

DOI:10.3969/j.issn.1674-8131.2023.03.007

# 灰色组合预测模型优化及科技人才需求预测

王晓颖<sup>1</sup>, 苟小义<sup>2</sup>, 曾波<sup>3</sup>

(1. 重庆财经学院 新媒体艺术学院, 重庆 401320; 2. 南京航空航天大学 经济与管理学院, 江苏 南京 211106; 3. 重庆工商大学 管理科学与工程学院, 重庆 400067)

**摘要:**科技人才需求预测是国家合理制订人才政策的重要依据。为此,本文基于科技人才需求的数据特征,构建适用于科技人才需求预测的新型离散灰色模型 FODGM( $r, 1, k^{\theta}, u$ ),该模型实现了系统发展灰信息非线性规律的较好反映以及累加阶数作用范围全实域拓展,缓解了原始序列中极值对模型性能的影响,能够有效模拟科技人才需求的发展趋势与演变规律。应用该模型对我国科技人才需求量进行预测,结果显示未来我国科技人才需求量呈逐步上升趋势,预计 2026 年我国科技人员全时当量将达 729.258 万人年,科技人才需求端压力较大。相关部门可以根据预测结果制定缓解我国科技人才需求端压力的对策。

**关键词:**科技人才需求预测;离散灰色模型;模型结构;非线性修正项;参数组合优化

**中图分类号:**F272.1;C964.2;N941.5 **文献标志码:**A **文章编号:**1674-8131(2023)03-0094-14

**引用格式:**王晓颖,苟小义,曾波.灰色组合预测模型优化及科技人才需求预测[J].西部论坛,2023,33(3):94-107.

WANG Xiao-ying, GOU Xiao-yi, ZENG Bo. Optimization of grey combination forecasting model and forecasting the demand for scientific and technological talents[J]. West Forum, 2023, 33(3):94-107.

## 一、引言

科技人才是一种广义的,随时代要求变化而改变的动态概念<sup>[1]</sup>,其常指拥有科学知识与技术、具有

\* 收稿日期:2023-02-23;修回日期:2023-04-28

**基金项目:**重庆市技术预见与制度创新专项项目(cstc2021jsyj-zdxwtAX0009, CSTB2022TFII-OIX0051);重庆市教委教育教学改革研究项目(222178);重庆市教委科学技术研究重点项目(KJZD-K202202102);重庆市研究生科研创新项目“数字经济和实体经济融合对成渝双城地区制造业绿色创新的影响研究”

**作者简介:**王晓颖(1977),女,重庆渝中区人;副教授,博士,主要从事管理学研究;Tel:023-88968667, E-mail: xiaoyingwang@hotmail.com。苟小义(1996),女,四川内江人;博士研究生,主要从事灰色预测建模研究;Tel:18328074625, E-mail: gxy18328074625@163.com。曾波(1975),男,四川内江人;教授,博士,主要从事预测理论与方法的研究;Tel:023-62769347, E-mail: 15922951705@163.com。

科学素养和能力及道德的人,是经济社会发展的重要资源与基础保障<sup>[2]</sup>。科技人才需求预测是人才培养、政策制定等的科学依据与重要基础。科技人才需求的预测方法主要有神经网络模型<sup>[3]</sup>、回归模型<sup>[4-5]</sup>以及灰色预测模型<sup>[6]</sup>等。然而,科技人才需求量具有非线性发展特征<sup>[7]</sup>,并受多种难以定量描述因素的影响<sup>[8]</sup>。因此,研究者常用GM(1,1)模型来预测科技人才需求量<sup>[9]</sup>,然而,GM(1,1)模型将灰作用量 $b$ 视为影响系统发展的所有灰信息<sup>[10]</sup>,导致建模过程中忽略了部分原始数据的非线性信息。

灰色系统理论是由华中工学院邓聚龙教授在1982年提出,主要解决“小样本,不确定性”系统的分析、预测、决策与控制问题<sup>[11]</sup>。灰色预测是灰色系统理论的一个重要分支,GM(1,1)模型因为其简单的模型构建和高效的预测能力获得了众多学者的青睐。然而,GM(1,1)模型本身也存在一些缺陷,例如对于原始数据要求高,模型预测结果不稳定等。在现实生活中,数据呈现复杂多变的情况,在GM(1,1)模型的基础上需要提高其模拟精度与适用范围。为此,学者们进行了GM(1,1)模型结构拓展的研究,取得了一系列的学术成果。其中,谢乃明和刘思峰提出了离散灰色预测模型DGM(1,1)<sup>[12]</sup>,并研究了其与传统GM(1,1)模型的关系。崔杰等针对近似非齐次指数特征序列构建了NGM(1,1,k)模型,进一步拓展了GM(1,1)模型的适用范围<sup>[13]</sup>。Zeng将线性修正项引入传统GM(1,1)模型,提出一种三参数离散灰色预测模型TDGM(1,1),显著提高了其建模能力和适用性<sup>[14]</sup>。虽然TDGM(1,1)模型较传统模型有了一定的改进,但是由于忽略了项数 $k$ 的非线性关系,在处理非线性原始序列时表现仍然较差。为了进一步提高模型的预测精度,Zeng在TDGM(1,1)的基础上增加了非线性修正项,发展出一种四参数离散灰色模型FDGM(1,1,k2,r)<sup>[15]</sup>,该模型对非线性原始序列具有更好的模拟性能。

这些拓展与改进不仅提高了GM(1,1)模型的预测精度和适用范围,还为灰色预测模型的发展提供了新的思路,但仍然存在一些改进的空间。为此,本文从科技人才需求量的演变特征出发,对FDGM(1,1,k<sup>2</sup>)模型的缺陷进行弥补,构建新型优化离散灰色预测模型FODGM( $r, 1, k^{\theta}, u$ )。新模型结构由线性、非线性作用项以及随机扰动项构成,由此反映影响系统发展的所有灰信息,实现了系统发展灰信息非线性规律的较好反映以及累加阶数作用范围全实域的拓展,缓解了原始序列中极值对建模的影响,可以更加全面、更加精准地进行预测。应用新模型对我国科技人才需求进行预测,并与其他几个预测模型的结果进行比较,验证了其预测的有效性,可以提供更加准确的结果,以帮助政府更好地了解和应对科技人才需求的变化。同时也验证了该预测模型具备科学性和实用性,为不同类型的预测系统提供了一种有效的方法,并对于灰色预测系统的发展具有积极的意义。

## 二、现有FDGM(1,1,k2)模型及其缺陷分析

在TDGM(1,1)模型基础上,将影响系统发展的灰信息由原有线性函数( $bk+c$ )拓展为非线性函数( $bk+c+dk^2$ ),并将原有的1阶拓展为 $r$ 阶( $r \in \mathbf{R}$ ),一种具有动态自适应性的灰色预测模型FDGM(1,1,k<sup>2</sup>)被提出<sup>[15]</sup>。本文首先介绍FDGM(1,1,k<sup>2</sup>)模型,再分析其存在的不足,旨在为进一步拓展模型的适应性及提高处理复杂序列的模拟能力奠定基础。

定义1:设 $X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)]$ 为建模序列,阶数 $r \in \mathbf{R}$ , $X_R^{(r)}$ 为新序列,且 $X_R^{(r)} = (x_R^{(r)}(1), x_R^{(r)}(2), \dots, x_R^{(r)}(n))$ <sup>[16]</sup>。其中:

$$x_R^{(r)}(k) = \sum_{i=1}^k \frac{\Gamma(r+k-i)}{\Gamma(k-i+1)\Gamma(r)} x_R^{(0)}(i), k = 1, 2, \dots, n \quad (1)$$

称公式(1)为序列 $X^{(0)}$ 的 $r$ 阶实数域灰色生成算子,简记为 $r$ -RGO( $r$ -order Real number field Generation Operator);称 $X_R^{(r)}$ 为原始序列 $X^{(0)}$ 的 $r$ -RGO生成新序列。定义1中阶数 $r$ 的正负性可用于判

定实数域灰色生成算子的功能。特别地, 当  $r < 0$ , 表示实数域灰色生成算子具备获取差异信息的新功能。

定义 2: 设  $X^{(0)}$  与  $X_R^{(r)}$  如定义 1 所示,  $Z^{(r)} = [z^{(r)}(2), z^{(r)}(3), \dots, z^{(r)}(n)]$  为  $X_R^{(r)}$  的紧邻均值生成序列<sup>[15]</sup>, 其中  $z^{(r)}(k) = 0.5 \times [x_R^{(r)}(k) + x_R^{(r)}(k-1)]$ ,  $k = 2, 3 \dots, n$ , 则有:

$$x_R^{(r-1)}(k) + az^{(r)}(k) = kb + c + k^2d \quad (2)$$

公式 (2) 为面向实数域灰色生成算子的四参数离散灰色预测模型 (Four-parameter discrete grey forecasting model), 简称 FDGM(1, 1,  $k^2$ ) 模型。该模型存在以下两方面的缺陷:

一是结构缺陷。FDGM(1, 1,  $k^2$ ) 模型中作用项  $dk^2$  是构成非线性灰信息的重要组成部分, 也是构造该模型参数矩阵不可或缺的部分, 对模型参数的大小以及模拟 (预测) 结果都至关重要。然而, 作用项  $dk^2$  的指数固化为 2, 对特定的序列表现良好, 但是固化的结构限制了模型性能和作用空间。该作用项指数无法依据原始序列的数据特征进行匹配, 缺乏动态可调性, 从而较大程度影响到模型性能。

二是参数缺陷。背景值系数是灰色预测模型的重要性能参数, 对构建模型及模型性能的发挥具有重要作用。然而, FDGM(1, 1,  $k^2$ ) 模型的背景值系数简化为 0.5, 这种简化处理缺乏针对具体特征序列的动态可变性, 模型的适应性无法体现, 故难以达到最佳的模拟效果。此外, 虽然 FDGM(1, 1,  $k^2$ ) 模型进行了单参数优化, 但参数组合优化更能体现全局性、整体性的优化作用, 因而该模型在参数优化方面仍有较大的改善空间。

### 三、一个新型灰色离散预测模型: FODGM( $r, 1, k^\theta, u$ ) 模型

自改革开放以来, 我国的科技人才全时当量呈现出快速增长的趋势。根据国家统计局发布的数据, 2000 年全国研究与开发人员全时当量为 92.2 万人年, 到 2010 年达到 255.4 万人年, 十年间增长了 277.01%。然而, 近年来我国 R&D 人员全时当量增速逐渐放缓, 从 2010 年到 2020 年增长了 198.27%, 增速明显下降。鉴于科技人才数量增长呈现出较为明显的非线性趋势, 传统的 GM(1, 1) 模型在建模时可能无法准确捕捉到这种非线性趋势。因此, 为了提高预测的精确度, 本文提出一种新型灰色离散预测模型——FODGM( $r, 1, k^\theta, u$ ) 模型。该模型在 FDGM(1, 1,  $k^2$ ) 模型的基础上, 为提高其作用范围及自适应性, 分别使用  $\theta$  ( $\theta \in \mathbf{R}$ ) 和  $u$  ( $u \in [0, 1]$ ) 替换 FDGM(1, 1,  $k^2$ ) 模型作用量  $k^2d$  的指数和背景值系数, 同时考虑指数  $\theta$ 、背景值系数  $u$ 、阶数  $r$  ( $r \in \mathbf{R}$ ) 的组合优化, 以期能动态匹配日益复杂的非线性序列。基于上述思路, 本文提出含待定指数作用项的参数组合优化离散灰色预测模型, 其构建及推导过程如下。

#### 1. FODGM( $r, 1, k^\theta, u$ ) 模型定义

定义 3: 设  $X^{(0)}$  与  $X_R^{(0)}$  如定义 1 所示,  $k^\theta d$  为含指数  $\theta$  ( $\theta \in \mathbf{R}$ ) 的非线性作用项,  $u$  ( $u \in [0, 1]$ ) 为背景值系数, 其中  $k = 2, 3 \dots, n$ , 则有:

$$x_R^{(r-1)}(k) + a[ux_R^{(r)}(k) + (1-u)x_R^{(r)}(k-1)] = bk + c + dk^\theta \quad (3)$$

公式 (3) 为面向实数域灰色生成算子含待定指数作用项的四参数优化离散灰色预测模型 (Four-parameter optimization discrete grey forecasting model), 简称 FODGM( $r, 1, k^\theta, u$ ) 模型。其中,  $ux_R^{(r)}(k) + (1-u)x_R^{(r)}(k-1)$  为含待定参数  $u$  的优化背景值。

FODGM( $r, 1, k^\theta, u$ ) 模型通过指数  $\theta$  的引入弥补了 FDGM(1, 1,  $k^2$ ) 模型的结构缺陷, 避免了固化的模型结构局限模型性能与作用空间。此外, 背景值系数  $u$  的设计弥补了 FDGM(1, 1,  $k^2$ ) 模型的参数缺陷, 提升了模型处理具体特征序列的动态可变性与模型的适应性。FODGM( $r, 1, k^\theta, u$ ) 模型较 FDGM(1, 1,  $k^2$ ) 模型的改进路线如图 1 所示。

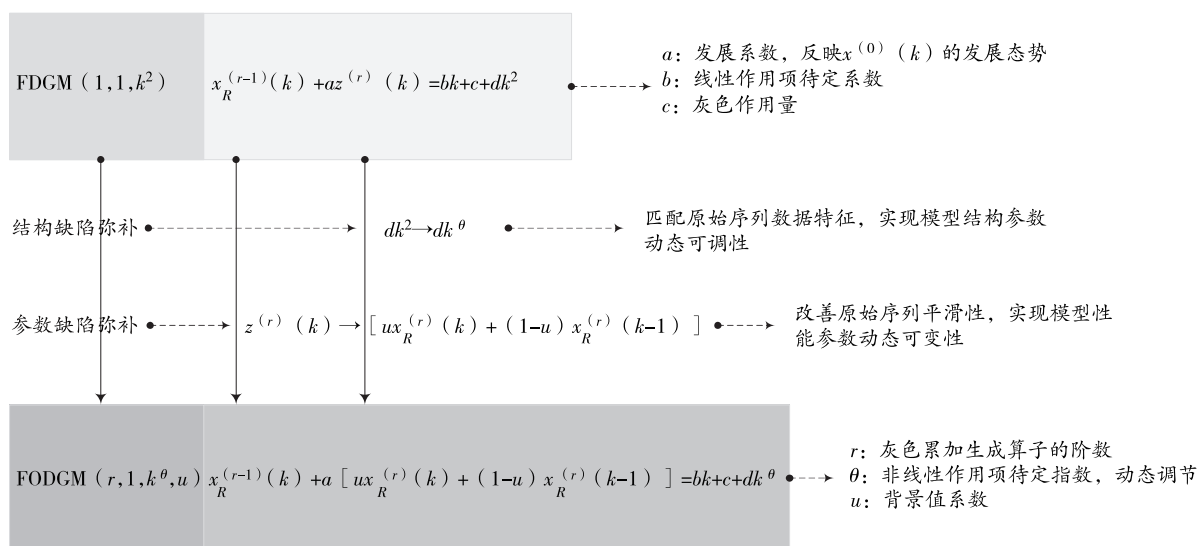


图 1 FODGM( $r, 1, k^\theta, u$ ) 模型改进路线图

单变量灰色系统中, 变量描述了系统的演化规律, 是系统在诸多复杂外部因素共同作用下的结果。系统发展的影响因素是“因”, 系统所体现出的变化结果是“果”。在控制论中, 前者称为输入, 后者称为输出<sup>[17]</sup>。对 FODGM( $r, 1, k^\theta, u$ ) 模型而言, 线性作用项  $bk$ 、非线性作用项  $k^\theta d$  及灰色作用量  $c$  组成的函数构成了输入, 代表了影响系统发展的所有灰信息。不同作用量分别表征了系统发展的线性、非线性及随机性, 特别地, 非线性作用项  $k^\theta d$  含待定指数  $\theta$ , 能动态调节系统输入。输出变量为  $x^{(0)}(k)$ , 其通过  $r$ -AGO 及  $u$ -MEAN 来动态调节灰因。其中,  $r$ -AGO 为实数域灰色生成处理, 能弱化随机性;  $u$ -MEAN 为含待定系数的紧邻生成处理, 能改善平滑性。系统输入、输出之间的关系如图 2 所示, 根据其关系可得 FODGM( $r, 1, k^\theta, u$ ) 模型的定义。特别地, 公式 (3) 含三大待定系数 ( $\theta, u, r$ ), 使用参数组合优化思路求解, 达到根据原始数据特征动态匹配从而得到全局最优模型的目的。

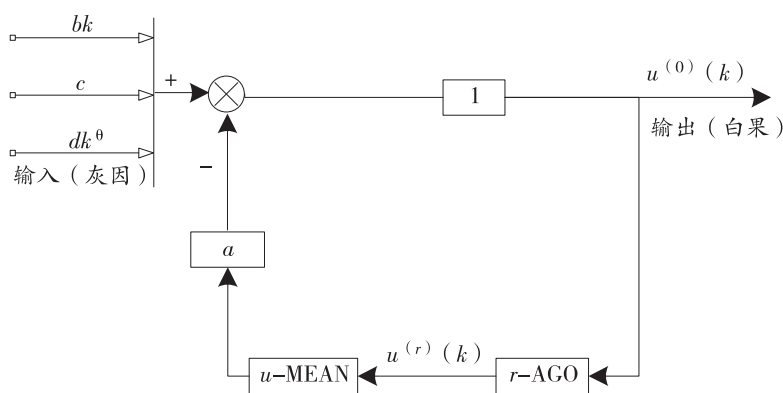


图 2 FODGM( $r, 1, k^\theta, u$ ) 模型的输入输出框图

## 2. FODGM( $r, 1, k^\theta, u$ ) 模型参数估计

定理 1: 设序列  $X^{(0)}, X_R^{(r)}$  如定义 1 所示,  $dk^\theta$  和  $ux_R^{(r)}(k) + (1-u)x_R^{(r)}(k-1)$  如定义 3 所示,  $\hat{p} = (a, b, c, d)^T$  为参数列, 参数列可以运用最小二乘法估计, 满足:  $\hat{p} = (a, b, c, d)^T = (B^T B)^{-1} B^T Y$ 。其中:

$$Y = \begin{bmatrix} x_R^{(r-1)}(2) \\ x_R^{(r-1)}(3) \\ \vdots \\ x_R^{(r-1)}(n) \end{bmatrix} = \begin{bmatrix} x_R^{(r)}(2) - x_R^{(r)}(1) \\ x_R^{(r)}(3) - x_R^{(r)}(2) \\ \vdots \\ x_R^{(r)}(n) - x_R^{(r)}(n-1) \end{bmatrix}; B = \begin{bmatrix} -ux_R^{(r)}(2) - (1-u)x_R^{(r)}(1) & 2 & 1 & 2^\theta \\ -ux_R^{(r)}(3) - (1-u)x_R^{(r)}(2) & 3 & 1 & 3^\theta \\ \vdots & \vdots & \vdots & \vdots \\ -ux_R^{(r)}(n) - (1-u)x_R^{(r)}(n-1) & n & 1 & n^\theta \end{bmatrix}$$

证明过程与参考文献[9]类似, 此处省略相关证明过程。

### 3. FODGM( $r, 1, k^2, u$ ) 模型求解

定理 2: 设  $B, Y, \hat{p}$  如定理 1 所示,  $\hat{p} = (a, b, c, d)^T = (B^T B)^{-1} B^T Y$ , 则有:

(1) FODGM( $r, 1, k^\theta, u$ ) 模型的时间响应式如公式(4)。

$$x_R^{(r)}(t) = u_1^{(t-1)} x_R^{(r)}(1) + \sum_{g=0}^{t-2} [(t-g)u_2 + u_3 + (t-g)^\theta u_4] u_1^g \quad (4)$$

$$\text{其中, } u_1 = \frac{1-a+au}{1+au}, u_2 = \frac{b}{1+au}, u_3 = \frac{c}{1+au}, u_4 = \frac{d}{1+au}$$

(2) FODGM( $r, 1, k^\theta, u$ ) 模型的累减还原式如公式(5)。

$$x_R^{(0)}(1) = x^{(0)}(1);$$

$$x_R^{(0)}(k) = (x_R^{(r)}(k))^{(-r)} = \sum_{i=1}^k \frac{\Gamma(-r+k-i)}{\Gamma(k-i+1)\Gamma(-r)} x_R^{(r)}(i), k=2, 3, \dots, n \quad (5)$$

证明(1): 根据定义 1, 有公式(6)。

$$x_R^{(r-1)}(k) = x_R^{(r)}(k) - x_R^{(r)}(k-1), k=2, 3, \dots, n \quad (6)$$

当  $k=2, 3, \dots, n$ , 将公式(6)代入公式(3), 可得公式(7)。

$$x_R^{(r)}(k) - x_R^{(r)}(k-1) + a(ux_R^{(r)}(k) + (1-u)x_R^{(r)}(k-1)) = kb + c + k^\theta d \quad (7)$$

整理得公式(8)。

$$x_R^{(r)}(k) = \frac{1-a+au}{1+au} x_R^{(r)}(k-1) + \frac{b}{1+au} k + \frac{c}{1+au} + \frac{d}{1+au} k^\theta \quad (8)$$

在公式(8)中, 参数  $a, b, c, d$  可根据定理 1 求解, 但  $x_R^{(r)}(k-1)$  未知, 无法计算  $x_R^{(r)}(k)$ , 需进一步分析。

$$\text{令 } u_1 = \frac{1-a+au}{1+au}, u_2 = \frac{b}{1+au}, u_3 = \frac{c}{1+au}, u_4 = \frac{d}{1+au}, \text{公式(8)简化为公式(9)。}$$

$$x_R^{(r)}(k) = u_1 x_R^{(r)}(k-1) + u_2 k + u_3 + u_4 k^\theta \quad (9)$$

根据公式(9), 当  $k=2, 3$  时, 可得公式(10)和(11)

$$x_R^{(r)}(2) = u_1 x_R^{(r)}(1) + 2u_2 + u_3 + 2^\theta u_4 \quad (10)$$

$$x_R^{(r)}(3) = u_1 x_R^{(r)}(2) + 3u_2 + u_3 + 3^\theta u_4 \quad (11)$$

公式(10)代入公式(11), 整理得公式(12)。

$$x_R^{(r)}(3) = u_1^2 x_R^{(r)}(1) + 2u_1 u_2 + u_1 u_3 + 2^\theta u_1 u_4 + 3u_2 + u_3 + 3^\theta u_4 \quad (12)$$

当  $k=4$ , 可得公式(13)。

$$x_R^{(r)}(4) = u_1 x_R^{(r)}(3) + 4u_2 + u_3 + 4^\theta u_4 \quad (13)$$

类似地, 将公式(12)代入公式(13), 整理得公式(14)。

$$x_R^{(r)}(4) = u_1^3 x_R^{(r)}(1) + 2u_1^2 u_2 + u_1^2 u_3 + 2^\theta u_1^2 u_4 + 3u_1 u_2 + u_1 u_3 + 3^\theta u_1 u_4 + 4u_2 + u_3 + 4^\theta u_4 \quad (14)$$

⋮

当  $k=t, (t = 2, 3, \dots, n)$ , 可得公式(15)。

$$x_R^{(r)}(t) = u_1^{(t-1)} x_R^{(r)}(1) + u_1^{(t-2)} (2u_2 + u_3 + 2^\theta u_4) + u_1^{(t-3)} (3u_2 + u_3 + 3^\theta u_4) + \dots + u_1^{(1)} ((t-1)u_2 + u_3 + (t-1)^\theta u_4) + u_1^{(0)} (tu_2 + u_3 + t^\theta u_4) \quad (15)$$

将公式(15)简化为公式(16)。

$$x_R^{(r)}(t) = u_1^{(t-1)} x_R^{(r)}(1) + \sum_{g=0}^{t-2} [(t-g)u_2 + u_3 + (t-g)^\theta u_4] u_1^g \quad (16)$$

证明结束。证明(2)与(1)类似, 证明过程略。

定理 3: 设序列  $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$  为原始序列<sup>[18]</sup>,  $p \in \mathbf{R}, q \in \mathbf{R}, X_R^{(p)}$  是  $X^{(0)}$  的  $p$  阶生成序列,  $X_R^{(q)}$  是  $X^{(0)}$  的  $q$  阶生成序列,  $X_R^{(p+q)}$  是  $X^{(0)}$  的  $p+q$  阶生成序列,  $(X_R^{(p)})^{(q)}$  是  $X_R^{(p)}$  的  $q$  阶生成序列,  $(X_R^{(q)})^{(p)}$  是  $X_R^{(q)}$  的  $p$  阶生成序列, 则多重生成算子满足交换律与指数率, 即公式(17)。

$$(X_R^{(p)})^{(q)} = (X_R^{(q)})^{(p)} = X_R^{(p+q)} \quad (17)$$

推论 1: 根据定理 3 可推导出公式(18)。

$$X_R^{(0)} = (X_R^{(r)})^{(-r)} = (X_R^{(-r)})^{(r)} \quad (18)$$

将公式(1)及公式(16)带入公式(18)即可得到累加还原式, 即公式(19)。

$$x_R^{(0)}(k) = (x_R^{(r)}(k))^{(-r)} = \sum_{i=1}^k \frac{\Gamma(-r+k-i)}{\Gamma(k-i+1)\Gamma(-r)} \sum_{j=1}^i \frac{\Gamma(r+i-j)}{\Gamma(i-j+1)\Gamma(r)} x_R^{(r)}(j) = \sum_{i=1}^k \frac{\Gamma(-r+k-i)}{\Gamma(k-i+1)\Gamma(-r)} x_R^{(r)}(i) \quad (19)$$

其中,  $k=2, 3, \dots, n, x_R^{(0)}(1) = x^{(0)}(1)$ 。至此, 证明结束。

#### 4. FODGM( $r, 1, k^\theta, u$ ) 模型建模流程

FODGM( $r, 1, k^\theta, u$ ) 模型中含有三个需借助寻优算法求解的待定参数(指数  $\theta$ 、实数域阶数  $r$ 、背景值系数  $u$ )。粒子群算法(the Particle Swarm Optimization, PSO)是一种效果良好且广泛使用的群体智能全局的优化算法<sup>[19]</sup>, 为此, 使用 PSO 实现三大待定参数的组合全局寻优。具体地, 在平均相对模拟误差最小的条件下寻求 FODGM( $r, 1, k^\theta, u$ ) 模型的最优  $r^*, u^*, \theta^*$ , 即满足公式(20)。

$$\min f(r, u, \theta) = \frac{1}{n-1} \sum_{k=2}^n \frac{|\hat{x}^{(0)}(k) - x^{(0)}(k)|}{x^{(0)}(k)}, r \in \mathbf{R}, \lambda \in [0, 1], \theta \in \mathbf{R} \quad (20)$$

此外, 模型误差检验是检验模型性能的重要手段, 其中, 平均相对百分比误差(MRPE)是评估灰色预测模型性能通用的方法。通常情况下, 为衡量模型的模拟性能、预测性能及综合性能, 需分别计算平均相对模拟百分比误差(MRSPE)、平均相对预测百分比误差(MRPPE)和综合平均相对百分比误差(GMRPE)。具体地, 设建模序列前  $t$  个元素组成的序列为模拟序列, 建模序列后  $(n-t)$  个元素组成的序列为预测序列, 则相关误差的计算方法如公式(21), 其中,  $\Delta k$  为模拟(预测)值  $\hat{x}^{(0)}(k)$  的相对误差(RPE)。

$$\bar{\Delta}_{MRSPE} = \frac{1}{t-1} \sum_{k=2}^t \Delta k, \bar{\Delta}_{MRPPE} = \frac{1}{n-t} \sum_{k=t+1}^n \Delta k, \bar{\Delta}_{GMRPE} = \frac{1}{n-1} \sum_{k=2}^n \Delta k \quad (21)$$

根据 FODGM( $r, 1, k^\theta, u$ ) 模型的建模过程及上述参数的最优求解和模型的检验方法, FODGM( $r, 1, k^\theta, u$ ) 的建模步骤如图 3 所示。

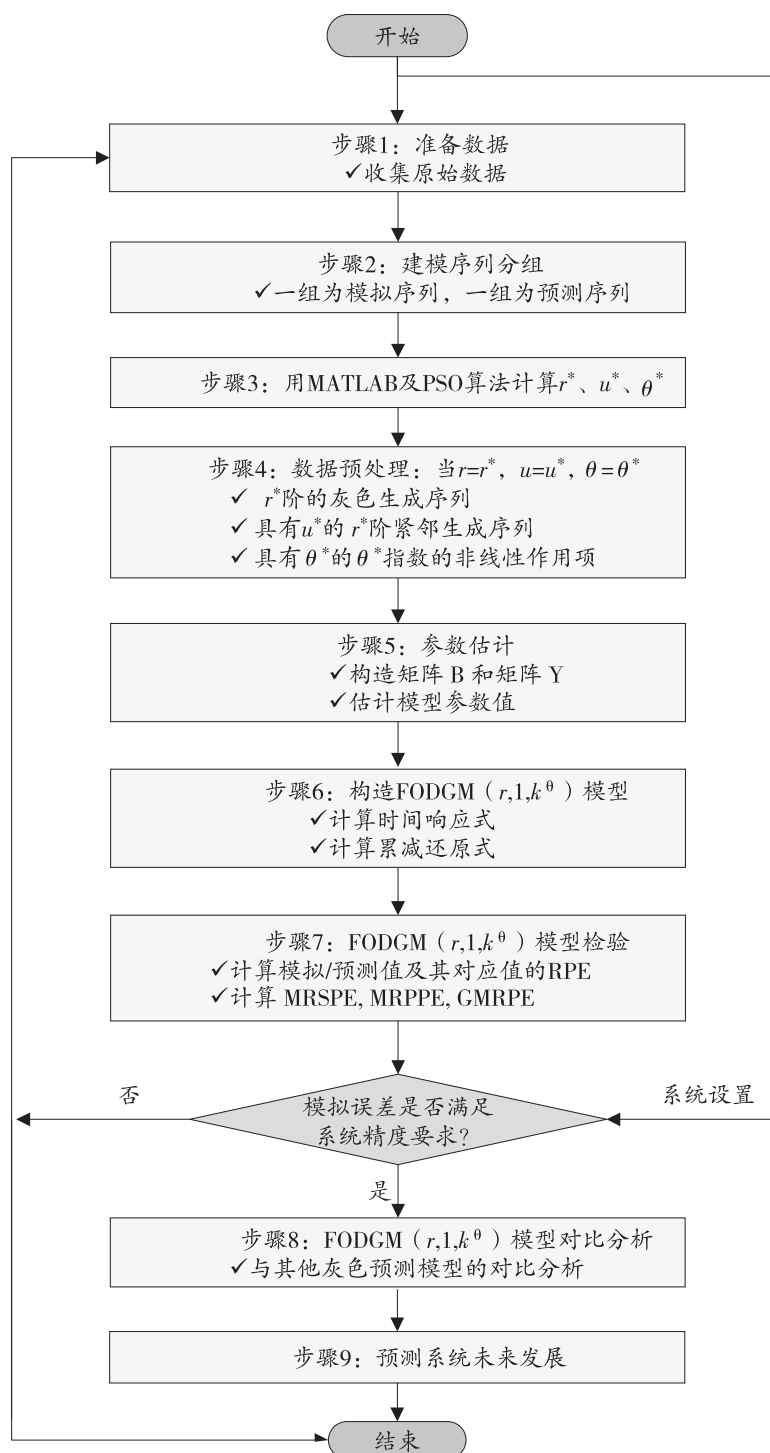


图3 FODGM( $r, 1, k^\theta, u$ )模型建模流程图

#### 四、FODGM( $r, 1, k^\theta, u$ )模型的应用: 我国科技人才需求预测

本文基于官方网站公布数据的可获得性,以 R&D 人员全时当量反映我国科技人才的需求量,数据来源于相应年度的统计年鉴。如表 1 所示。我国 R&D 人员全时当量的增长具有较大的波动性,最低增

长率为 1998 年的 -9.15%, 最高增长率为 2005 年的 18.39%, 增长率区间跨度超过 27%, 具有较明显的非线性特征, 容易导致传统 GM(1,1) 模型的预测效果不理想。

表 1 1995—2021 年我国 R&D 人员全时当量(万人年)

年份	R&D 人员全时当量	增长率	年份	R&D 人员全时当量	增长率
1995	75.2	-	2009	229.1	16.59%
1996	80.4	6.91%	2010	255.4	11.48%
1997	83.1	3.36%	2011	288.3	12.88%
1998	75.5	-9.15%	2012	324.7	12.63%
1999	82.2	8.87%	2013	353.3	8.81%
2000	92.2	12.17%	2014	371.1	5.04%
2001	95.7	3.8%	2015	375.9	1.29%
2002	103.5	8.15%	2016	387.8	3.17%
2003	109.5	5.8%	2017	403.4	4.02%
2004	115.3	5.3%	2018	438.1	8.6%
2005	136.5	18.39%	2019	480.1	9.59%
2006	150.2	10.04%	2020	523.5	9.04%
2007	173.6	15.58%	2021	571.6	9.19%
2008	196.5	13.19%			

### 1. FODGM( $r, 1, k^\theta, u$ ) 模型建模

为同时验证 FODGM( $r, 1, k^\theta, u$ ) 模型的模拟及预测性能, 本文将表 1 中的原始数据进行分组:  $X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(t), x^{(0)}(t+1), \dots, x^{(0)}(n)]$ 。为体现新模型建模的稳健性, 根据 R&D 人员全时当量数据, 预留不同长度数据进行建模。为验证 FODGM( $r, 1, k^\theta, u$ ) 模型较现有灰色预测模型的优越性及其对 FDGM( $1, 1, k^2$ ) 模型结构和参数缺陷的优化, 将常用预测模型 GM(1,1)、新模型的基础模型 FDGM( $1, 1, k^2$ )、仅优化单一参数  $u$  的 FODGM( $1, 1, k^2, u$ ) 模型作为对比模型。各模型建模结果如表 2 所示, 并根据表 2 数据绘制各模型不同样本区间下的误差折线图, 如图 3 所示。

根据表 2 与图 3, 新模型在不同的样本区间下均具有最小的综合误差, 表明新模型较其余模型具有最优的建模性能及稳健性。由于篇幅所限, 后文仅展示选取前 26 个数据组成的建模序列  $X_s^{(0)}$ 、预留后 1 个数据组成预测序列为  $X_f^{(0)}$  的详细建模过程。根据 FODGM( $r, 1, k^\theta, u$ ) 模型的建模步骤, 使用 MATLAB 及 PSO 算法 (MATLAB 自带), 计算得 FODGM( $r, 1, k^\theta, u$ ) 模型最优参数组合数值为  $r^* = -0.373\ 333$ 、 $u^* = 0.314\ 490$ 、 $\theta^* = -0.016\ 398$ 。根据定理 1 构造模型基本参数估计矩阵 B 和 Y, 并计算得出基本参数列如下:

$$\hat{p} = (a, b, c, d)^T = (B^T B)^{-1} B^T Y = [0.121\ 4, 0.498\ 2, 758.860\ 8, -780.187\ 6]^T$$

进一步根据定理 2 可得 FODGM( $r, 1, k^\theta, u$ ) 模型时间响应式中所需参数如下:

$$u_1 = \frac{1-a+au^*}{1+au^*} = 0.883\ 1, u_2 = \frac{b}{1+au^*} = 0.479\ 9,$$

$$u_3 = \frac{c}{1+au^*} = 730.960\ 4, u_4 = \frac{d}{1+au^*} = -751.503\ 1$$



将上述参数代入公式(4)即可得  $FODGM(r, 1, k^\theta, u)$  模型时间响应式。

表 2 不同样本区间下各模型建模的综合误差

预测检验长度	0	1	2	3	4	5	6	7
$FODGM(r^*, 1, k^{\theta^*}, u^*)$	4.268 8%	4.310 2%	4.247 1%	4.508 5%	4.743 4%	4.566 2%	4.499 6%	4.945 6%
$FDGM(1, 1, k^2)$	4.648 1%	4.659 3%	4.287 8%	4.539 9%	4.746 8%	4.833 1%	5.536 8%	5.571 2%
$FODGM(1, 1, k^2, u)$	4.834 5%	5.081 9%	5.200 1%	5.182 7%	5.109 6%	5.256 2%	6.070 7%	8.199 3%
$GM(1, 1)$	11.756 1%	11.460 2%	11.197 9%	10.923 1%	10.688 1%	10.477 4%	11.243 5%	12.724 8%

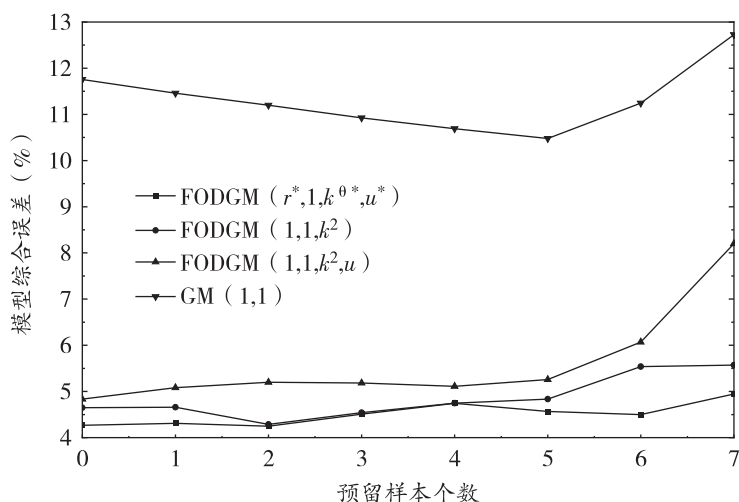


图 4 不同样本区间下模型误差折线图

## 2. $FODGM(r, 1, k^\theta, u)$ 模型检验及对比分析

同样地,使用 MATLAB 自带 PSO 程序对  $FDGM(1, 1, k^2)$  和  $FODGM(1, 1, k^2, u)$  模型中待定参数寻优,分别为  $r=0.197\ 829$ 、 $u=1$ 。其余建模结果信息如表 3 所示,并绘制各模型建模误差柱状图如图 4 所示,其中  $\hat{x}^{(0)}$  为模型得到的模拟(预测)数据,  $\Delta k$  为  $\hat{x}^{(0)}$  相对误差 ( $\Delta k = |\hat{x}^{(0)}(k) - x^{(0)}(k)| / x^{(0)}(k) \times 100\%$ )。

表 3 我国 R&D 人员全时当量模拟及预测结果

序号	$x^{(0)}$	$FODGM(r^*, 1, k^{\theta^*}, u^*)$		$FDGM(1, 1, k^2)$		$FODGM(1, 1, k^2, u)$		$GM(1, 1)$	
		$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$
模拟检验									
$k=2$	80.4	83.393	3.723	81.579	1.466	84.297	4.847	76.601	4.726
$k=3$	83.1	83.096	0.005	83.100	0.000	77.876	6.287	83.245	0.174
$k=4$	75.5	81.934	8.522	84.341	11.710	75.829	0.436	90.465	19.822
$k=5$	82.2	82.294	0.114	86.712	5.490	77.628	5.562	98.312	19.601
$k=6$	92.2	85.058	7.747	90.802	1.516	82.807	10.188	106.840	15.878
$k=7$	95.7	90.518	5.414	96.874	1.227	90.956	4.957	116.107	21.324

续表

序号	$x^{(0)}$	FODGM( $r^*, 1, k^{\theta^*}, u^*$ )		FDGM( $1, 1, k^2$ )		FODGM( $1, 1, k^2, u$ )		GM( $1, 1$ )	
		$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$	$\hat{x}^{(0)}$	$\Delta k(\%)$
$k = 8$	103.5	98.701	4.637	105.034	1.482	101.716	1.724	126.178	21.911
$k = 9$	109.5	109.503	0.003	115.300	5.297	114.771	4.814	137.123	25.226
$k = 10$	115.3	122.760	6.470	127.641	10.703	129.843	12.613	149.017	29.243
$k = 11$	136.5	138.283	1.306	141.991	4.022	146.687	7.463	161.942	18.639
$k = 12$	150.2	155.874	3.778	158.265	5.369	165.089	9.913	175.989	17.170
$k = 13$	173.6	175.341	1.003	176.365	1.593	184.861	6.487	191.254	10.169
$k = 14$	196.5	196.500	0.000	196.187	0.160	205.836	4.751	207.843	5.773
$k = 15$	229.1	219.179	4.331	217.618	5.012	227.870	0.537	225.872	1.409
$k = 16$	255.4	243.220	4.769	240.548	5.815	250.832	1.788	245.464	3.891
$k = 17$	288.3	268.478	6.876	264.864	8.129	274.613	4.748	266.755	7.473
$k = 18$	324.7	294.822	9.202	290.452	10.548	299.111	7.881	289.893	10.720
$k = 19$	353.3	322.134	8.822	317.203	10.217	324.240	8.225	315.038	10.830
$k = 20$	371.1	350.306	5.603	345.009	7.031	349.925	5.706	342.364	7.743
$k = 21$	375.9	379.245	0.890	373.762	0.569	376.097	0.052	372.061	1.021
$k = 22$	387.8	408.864	5.432	403.362	4.013	402.697	3.841	404.333	4.263
$k = 23$	403.4	439.086	8.846	433.708	7.513	429.674	6.513	439.405	8.925
$k = 24$	438.1	469.846	7.246	464.703	6.072	456.983	4.310	477.519	8.998
$k = 25$	480.1	501.081	4.370	496.254	3.365	484.582	0.934	518.938	8.090
$k = 26$	523.5	532.739	1.765	528.273	0.912	512.437	2.113	563.951	7.727
$\bar{\Delta}_{MRSPE}$			4.435		4.769		5.068		11.630
预测检验									
$k = 27$	571.6	564.772	1.195	560.672	1.912	540.517	5.438	612.867	7.220
$\bar{\Delta}_{MRPPE}$			1.195		1.912		5.438		7.220
$\bar{\Delta}_{GMRPE}$			4.310		4.659		5.082		11.460

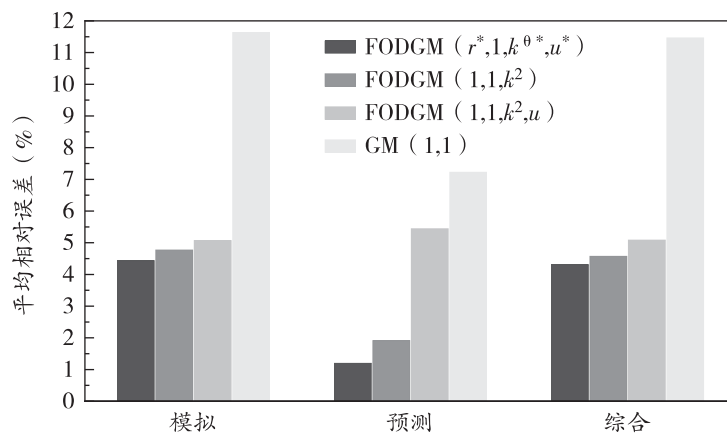


图5 模型建模误差柱状图

根据上述建模结果误差, 新模型的模拟检验、预测检验及综合性能较其余模型均更优, 表明新模型对我国科技人才需求的预测有着显著的优越性, 更能准确预测未来趋势。为了进一步说明本文提出的新模型在科技人才需求预测中的优势, 与已有的相关研究结果进行比较, 如表 4 所示。可以看出 GM(1, 1) 模型预测的结果不稳定, 一是因为预测的对象不一样, 二是因为 GM(1, 1) 模型本身对原始数据要求较高。在本文中, 应用 GM(1, 1) 模型的预测效果远不如新模型 FODGM( $r, 1, k^\theta, u$ )。

表 4 相关科技人才预测的综合平均百分误差

序号	方法	作者	预测对象	综合平均百分比误差
1	GM(1,1) 模型	姚娟等 <sup>[20]</sup>	科技创新人才	7.270 5%
2	GM(1,1) 模型	卞永峰, 李恩平 <sup>[21]</sup>	山西省转型期紧缺科技人才	8.769 3%
3	GM(1,1) 模型	王维, 李钰 <sup>[22]</sup>	四川省从业人才	0.862 3%
4	离散灰色预测模型	尹逊震, 门可佩 <sup>[9]</sup>	全国科技人力资源	0.889 5%
5	MGM(1, n) 模型	王五祥等 <sup>[23]</sup>	天津市 R&D	4.791 3%

综上所述, 与现有的相关研究相比, 本文提出的科技人才需求量预测模型考虑了更多因素的影响, 具有更高的预测精度和可靠性, 有助于政策制定者更好地把握科技人才市场的变化趋势和未来发展方向, 进而促进我国科技人才的可持续高质量发展。

### 3. FODGM( $r, 1, k^\theta, u$ ) 模型预测结果

如表 5 所示, 从预测未来的数据可以看出, GM(1, 1) 模型受其模型结构的制约, 预测结果呈现指数增长趋势。FDGM( $1, 1, k^2$ )、FODGM( $1, 1, k^2, u$ ) 与新模型由于非线性项的引入, 具有处理复杂序列规律的能力, 呈现出更科学的预测结果。然而, FDGM( $1, 1, k^2$ ) 与 FODGM( $1, 1, k^2, u$ ) 模型固化的非线性项结构限制了模型性能和作用空间, 导致作用项指数无法依据原始序列的数据特征进行匹配, 缺乏动态可调性。此外, 根据现有数据的模拟及预测检验情况来看, 新模型较 FDGM( $1, 1, k^2$ ) 与 FODGM( $1, 1, k^2, u$ ) 具有更优的预测检验性能。因此, 新模型的预测结果更具科学性与合理性。根据预测结果可知, 未来我国 R&D 人员全时当量呈现逐步上升趋势, 表明我国未来科技人才需求量处于上升阶段, 且增幅与历年趋势相近, 科技人才需求端压力显著。

表 5 我国 R&D 人员全时当量未来预测数据信息表

模型	2022 年	2023 年	2024 年	2025 年	2026 年
FODGM( $r^*, 1, k^{\theta*}, u^*$ )	597.139	629.801	662.728	695.888	729.258
FDGM( $1, 1, k^2$ )	593.368	626.282	659.336	692.458	725.577
FODGM( $1, 1, k^2, u$ )	568.795	597.246	625.849	654.587	683.443
GM(1, 1)	666.027	723.798	786.579	854.807	928.952

## 五、结论

科技人才需求量具有非线性发展的特征, 并受多种难以定量描述因素的影响, 现有 GM(1, 1) 模型将灰作用量  $b$  视为影响系统发展的所有灰信息, 建模过程中容易忽略部分原始数据间的非线性关系, 从而导致建模效果不理想。为此, 本文在分析现有模型缺陷的基础上, 构建了新型优化离散灰色预测模型

FODGM( $r, 1, k^\theta, u$ )。新模型弥补了现有模型的不足, 实现了系统发展灰信息非线性规律的较好反映, 并通过参数组合优化进一步提升了模型建模性能。将新模型用于我国科技人才需求量预测, 预测结果显示未来我国科技人才需求量将会呈现持续稳步上升的趋势。预计到 2026 年, 我国科技人员全时当量将达到 729.258 万人年。这一结果表明科技人才需求端的压力将会显著增加, 也为进一步合理制定相关政策提供了科学依据。相关部门可以基于预测结果制定相关政策, 缓解科技人才需求端的压力, 避免人才资源的浪费, 为我国科技事业发展提供有力的人才支持, 进而促进科技进步和创新驱动发展。

本文提出的新预测模型 FODGM( $r, 1, k^\theta, u$ ) 在预测科技人才需求量方面具有一定的科学性和实用性, 但也存在一些不足之处。首先, 新预测模型建立在历史数据的基础上, 因此对于突发事件、政策变化等因素的影响预测能力相对较弱。其次, 预测模型的精度还有待提高。为进一步提高新模型预测的准确性和实用性, 可以从加强对新技术、新趋势的研究和建立更加完整、准确的数据集两个方面进行探索和改进。综上所述, 灰色预测是一个复杂的问题, 需要考虑数据的演变规律以及模型的适用范围。随着技术的不断发展和经验的不断积累, 灰色预测模型的准确性和实用性将会不断提高。

#### 参考文献:

- [1] 杜谦, 宋卫国. 科技人才定义及相关统计问题[J]. 中国科技论坛, 2004(5): 137-141.
- [2] 瞿群臻, 王嘉吉, 唐梦雪, 等. 基于组合模型的“十四五”期间中国科技人才需求预测[J]. 科技管理研究, 2021, 41(21): 129-135.
- [3] 张建勇, 赵涛. 基于 BP 神经网络的科技人才需求预测[J]. 科技管理研究, 2009, 29(8): 501-502.
- [4] [1] 李锡元, 王艳姣. 重点产业海外高层次人才需求研究——以湖北省汽车和钢铁产业为例[J]. 科技进步与对策, 2014, 31(8): 135-139.
- [5] AKINULI B O, APALOWO R K. Regression analysis based effective manpower planning methodology: a case study[J]. Journal of Engineering Research and Reports, 2018: 1-12.
- [6] LI L, WANG R, LI X. Grey GM(1, 1,  $\beta k$ ) model and its application in R&D personnel[J]. Journal of Grey System, 2017, 29(1): 120-134.
- [7] 白旭光. 基于灰色系统理论的风险投资人才需求预测研究——以山西省为例[J]. 系统科学学报, 2017, 25(4): 67-69.
- [8] 闵惜琳. 基于灰色预测模型 GM(1, 1) 的人才需求分析[J]. 科技管理研究, 2005(6): 72-74+77.
- [9] 尹逊震, 门可佩. 离散灰色模型在科技人才资源预测中的应用[J]. 科技进步与对策 2007(10) 204-206.
- [10] YIN K, GENG Y, LI X. Improved grey prediction model based on exponential grey action quantity[J]. Journal of Systems Engineering and Electronics, 2018, 29(3): 560-570.
- [11] 邓聚龙. 灰色控制系统[J]. 华中工学院学报, 1982(3): 9-18.
- [12] 谢乃明, 刘思峰. 离散 GM(1, 1) 模型与灰色预测模型建模机理[J]. 系统工程理论与实践, 2005(1): 93-99.
- [13] 崔杰, 党耀国, 刘思峰. 一种新的灰色预测模型及其建模机理[J/OL]. 控制与决策, 2009, 24(11): 1702-1706.
- [14] ZENG B, LI C. Forecasting the natural gas demand in China using a self-adapting intelligent grey model[J/OL]. Energy, 2016, 112: 810-825.
- [15] ZENG B, LI H. Prediction of coalbed methane production in China based on an optimized grey system model[J]. Energy & Fuels, 2021, 35(5): 4333-4344.
- [16] 曾波, 余乐安, 刘思峰, 等. 灰色累加算子与灰色累减算子的统一及其应用[J]. 系统工程理论与实践, 2021, 41(10): 2710-2720.
- [17] 曾波, 李树良, 孟伟. 灰色预测理论及其应用[M]. 北京: 科学出版社, 2020.

- [18] 孟伟, 曾波. 分数阶算子与灰色预测模型研究[M]. 北京: 学出版社, 2015.
- [19] 万春林, 张卫. 基于改进粒子群算法的基数受限最优化问题研究[J/OL]. 统计与决策, 2021, 37(20): 20-24.
- [20] 姚娟, 刘鸿渊, 刘建贤. 科技创新人才区域性需求趋势研究——对四川、陕西、上海的预测与比较分析[J]. 科技进步与对策, 2019, 36(14): 46-52.
- [21] 卞永峰, 李恩平. 基于组合预测模型的山西省转型期紧缺科技人才需求预测研究[J]. 科技管理研究, 2013, 33(21): 41-45.
- [22] 王维, 李钰. 最优组合预测在四川省人才需求预测中的应用[J]. 价值工程, 2005(3): 9-12.
- [23] 王五祥, 张维, 崔和瑞, 等. 多变量灰色模型 MGM(1, n) 在 R&D 投资预测中的应用[J]. 研究与发展管理, 2006(2): 92-96+103.

## Optimization of Grey Combination Forecasting Model and Forecasting the Demand for Scientific and Technological Talents

WANG Xiao-ying<sup>1</sup>, GOU Xiao-yi<sup>2</sup>, ZENG Bo<sup>3</sup>

(1. School of New Media Art, Chongqing Finance and Business College, Chongqing 401320, China;

2. College of Economics and Management, Nanjing University of Aeronautics and Astronautics,

Nanjing 211106, Jiangsu, China; 3. School of Management Science and Engineering, Chongqing

Technology and Business University, Chongqing 400067, China)

**Abstract:** In recent years, China has proposed the task of promoting technological innovation and improving the quality of development to achieve the innovation-driven development strategy and consolidate the foundation of the real economy. As an important power source and core element to promote the innovation-driven development strategy, scientific and technological talents are conducive to stimulating new economic growth points and releasing the endogenous power of the economy while promoting scientific and technological progress, thus promoting the structural transformation and high-quality economic development of China. The demand forecast for scientific and technological talents is an important basis for the cultivation of scientific and technological talents and policy formulation. The demand for scientific and technological talents has typical non-linear development characteristics and is also influenced by a variety of factors that are difficult to describe quantitatively. Therefore, the grey prediction model with the characteristics of small data and poor information modeling is gradually becoming a common method for forecasting the demand for scientific and technological talents.

Although the grey prediction model has the advantages of simple structure and wide applicability, there are still some shortcomings, such as unstable model prediction performance and poor model compatibility. Therefore, starting from the characteristics of the demand for scientific and technological talent data, this article proposes a new grey prediction model FODGM( $r, 1, k^{\theta}, u$ ) with dynamic adaptability based on the defect analysis of the FDGM( $1, 1, k^2$ ) model. The new model adds a nonlinear correction term to the FDGM( $1, 1, k^2$ ) model, which improves the model's simulation performance of nonlinear original series; the dynamic adaptivity of the model is improved by extending the original first-order accumulation to fractional-order. The new model

achieves a better simulation of the nonlinear law of grey information of system development and the totally real number fields expansion of the accumulating order action range, which alleviates the influence of extreme values in the original sequence on the model performance and improves the compatibility of the model structure and the stability of the prediction results.

The national research and development (R&D) personnel of full-time equivalent is selected to reflect the demand for scientific and technological talents in China. In order to reflect the robustness of the new model, data of different lengths are reserved for simulation and prediction modeling, and compared with other models. The results show that the  $FODGM(r, 1, k^\theta, u)$  model has the smallest comprehensive error in different sample intervals. Finally, the  $FODGM(r, 1, k^\theta, u)$  model was applied to predict the future demand for R&D personnel in China. The results showed that the demand for R&D personnel in China is gradually increasing in the future. It is expected that the full-time equivalent of R&D personnel in China will reach 7.29258 million person-years in 2026, indicating significant pressure on the demand side of scientific and technological talents. To this end, relevant countermeasures and suggestions have been proposed. The research results have positive significance in promoting the sustainability of China's science and technology industry, alleviating the pressure on the demand for scientific and technological talents, and promoting the healthy development of China's science and technology industry. It has important value for enriching, developing, and improving the theoretical system of grey prediction models.

**Key words:** demand forecast of scientific and technological talents; discrete grey model; model structure; nonlinear correction term; parameters combination optimization

**CLC number:** F272. 1; C964. 2; N941. 5 **Document code:** A **Article ID:** 1674-8131(2023)03-0094-14

(编辑:朱德东)

## 声 明

本刊已许可中国学术期刊(光盘版)电子杂志社在中国知网及其系列数据库产品中以数字化方式复制、汇编、发行、信息网络传播本刊全文;同时,本刊为《国家哲学社会科学学术期刊数据库》《万方——数字化期刊群》《中文科技期刊数据库》《科技论文在线》《超星数字图书馆》《国研网》《龙源期刊网》《教育阅读网》《博看网》等数据库全文收录期刊(其中《国研网》为选择性收录),论文在本刊发表后将通过上述数据库传播。

文章凡经本刊选用,即视为作者同意本刊代理该作品电子版的信息网络传播权,并且本刊有权授权其他机构进行该作品电子版信息的网络传播。

作者向本刊提交文章发表的行为即视为同意本刊上述声明。若作者不同意其作品收录入上述或其他数据库,请在来稿时说明,我们可做相应处理。

西部论坛编辑部