一种衡量指标, e^2 越小, 模拟精度越高。 e^2 显然是模型参数的二次函数, 必具有一个极小值。这样就自然会想到, 如果寻求的模型参数只要求使 e^2 极小, 那么流域水文模型参数率定这个反问题就转换为求式(7)的极小值问题了。这是一个易于被接受的常识观点。众所周知, 式(7)取得极小值的充要条件是

$$\begin{cases} \frac{\partial e^2}{\partial a} = 0 \\ \frac{\partial e^2}{\partial b} = 0 \end{cases} \tag{8}$$

将式(7)代入式(8), 可得下列以 a 和 b 为未知量的方程组:

$$\begin{cases} na + b \sum_{i=1}^n x_i - \sum_{i=1}^n y_i = 0 \\ b \sum_{i=1}^n x_i^2 + a \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i = 0 \end{cases} \tag{9}$$

求解式(9)得

$$\begin{cases} a = \frac{1}{n} \left\{ \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \left[\frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \right] \right\} \\ b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \end{cases} \tag{10}$$

式(9)是适定的, 其唯一解就是式(10), 这就表明若利用离差平方和极小的条件确定参数, 则流域水文模型参数率定这一反问题, 通过转换成最优化问题, 就由不适定变成为适定了。由于这个缘故, 最优化方法已成为解决流域水文模型反问题的重要方法之一。

用最优化方法确定流域水文模型参数有间接法

(亦称解析法) 和直接法(亦称数值法)之分。利用由离差平方和达到极小值的充要条件导出的方程组求解模型参数是为间接法。先对待求的模型参数值进行设定, 通过使模型的计算输出与原型的实测输出的离差平方和达到极小值求解模型参数的方法称为直接法。由于流域水文模型一般难以表达成解析式, 故间接法在求解流域水文模型反问题中没有普遍性, 于是直接法就受到青睐。原始的直接法就是试错法, 因其犹如大海捞针, 应用十分不便, 为克服试错法这一缺点, 已提出了名目繁多的直接法^[9-11], 如遗传算法、粒子群算法、蚁群算法等。间接法之所以能将流域水文模型反问题由不适定转换成适定已如上述。应用直接法会有新问题出现吗? 现通过一个实例来回答这个问题。对式(4)表达的模型所对应的原型进行观测, 获得了 10 组观测资料(表 1)。若将这些观测资料代入式(10), 则可得间接法确定的模型参数为 $a=2.33, b=0.67$ 。若用试错法求解, 则可得表 2。表 2 是函数 $e^2=f(a,b)$ 的列表形式。由表 2 可以得到: ①若数值计算精度不高, 则 e^2 的极小值将近似地在一个区域中, 相当于一块“洼地”(表 2 中黑斜体数字所在区域), 在本例中, 这块“洼地”的范围大概为 $a=2 \sim 2.5, b=0.67 \sim 0.72$, 在这个范围内平均的 e^2 约为 3.41。因此, 对于直接法, 若计算精度较低, 则有可能出现不适定。②对任一 a 值或 b 值, 都可以找到相应的 e^2 极小值, 因此, 若直接法缺乏在全局范围内寻优的功能, 则有可能出现不适定。若计算布点较稀疏, 例如 a 和 b 的计算间隔分别取 $\Delta a=1$ 和 $\Delta b=1$, 则 e^2 的极小值为 11.60, 相应的参数值变为 $a=1, b=1$ (表 3)。直接法的这种不适定与前述模型结构误差、观测资料误差和信息不充分导致的不适定不同, 是由于数值计算精度和直接法寻优的功能缺陷而导致的, 因此, 研制适应流域水文模型反问题特点的最优化方法也应引起水文学家的重视。

表 1 数学表达式为式(4)的模型对应的原型观测资料

x	0.5	1.5	3.0	4.0	4.5	6.0	6.0	7.5	9.0	9.5
y	3.0	2.5	5.2	4.5	5.5	6.0	7.0	7.4	8.0	9.0

表 2 不同参数所得之离差平方和

a	e^2							
	$b=-1$	$b=0$	$b=0.5$	$b=0.67$	$b=0.72$	$b=1$	$b=1.3$	$b=2$
-1	1604.81	510.98	185.40	116.22	101.67	41.45	34.90	268.40
0	1435.65	378.61	110.46	58.80	61.62	16.05	42.43	348.00
1	1226.25	272.20	55.60	21.46	15.32	11.60	70.20	470.80
2	1018.58	185.15	22.35	3.88	3.37	21.18	117.81	576.95
2.33	973.03	161.68	13.60	2.89	3.33	37.46	122.85	611.05
2.5	948.50	150.10	9.55	2.95	4.05	40.90	148.55	633.96
3	866.40	118.90	5.56	6.71	9.48	63.15	175.80	696.55

表 3 $\Delta a=1, \Delta b=1$ 时不同参数之离差平方和

a	e^2			
	$b=-1$	$b=0$	$b=1$	$b=2$
-1	1 604. 81	510. 98	41. 45	268. 40
0	1 435. 65	378. 61	16. 05	348. 00
1	1 226. 25	272. 20	11. 60	470. 80
2	1 018. 56	185. 15	21. 18	576. 95
3	866. 40	118. 90	63. 15	696. 55

采用最优化方法解流域水文模型反问题只是为了将反问题由不适定转换为适定,离差平方和最小仅仅是在模型结构已确定的前提下率定参数应满足的条件。模型结构一旦改变,不仅参数会变动,离差平方和的最小值也随之改变。因此,最优参数组相应的离差平方和最小值并不表明模型结构误差是最小的。

5 解反问题的“大数据”方法

“大数据”时代已经到来,它催生了根据密集的观测数据描述与揭示自然规律的科学研究范式。笔者曾在文献[12]中指出: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下的解。虽然迄今为止人们还不能导出这个微分方程,也不知道精确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精确的初始条件和边界条件,但它的解却被人们真真切切地捕捉到了,这就是观测数据。由此,人类是否应当改变一下思维方式,从观测到的现象(数据)入手,也就是从描写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来描写其时空变化,揭示其规律,预测预报其未来状态。“大数据”也许是实现逆向思维,由“果”寻“因”的有效途径之一。建立微分方程原本不是目的,而是揭示现象时空变化规律的一种手段,求得微分方程的解,得到现象的时空变化规律才是目的。“大数据”方法将会成为解决用传统科学思维方式难以解决、甚至无法解决的复杂自然现象规律问题的有效方法。笔者认为,“大数据”方法也是可以将反问题由不适定转换为适定的一种方法。

现有的流域水文模型模拟复杂的流域降雨径流形成都是近似的,有的甚至是粗糙的。但一个流域所积累的降雨、蒸发、流量等观测资料,以及所掌握的下垫面特征和人类活动信息却真实地记录了流域降雨径流形成过程中过去和现在的“因”与“果”的关系。如果一个流域已经积累了 n 场降雨及其所形成的流域出口断面流量过程的完整资料,那么它们实际上是描写该流域降雨径流形成规律的微分方程在不同初始条件和边界条件下的解。 n 越大,所掌

握的流域降雨径流形成的信息就越丰富,未来将发生的情况就越有可能在历史资料中找到。当 n 大到“大数据”级别时,这种可能性就几乎接近必然性。

人类在与自然共处中发现,在完全相同的条件下现象几乎可以重现。这似乎是一个不以人们意志为转移的自然法则。例如:人们就牛顿第二定律做实验,只要实验条件完全相同,结论必然是作用在物体上的合力等于该物体的质量与产生的加速度的乘积。对于流域降雨径流形成,只要降雨、蒸发等气象因子的时空分布完全相同,包括流域形状、大小、冰系分布、地形地貌、土壤植被、地质和水文地质等的下垫面条件完全相同,人类活动影响相同,以及初始条件相同,所形成的流域出口断面流量过程也会相同。这就是利用“大数据”解决流域降雨径流形成模拟反问题的哲学基础。

有人将在相同条件下,流域出口断面流量过程可以重现这一自然法则说成是“水文相似性”,这是不合适的,有误导之嫌。所谓水文相似性是指在相同或成比例的降雨、蒸发、下垫面条件、人类活动影响及初始条件的前提下,两个几何相似的流域所形成的出口断面流量过程之间的关系也成一定比例的性质。水文学家早就意识到,对于流域降雨径流形成现象,所谓水文相似性是不存在的。一个典型的例子是一场时空分布相同的双峰雨,降落在两个形状相同、大小成比例,而且其他影响降雨径流形成的因子也完全相同的流域,其结果将是:对于较小的那个流域将形成双峰的出口断面流量过程,而对于较大的那个流域也许只能形成单峰的出口断面流量过程。这两个相似流域形成的出口断面流量过程在形状上是无“相似”可言的。为什么水文相似性并不存在呢?笔者认为,流域调蓄作用是主要原因。流域调蓄作用产生的原因,一是降落在流域上的雨滴距流域出口断面有远近之分;二是每个雨滴向流域出口断面汇集的速度有大小之分。流域调蓄作用的结果必然是,流域出口断面流量过程比相应的降雨过程峰值减小,峰现时间滞后,形状坦化。由于流域面积大,一般流域调蓄作用也大,因此,对于同样一场双峰雨型,较小流域形成的出口断面流量过程可能仍为双峰,而较大流域也许就变成单峰了。

笔者设想的基于“大数据”方法解流域水文模型反问题的思路如下:首先将该流域历史上发生的所有场次洪水的降雨时空分布及相应的流域出口断面流量过程从积累的观测资料中提取出来;其次通过对流域降雨径流形成机理及影响因素的分析,将每场洪水的下垫面情况、人类活动信息和初始条件提取出来;第三,建立专门数据库,将以上提取所得

的每场洪水的全部信息经数据挖掘、加工编辑后存放于其中。这样,便完成了流域水文模型的“建模”任务,并事实上完成“建模”过程中必须涉及的反问题求解。如果欲解决诸如洪水预报等正问题,那么只要给出面临的一定时空分布的实测暴雨或预报暴雨及相应的下垫面情况、人类活动信息和初始条件就可从数据库中寻找答案。不言而喻,用“大数据”解流域水文模型反问题必然要遇到广泛而深入的计算机技术问题,因此,水文学家与计算机专家的合作是十分重要的。

可以看出,利用“大数据”解流域水文模型反问题,不必纠缠于模型结构的构建和模型参数的率定。事实上,这种表面上看不见具体结构和参数的“流域水文模型”其实已经将它们融合在历史“大数据”之中了。因此,可以避免模型结构不合理和某些最优化方法可能带来的反问题的不适宜。与最优化方法相比,“大数据”方法是一种无需模型结构和参数,就可将流域水文模型反问题由不适宜转换为适宜的方法。

笔者乐观地认为:“大数据”技术在解水文模型反问题和正问题中均有着美好的应用前景!

6 结 语

模型是认识世界的一种重要工具,既从物理、数学角度又从系统角度理解模型,可加深对模型本质的认识。模型总是原型的近似,要求模型必须精准反映原型的所有特征是不可能的,尤其像流域降雨径流形成这样复杂的现象。但是,根据不同目的或精度要求构建相适应的模型也许应成为一项基本原则。为水资源和水环境评价目的而构建的流域水文模型,要求它用于洪水预报也能取得令人满意的精度就不现实。不要刻意去追求流域水文模型的普适性。普适性是从许多个性中归纳和总结出的共性,对于一个具体的流域水文模型,个性一般比共性突出,要获得其普适性是十分困难的。

通过“分解”可精细地描写和揭示各种基本水文现象的发生机理和时空变化规律,这是产汇流理论的研究任务。将揭示所得的各种基本水文现象的机理和规律集成起来用于模拟流域降雨径流形成,属于流域水文模型的研究任务。没有产汇流理论的创新,就不可能有流域水文模型的实质性进步,忽视产汇流理论必将阻碍流域水文模型的发展。

在流域水文模型的研制和应用中涉及许多反问题,其中最为常见和重要的是模型参数的率定。反问题的求解难度远大于正问题,这是因为反问题一般是不适宜的。造成流域水文模型反问题不适宜的

主要原因是模型结构误差,甚至错误,以及观测资料误差。最优化方法是在满足一定精度前提下使反问题由不适宜转换为适宜的一种有效方法。对于最优化方法中的直接法,若不能保证搜索的方向数目与参数数目相同,又不能保证这些搜索方向交汇于唯一点,则也会出现不适宜性。因这种不适宜产生的原因与前述不同,故称为假性不适宜。试图用“大数据”方法将流域水文模型反问题由不适宜转换为适宜,是一种新的思路,尚缺乏实践经验,有待通过进一步研究来验证和完善。

对于流域水文模型,既不可小视其潜在价值,也不能忽视其先天缺陷。

参考文献:

- [1] 芮孝芳.论流域水文模型[J]. 水利水电科技进展, 2017,37(4):1-7. (RUI Xiaofan. Discussion of watershed hydrological mode [J]. Advances in Science and Technology of Water Resources, 2017, 37(4):1-7. (in Chinese))
- [2] 段德智.莱布尼茨哲学研究[M]. 北京:人民出版社, 2011.
- [3] 芮孝芳.水文学原理[M]. 北京:高等教育出版社, 2013.
- [4] 芮孝芳.产汇流理论[M]. 北京:水利电力出版社, 1995.
- [5] 芮孝芳.随机产汇流理论[J]. 水利水电科技进展, 2016,36(5):8-12. (RUI Xiaofang. Random theory of runoff yied and flow concentration [J]. Advances in Science and Technology of Water Resources, 2016, 36(5):8-12. (in Chinese))
- [6] 芮孝芳.单元嵌套网格产汇流理论[J]. 水利水电科技进展,2017,37(2):1-6. (RUI Xiaofang. Theory of runoff yield and confluence by grid inlaid in watershed basic unit [J]. Advances in Science and Technology of Water Resources, 2017,37(2):1-6. (in Chinese))
- [7] 田立平.数学物理方程及其反问题研究[M]. 北京:机械工业出版社,2010.
- [8] 钱志.数学物理反问题的正则化[D]. 兰州:兰州大学, 2008.
- [9] 华尔德 D J,皮特勒 C S.优选法基础[M]. 尤云程,译. 北京:科学出版社,1978.
- [10] 王开荣.最优化方法[M]. 北京:科学出版社,2012.
- [11] 张超.流域水文模型参数自动优化率定及不确定性研究[D]. 南京:河海大学,2010.
- [12] 芮孝芳.水文学与“大数据”[J]. 水利水电科技进展, 2016,36(3):1-4. (RUI Xiaofang. Hydrology and big data [J]. Advances in Science and Technology of Water Resources, 2016,36(3):1-4. (in Chinese))

(收稿日期:2017-11-15 编辑:骆超)