

# 地理学论文的数学模型写作要求

姚鲁烽, 何书金, 赵 歆

(《地理学报》编辑部, 北京 100101)

随着地理学的发展, 越来越多的数学模型在地理学研究论文中被采用。有的论文是通篇围绕一个数学模型的构建与应用来论述, 有的论文则是在其中一个章节中进行模型的应用。但是, 由于有些稿件在介绍数学模型的方法和应用中存在着一些不当之处, 使得论文在写作中出现各种问题。为了帮助论文初写者了解地理学模型的写作格式要求, 我们对近 10 年来《地理学报》来稿评审中发现的主要问题进行了归纳, 总结论文写作中数学模型的选择与构建、参数系数的设置与定义、模型应用的实例与研究、模拟误差的检验与分析、计算结果的图形与检查等方面的格式要求。

## 1 模型的选择和构建

论文中对所用模型的结构、功能和使用方法等一般应进行简要介绍, 说明该模型是什么? 能做什么? 如何保证模型使用的精度? 对数学模型使用一般分为三种情况: 使用前人已建的模型、对前人已建模型进行改进后使用、作者使用自己建立的模型。

### 1.1 对已有模型的选择使用

由于前人对于地理学各学科中数学模型的使用已进行了广泛的研究, 使后来者可以根据自己的研究需要采用前人已确立的模型。在使用前人已建立的数学模型时, 对模型的构建方法和步骤, 可用参考文献标出, 不必做重复论述说明。近年已出版的有关地理学数学模型专著有:

(1) 刘昌明、岳天祥、周成虎主编的《地理学的数学模型与应用》, 科学出版社 2000 年 10 月出版。书中汇集了 1934~1999 年《地理学报》中发表各类数学模型和公式。

(2) 岳天祥主编的《资源环境数学模型手册》, 科学出版社 2003 年 10 月出版。书中汇集了国内外资源环境领域主要期刊发表各类数学模型和公式。

(3) David R. Maidment 主编、张建云等译的《水文手册》, 科学出版社 2002 年 10 月出版。全书总结了 20 世纪 60~90 年代国外发表的有关水文模拟模型和计算公式。

在对各种数学模型和公式选择使用时, 要全面分析各种模型的使用条件、数据要求、精度保证。一般来讲, 各类数学模型都存在着一定的适用范围, 有的受学科限制, 有的受地区影响。所以, 在选择数学模型时必须考虑该数学模型的适用范围。其中, 数学模型应用的区域范围、学科范围和数据条件是检查的重点。数学公式使用前要进行验证, 引入国外的数学公式要先做论证说明。应该用实际观测值与计算值进行比较。

对选用的模型应进行显著性分析, 说明这些关系式的误差水平和适用范围。这是文中计算和讨论的依据, 如果显著性检验通不过, 那么之后的分析均不成立。

1.1.1 根据研究区的地理环境背景选择模型 采用前人建立的数学模型, 要说明前人构建模型的研究区环境背景, 说明本文研究区与原模型使用环境的异同之处。分析引用模型的可行之处, 在正式计算前, 要对公式进行验证计算。至少要与前人在该地区该要素的研究结果进行对比。例如, 在计算河流夹沙力时, 现有的河流夹沙力计算公式很多, 而对于同一水流边界条件来说, 采用不同的夹沙力公式得到的结果相差很大。要考虑研究区的环境背景与该模型以前应用的环境背景的差异。

1.1.2 根据数据的类型和数量条件选择模型 数据的类型与数量对模拟的方法也有影响。例如：在对树木年轮进行拟合时，对于年轮数量少，长度短的样芯，一般用负指数或者直线方法拟合，如果没有明显下降趋势，可以直接用直线拟合。

1.1.3 根据研究要素的形成条件选择模型 对地理要素的特征进行数学分析时要考虑要素形成的影响条件。例如，通过计算黄土粒度分维来反映黄土形成机制，如果黄土粒度分维没有与决定黄土性质的气候、成壤强度相联系，则得出的结论显得空洞，说明不了具体问题。

## 1.2 对已有数学模型的改进

1.2.1 对原有模型系数的改变 在中国引入国际模型计算方法，通常需要对系数做必要的调整。即使在中国构建的模型，在不同区域范围使用，系数都可能有所不同。例如：

计算河流输沙量在不同地区要用不同模型。在黄土高原，地形坡度远大于黄河下游，而河道输沙能力大小，主要取决于河道坡度和来沙量与含沙量。

1.2.2 对原有模型子模块的改变 对模型各子模块的适用性需要进行验证。对于模型的验证一般应该采用一段时间的连续观测资料来进行。在使用中要建立不同的系数系列。例如，降水侵蚀力计算模型在应用于不同气候带时，要根据本地区的降水资料修正其中的参数、系数值。再如，在研究青藏高原使用各类数学模型时，应阐明该方法对于高原特殊环境的适用性和检验效果。

1.2.3 与原有模型的对比说明 对原有模型进行新的数据处理改进时，应该与原方法进行比较分析，并说明新方法比原有方法在数据替代、模拟精度等方面的优越性。而不仅仅是对已有方法进行一种简单的验证。这样才会增大论文的价值。

## 1.3 对新建数学模型需要的说明

由于数学模型具有抽象性，所以有些数学模型可以在不同学科之间进行移植使用。有的论文是在基于一些通用数学模型和其他学科数学模型的基础上，建立新的地理学模型。对建立模型的思路和技术路线可以用框图表示。对于一个可操作的模型，模型的数学描述非常重要。对关键性的模拟不应只有几行文字性的表述，应给予学科专业方面的论证说明。

1.3.1 数学模型的表达规范 一般数学模型的表达形式通常是由一系列的数学公式组成，其中结论性的关键公式通常应列在数学公式组的最前面或最后面，这样可以使读者能够迅速、准确地找到数学模型的基本关系式。

当结论性的关键公式列在数学公式组的最前面时，其后的计算公式依次分别说明关键公式中各项的计算方法；当结论性的关键公式列在数学公式组的最后面时，数学公式组的排列是按照推理过程逐步进行。

### 1.3.2 数学公式的形式和概念表达

一般数学公式或模型由自变量、因变量、常数和等号等部分组成。公式列举前应简要说明模型的基本概念。例如，列举生态足迹的测算模型。生态足迹的计算基于以下两个基本事实：人类可以确定自身消费的绝大多数资源及其所产生废弃物的数量；这些资源和废弃物的量能转换成相应的生态生产性土地面积。其模型主要用来计算在一定的人口和经济技术水平条件下，维持资源消费和废弃物吸收所必需的生态生产性土地面积。具体计算公式为：

$$EF = Nef = N \sum (a_i) = N \sum (C_i / P_i) \quad (1)$$

式中：EF 为总的生态足迹；N 为人口数；ef 为人均生态足迹；i 为消费品的类型 (i = 1, 2, ..., n)； $a_i$  为 i 种消费品人均消费量折算的生态生产性土地面积； $C_i$  为 i 种消费品的人均消费量； $P_i$  为 i 种消费品的全球单位面积平均生产能力。

1.3.3 相关关系的检查 数学模型中要分析因变量与自变量的关系是正相关还是负相关。由此确定相关因子在数学公式中是处于分子还是分母的位置。例如,在一定地域范围内,土壤侵蚀量基本与植被覆盖率呈负相关关系。因此,在构建土壤侵蚀模型时,植被覆盖率应该放在分母的位置。也有些要素不是简单地正负相关关系,例如,土壤 pH 值与土壤肥力关系,不是越大越好或越小越好,而是适中为好。这种相关分布关系在模型构建中应特别注意。对数学模型中各项要素相关关系的定性检查,是保证数学模型不出逻辑性错误的重要方法之一。

1.3.4 主次关系的检查 地理学研究的要素往往是受到多种因素的影响,这些要素有些是起决定性的全局影响,有些则是起着次要性的局部影响。在构建模型中应该充分考虑到。例如,在建立计算蝗虫分布的数学模型中,由于蝗虫一般不能在海拔 4500 m 以上生存。因此,数学公式中应建立其他要素与海拔高度相乘的关系。应定义当高程参数达到 4500 m 时,蝗虫数量为零。

1.3.5 数学方程的量纲检查 数学方程式左右两端的计量单位应一致。就量纲分析而言,如果方程左式的单位和方程右式不同,说明该模型仍旧是统计意义上的,其物理意义模糊。

## 2 数学模型或公式中参数系数的设置与定义

### 2.1 参数类型的设置

自然界任何一种要素的形成和变化,都会受到其他多种要素的影响与控制。因此,在模型建立中要全面考虑各影响要素的设置。例如,植物的生长与温度、降水、海拔高度、土壤质地、地下水埋深等要素的影响。这些要素就构成了有关植物生长模型的计算参数。在有关植被恢复的数学模型就应当有温度、降水、海拔、土质、地下水位等参数,因为这些要素是植被恢复生长的重要条件。

### 2.2 参数符号的使用要求

在数学模型中,各项参数是其组成的基本要素。参数符号的表达形式和定义说明应符合科技论文写作的基本格式。

2.2.1 参数符号的表达形式 根据国家出版标准和科技论文的表达惯例,数学公式中参数符号通常采用参数的英文首字母,变量符号必须用斜体表示。参数属性在参数符号右下角用英文字母下标表示,一般为参数类型的英文首字母。

例如: $I_w$  为水污染指数, ( $I$  为 Index 的首字母,下角标  $w$  为指数类型 water 的首字母。)  $E_s$  为土壤蒸发量, ( $E$  为 Evaporate 的首字母,下角标  $s$  为蒸发类型 soil 的首字母。)

以上二个参数符号如果分别用 WINDEX、SOILE 表示,就显得不够简明。

下标符号表达的原则为:表示变量符号的下标用斜体;其他下标用正体。例如,正体下标:  $C_g$  ( $g$  为气体 gas)、 $G_n$  ( $n$  为标准 normal)、 $G_{\max}$  ( $\max$  为最大值)等;斜体下标:  $C_p$  ( $p$  为压力)等。一些约定的常数,例如:  $Re$  (雷诺数),  $Fo$  (傅里叶数)等及动植物等的属以下(含属)的拉丁文名称(如:  $E. coli$ ,  $Oryza$ )也需要使用斜体形式。有关矢量、张量和矩阵等符号,需要采用黑斜体形式。

除上述需要用斜体表示的变量符号外,其他表示非变量的符号需用正体表示,主要有: 计量单位和词头符号,如  $s$  (秒)、 $L$  (升)、 $Pa$  (帕)、 $k$  (千)、 $m$  (毫)、 $M$  (兆)等;

缩写符号或特殊函数符号,如  $\min$  (最小)、 $\lim$  (极限)、 $\text{Im}$  (虚部)、 $T$  (转置)、 $\exp$ 、 $\ln$ 、 $\sin$ 、 $\cos$ 、 $\pi$  (圆周率)、 $e$  (自然对数的底)、 $i$  (虚数单位)等。 化学元素符号、层壳、电子态、电子轨道,如  $s$ 、 $t$ 、 $S$ 、 $f$  等。某些特殊的集符号需要用空心正体形式。

2.2.2 参数符号的定义要求 在数学模型中使用参数符号时要注意以下几点:

(1) 参数定义在全文中应是统一和唯一的。例如: 在同一论文中不能在一处将  $r$  定义为耗水率, 在另一处又将  $r$  定义为回归系数; 也不能在同一论文中一处将  $r$  定义为耗水率, 在另一处又将  $p$  定义为耗水率。

(2) 参数的定义不能自我定义。例如, 式中: GDP 为流域 GDP。

(3) 参数符号在数学公式中与文字中要一致。

(4) 对参数的说明要准确。例如, “\*\*\* 降水参数评估会对 \*\*\* 模型径流和侵蚀产生影响”, 说法不对。应为: “\*\*\* 降水参数评估会对 WEPP 模型模拟径流和侵蚀的精度产生影响”。

2.2.3 参数符号的定义顺序 参数符号的定义应紧接在列举的数学模型后, 应当按照自左而右、自上而下的顺序依次定义数学公式中所有的参数和量纲(计量单位)。这样做可以使作者不易漏掉需要说明的参数, 也使读者能很快理解数学公式的所有参数含义。

例如, 冰川作用的径流模拟计算公式为:

$$R = \frac{T}{L_m}(H_n + H_s) - \frac{L_e}{L_m}E + P_L \quad (5)$$

式中:  $R$  为径流量 (mm),  $T$  为计算时段,  $L_m$  为冰雪融化潜热,  $H_n$  为净辐射 ( $W/m^2$ ),  $H_s$  为感热通量 ( $W/m^2$ ),  $L_e$  为蒸发潜热,  $E$  为蒸发量 (mm),  $P_L$  为液态降水 (mm)。

### 2.3 参数确定的基本方式

模型中参数的确定通常有三种方式: (1)直接定值、(2)相应赋值、(3)计算取值。在论文中应分别说明各项参数的确定方式。

2.3.1 模型参数的直接定值 模型的计算值依据气温、降水量、高程、人口、产值、土地面积等定量数据直接确定。例如, 式中:  $T$  为年均温 ( $^{\circ}C$ ),  $P$  为年降水量 (mm) 等。

2.3.2 模型参数的相应赋值 有的模型参数赋值方法是依据该参数的特性和重要性确定的。例如, 在评价模型中, 各要素的权重值通常是依据该要素的影响程度来确定的, 对于数学公式中评价指标权重选取的原则与步骤应说明, 并且最好做出相关的灵敏度分析。

也有的参数量值根据某种数据进行分段赋值, 这种情况一般可通过列表说明, 例如: 海拔范围、化学离子浓度范围。也可依据要素的类型进行分类赋值, 例如, 根据土壤类型或植被类型赋以不同的数值。

2.3.3 模型参数的计算取值 有的模型参数的量值不是直接由观测或统计数据直接输入, 而是通过数学公式计算出来。例如, 通过建立指数来进行定量取值。在这种情况下, 要说明各种指数的物理意义。各种参数计算公式往往又构成了数学模型的子模块。

模型中参数的算法如果是自创的, 要说明设定的依据, 给出实验依据, 说明这些参数设置的不同对于模拟结果影响有多大。如果是引用前人的, 则要标注文献出处。

### 2.4 系数的检查

有的计算公式不是理论公式, 属于经验公式, 其参数需要根据不同的实例来进行率定, 不能任意套用。在方程列举中, 应该提供相应的参数以供参考。一般可对相关数据进行回归分析, 确定计算公式的系数。

## 3 数学模型的应用与检查

论证数学模型在地理研究中应用的可行性, 不仅要通过科学的逻辑判断和正确的数学公式推导, 还要在具体研究区中通过实例计算来验证模型的有效性与合理性。应用举例中要定量阐述研究要素的分布、变化特征及其原因。应用实例举证应介绍具体的计算步骤, 以供读者借鉴。

### 3.1 模型应用的数据要求

有的作者将精力集中于程序编制,系统设计很先进。但由于缺少应用充足的数据进行实际计算检验,使得模型的可信性受到影响。应利用模型输入和处理大量数据,进行具体实例计算分析,以证明模型的可靠性和先进性。

**3.1.1 模型应用中的数据样本说明** 在模型应用中要明确说明采集数据样本的起讫时间、时段长度、研究范围、资料数量、数据类型。例如,“将 1955~2005 年(起讫时间)50 年间(时段长度)黄河中游(研究范围)180 处气象站(资料数量)的年降水量(数据类型)数据输入模型中”。

对各类数据的出处应说明,如果数据是来自统计年鉴、前人文献或相关网站时,要用参考文献标注。

**3.1.2 模型应用中的数据匹配要求** 模型中使用的各类相关数据在时间尺度和地域空间上应具有有一致性,这样才能保证地理要素空间变化和相关分析的结果是可靠的。

输入模型的资料要尽可能地包括近几年的最新数据。例如,有关区域经济的模拟,无论模拟模型建立的如何可靠,如果没有近年的经济数据输入,就难以真正模拟出未来的经济趋势。

**3.1.3 模拟数据的时间长度保证** 模拟数据要有一定的时间长度,才能对变化周期、未来预测等模拟有可靠的结果。例如,中国农牧交错地带是年降水变率较大的地区,进行水文模拟要用多年的降水径流数据,否则模型的可靠性不足。再如:对于人工林的模拟最好有 20 年以上的资料,否则可能存在问题。对于模型的验证一般应该采用一段时间的连续观测资料来进行验证。如果输入的数据过旧或太少,仍然不能得出很好地模拟出研究要素的未来演化趋势。

**3.1.4 模拟数据的空间范围保证** 为保证模型应用的区域兼容性和可推广性,对模型应用的选择范围不宜太小。过小的研究范围难以涵盖多种要素类型,也难以显示模型的应用意义。

**3.1.5 对拟合统计检验结果说明** 为检验数学模型的可靠性,需要将各类因子的实际数据代入公式中进行实际运算,并将模拟的计算结果与实测值进行对比,确定误差范围。例如,计算区域生态需水量,需要考虑降水、径流、蒸发、生产用水和生活耗水等要素。

### 3.2 模拟精度的误差检查

应当依据同一时间、同一地点的实测数据对模拟计算值进行误差检验。如模拟结果的误差很大,是模型结构的问题还是其他问题?如何能够提高模型的模拟精度?应进行深入研究,并给出有建设性的改进意见。

模型估算的验证涉及到时空尺度一致的问题,对验证过程应交待清楚,可以和别的模拟数据进行一定的比较。在提出新的测量技术时,应通过比测来确定该新方法给出的结果的精度。既在同样的条件下,既用公认的传统方法,也用新方法来进行对比测量。然后进行误差分析,从而确定新方法的精度。并对此进行讨论和精度评价。

**3.2.1 模型的方差解释量和 F 检验** 模型在进行方差解释量和 F 检验时要说明:如果是 F 检验, F 检验反映的是方程的稳定性与总体线性显著性,而方差解释量反映的是方程的拟合程度。一般提法应为“模型在 \*\* (例如 0.001) 的 F 检验显著水平下,方差解释量达到 \*\* (例如 90.9%)”。

**3.2.2 模拟精度与实际情况变化幅度对比** 要考虑模型的模拟精度与实际情况变化幅度。例如:作物模型的模拟结果为:生育期误差在 10%左右,产量误差在 15%左右;而气候变化对冬小麦潜在产量的可能影响幅度为 1%~3%。作物模型模拟精度远高于气候变化影响幅度,则结论的可靠性值得推敲。

3.2.3 模型的置信水平检验 模拟结论的论据要充分。例如,在164个气象台站中,如果冬半年降水总量只有15个站有明显变化趋势的置信水平超过95%;降水日数有18个站;日降水强度有19个站;强降水日数有15个站,只占总气象站的9%~12%。据此认为区域降水有怎样的变化特征,缺乏必要的论据。

### 3.3 模拟结果的图形检查

将模拟结果绘制成相应的曲线图或分布图,可以直观地分析研究要素的时空变化过程,找出相关规律、发现其中问题。

3.3.1 回归分析曲线图的检验 重点检查数值点的数量、种类、回归曲线与数值点的分布关系。在回归模拟图中,如有不少点子偏离甚大,应对其原因进行解释。

3.3.2 空间分布变化图的检验 分析空间分布变化趋势与实际基本情况是否相符。例如,在进行华北平原高程趋势面分析时,如出现东高西低的情况,就表明模拟计算中存在问题。

3.3.3 正负变幅柱状图的检验 有些模型的计算结果可以利用数据的性质来判断正误,特别是计算一些单一流向的矢量数值。例如,计算降雨的时空分配特征,得出的降雨数值应当为正值,因为如果计算的数值是负值,就不是降雨值,而是蒸发值。再如,对于山区径流的模拟也是如此,河道断面通过的径流量应该为正值,如果径流计算结果为负值就意味着出现山区水流在逆流而上,其结果是不合理的。

### 3.4 模拟结果的学术分析

对模拟结果的地理规律分析 对地理模型使用结果的说明,不仅要说明模型的可靠性、有效性,更重要的是应说明模型所揭示的研究区各要素的时空变化规律和相互影响程度。对模拟结果要有独立的思考,对数据处理流程的理解和实际结果要有到位的思考。不仅有纯数据的分析,而且要有地理过程方面的分析与思考。

由于有些作者是非地理专业的学者,对地理专业的知识相对不熟。所以在利用数学模型研究各类地理要素时,要查阅相关的专业书,以免出现与学科常识不符的错误结论。

3.4.1 对模拟结果的认识深度说明 利用数学模型分析应当得出与过去不同的认识,根据计算结果要看出与过去的不同认识,得到研究要素的实际相关特征。结论部分应该对本模型的特点和功能进行归纳和总结,点出创新之处和有待改善的地方。结论部分一方面是阐明模型的应用效果,另一方面尽量用定量化的语言对专业成果进行描述。

3.4.2 对偏离趋势的特殊样本解释 对数值进行相关关系的回归模拟时,既要多数样本的数值分布基本趋势进行分析,也要对少数偏差较大的样本进行解释,说明其出现偏差的环境背景和特殊原因。

3.4.3 对模拟结果的分级阈值说明 在用数学模型确定分级分类的阈值时,由于阈值通常是有有一定变化幅度的。所以要说明阈值的变化范围、影响因素。

## 4 总结

对数学模型的写作提出规范性要求,有利于作者了解数学模型写作的基本格式,减少和避免论文表达中的错误。也有助于读者在阅读文献时能够完整、准确、方便地掌握相关数学模型的学术概念和使用方法。

本文总结的只是《地理学报》来稿中发现的一些常见问题,提出这些问题的都是地理学各专业的审稿专家。我们对以往积累的审稿意见进行总结,主要目的是归纳写作要点、规范论文格式,使更多的学术成果能够尽快发表。