https://doi.org/10.66106/tiyua5.20250106

·竞技体育研究·

基于随机森林和灰色GM（1,1）对四人雪车成绩分析及米兰冬奥会混合预测研究

徐嘉欣[1]，尹一全[2*]
1.烟台南山学院体育学院，山东烟台，265713；
2.北京体育大学中国冰雪运动学院，北京，100084

摘要：为提高我国男子四人雪车项目的竞技表现提供理论参考，本研究提出了一种融合随机森林回归与灰色GM(1,1)模型的混合预测方法。基于2010-2022年四届冬奥会男子四人雪车项目前三名单轮成绩数据，结合赛道物理参数，构建赛道难度系数与加权指标，利用随机森林量化赛道特征对成绩的影响，并通过灰色模型捕捉成绩的长期趋势。研究结果显示，赛道垂直落差与弯道数量对成绩影响最为显著，灰色模型预测的平均相对误差为2.88%，关联度为0.833，方差比为0.307，满足一级精度标准。最终预测2026年米兰冬奥会男子四人雪车前三名最优轮次成绩分别为58.66秒、60.72秒和60.99秒。该方法有效解决了小样本条件下冬季项目成绩预测的难题，为我国雪车项目备战米兰冬奥会提供了科学依据，并为同类冰雪竞速项目提供了可借鉴的方法论框架。

关键词：随机森林；灰色GM(1,1)模型；四人雪车；成绩预测；米兰冬奥会

Performance Analysis of Four-man Bobsleigh and Hybrid Prediction for Milan Winter Olympics Based on Random Forest and Grey GM(1,1) Model

Jiaxin Xu[1]，Yiquan Yin[2*]
1.School of Physical Education, Yantai Nanshan University，Yantai Shandong 265713，China；
2.China Snow and Ice Sports College, Beijing Sport University，Beijing 100084，China

Abstract：To provide a theoretical reference for improving the competitive performance of China's four-man bobsleigh team, this study proposes a hybrid prediction method integrating random forest regression and the grey GM(1,1) model. Based on the single-run time data of the top three athletes in the four Winter Olympics from 2010 to 2022, combined with track physical parameters, a track difficulty coefficient and a weighted index were constructed. The random forest model was used to quantify the influence of track characteristics on performance, while the grey model captured long-term performance trends. The results show that vertical drop and number of curves were the most significant factors affecting performance. The grey model achieved a mean relative error of 2.88%, a relational degree of 0.833, and a variance ratio of 0.307, meeting the first-class accuracy standard. The predicted optimal run times for the top three in the 2026 Milan Winter Olympics are 58.66 s, 60.72 s, and 60.99 s, respectively. This method effectively addresses the challenge of predicting performance in winter sports with small sample sizes, providing a scientific basis for China's bobsleigh preparation and a methodological reference for similar ice and snow racing sports.

Keywords：Random Forest；Grey GM(1,1) Model；Four-man Bobsleigh；Performance Prediction；Milan Winter Olympics

Copyright © 2025 by author(s) and Quest Press Limited.This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
https://creativecommons.org/licenses/by/4.0/

竞技体育中对项目成绩预测方法研究是该领域一类重要的理论研究，随着信息技术的发展及统计学方法的不断优化，竞技体育成绩预测的准确性不断提升[1]。对奥运会成绩的预测不仅直接影响项目的训练准备，还对关系到各个国家奥运重点项目的主攻方向性选择以及项目的布局谋划[2]。我国雪车项目发展虽晚但进步迅速[3]，自2015年雪车国家队组建以来，已有多名选手在IBSF世界杯屡次跻身八强，并在2022年北京冬奥会上实现了历史性突破[4]。由于该项目成绩受到差异化的场地器械影响较大，因此选择合适的工具，通过预测成绩厘清成绩应先各因素就更具重要性。

随着大数据技术的快速发展，机器学习方法不断优化应用于体育领域[5-7]，随机森林算法凭借其优秀的非线性建模能力，在竞技结果预测领域取得突破性进展。学者们通过对随机森林算法的优化，在网球、篮球、曲棍球等项目中等项目中，对竞技能力影响因素、伤病影响因素以及运动表现预测评估方面取得诸多进展[8-10]，并提高了体育大数据挖掘准确率，为运动表现分析提供了新范式[11-13]。灰色系统理论则主要应用于竞技成绩预测[14-15]。我国学者在田径项目的成绩演化规律以及成绩预测方面都做到了较高的预测成功率[16-18]，并直接服务于奥运会的备战和战略选择。

由于竞技体育领域高质量数据量的限制，通过结合两种数据算法，机器学习与灰色系统理论在成绩预测和因素甄别研究中能够形成方法学互补[19]。随机森林算法在复杂特征挖掘方面优势显著，而灰色系统理论在小样本建模中展现独特价值。为提高我国男子四人雪车项目的运动员的竞赛成绩提供理论参考，本研究旨在设计合理的方法将两种模型混合提高预测效果，一方面完成对2026 年米兰冬奥会男子四人雪车项目的成绩预测，另一方面验证该方法的适用性和可靠性。为我国备战工作提供借鉴的同时，为同类项目的研究提供方法学参考。

1 研究设计

1.1 研究对象

基于2010—2022年四届冬奥会男子四人雪车项目前三名选手的单轮时间数据，结合赛道物理参数与灰色GM(1,1)小样本预测优势，构建随机森林回归与灰色模型的混合预测框架，验证混合模型在成绩预测中的可靠性与适用性。

1.2 研究总体设计

在雪车运动项目的竞技成绩研究中，影响因素涵盖预赛与决赛的竞技表现、不同赛季的周期性特征、分段计时数据、赛道参数及推车起跑阶段的时间效率等多重维度。基于冬奥会决赛前三名运动员的比赛数据进行实证分析，鉴于冬奥会赛事固定于2月举办的特征有效控制了赛季变量对成绩的潜在干扰，同时依据分段计时指标和推车时间两个变量与总成绩呈现显著正相关的理论前提，亦不作为独立分析对象。因此，通过将研究焦点集中于赛道难度系数这一核心变量。为了提高模型预测精度，准确反映雪车项目竞技表现的内在规律，本文提出了一种基于赛道参数识别

基金项目：
中央高校基本科研业务费专项资金资助课题（the Fundamental Research Funds for the Central Universities）
课题名称：
运动表现促进的关键技术研究（课题编号：2025KYPT04）
平台：
国家体育总局运动训练重点实验室（Key Laboratory of Sport Training of General Administration of Sport of China）
第一作者：
徐嘉欣（1998–），女，硕士研究生，助教，研究方向：冰雪运动。
***通讯作者**：
尹一全（1991–），男，博士，讲师，研究方向：运动训练。

和改进混合模型的冬奥会成绩预测方法（图1）。该方法先通过物理参数融合算法量化赛道难度系数，并构建赛道加权指标作为核心预测指标，避免了传统研究中赛道特征与滑行效率的孤立分析；然后，在模型构建阶段充分考虑到雪车运动的多维影响因素，构建了涵盖赛道动力学特征与历史成绩演化的综合变量数据集；相对于单一预测模型，该方法针对小样本数据的局限性，结合灰色GM(1,1)的累加生成机制与随机森林的特征重要性评估，通过加权融合策略改进了传统模型对非线性关系的刻画能力，显著提升了预测结果的稳健性。

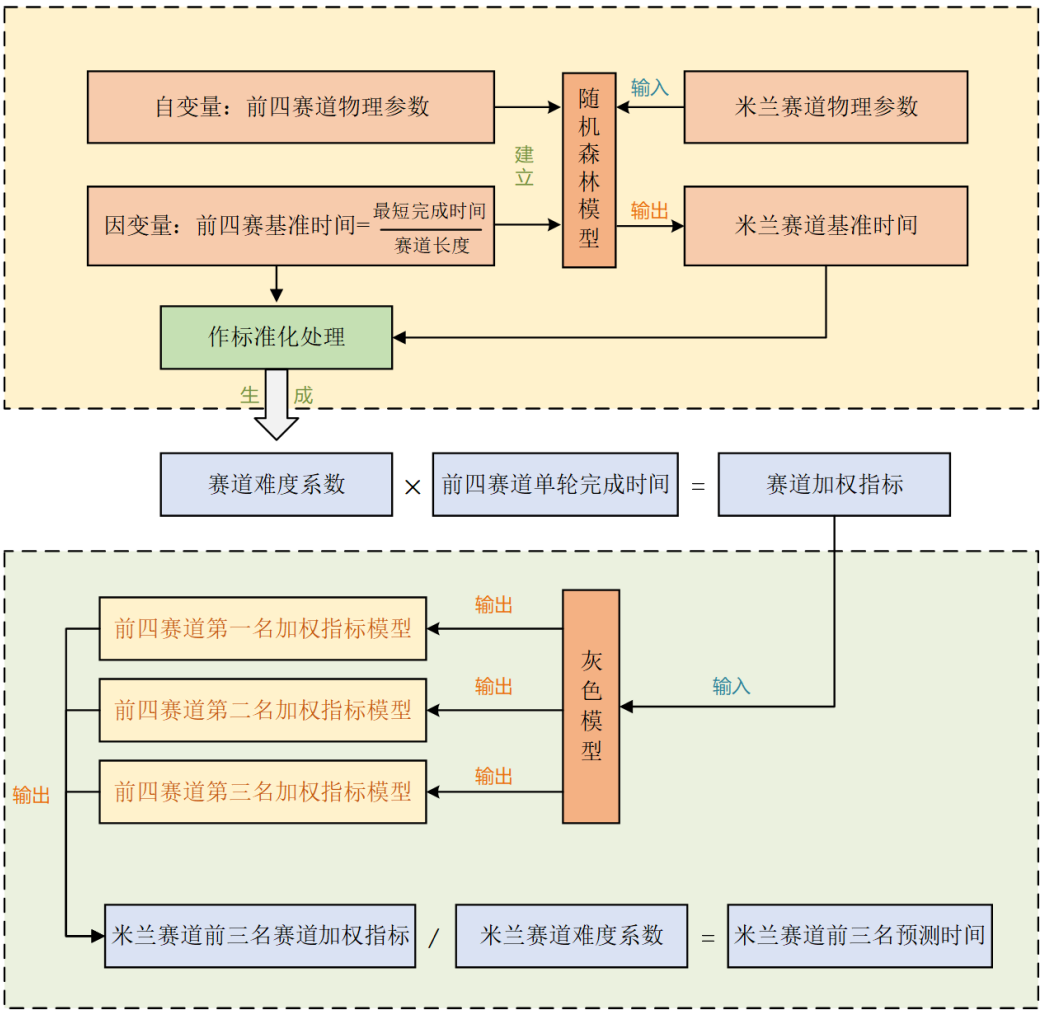

图1 本文方法总体流程

1.3 研究数据来源

数据来源于国际有舵雪橇和钢架雪车联合会（IBSF）官方数据库，该数据库收录了2004年以来所有国际A级赛事成绩记录。需特别说明的是，由于都灵冬奥会比赛场地切萨纳·帕里奥赛道自2006年退役后相关数据未被完整收录，为确保研究数据的完整性和可比性，本研究选取2010年温哥华冬奥会至2022年北京冬奥会连续四届冬季奥运会（第21-24届冬奥会）男子四人雪车项目前三名完整赛事的单轮时间成绩数据作为分析样本。赛道物理参数数据采集严格遵循国际雪车联合会（IBSF）的赛道认证标准，所有参数均取自官方发布的赛道拓扑测绘报告。米兰赛道数据来源于试滑认证运动员的场地手册。

1.4 数据模型与数据定义

1.4.1 随机森林算法

随机森林（Random Forest）是由统计学家Leo Breiman于2001年提出的一种集成学习模型，其核心思想是通过构建多棵决策树实现协同决策。该算法基于Bootstrap重抽样技术生成差异化

OPEN ACCESS

训练集，并结合特征随机选择机制，显著提升了模型的鲁棒性与计算效率。相较于单一决策树，随机森林在分类任务中表现出更高的准确率，尤其对含噪数据及缺失值场景具有较强耐受性；在回归任务中，其通过多树均值预测可有效解释自变量对因变量的影响关系，同时提供特征重要性排序功能，为变量筛选提供量化依据。从算法原理来看，随机森林通过集成策略降低模型过拟合风险，具体优势体现于三方面：其一，集成多棵决策树的投票或均值机制抑制了单棵树的不稳定性；其二，特征随机选择降低了维度对计算复杂度的依赖；其三，基于袋外数据（Out-of-Bag, OOB）的误差估计方法增强了模型评价的可信度。根据因变量类型的不同，随机森林可分为分类与回归两类：当因变量为离散型时，采用多数表决机制确定最终分类结果；当因变量为连续型时，通过回归树均值输出预测值。

随机森林回归算法的流程：（1）从原始数据集的 n 个个案中，应用 Bootstrap 法重复抽取 b 个训练样本集，据青年人体质随机森林回归模型构建与评价此构建 b 棵回归树。而每次抽取训练样本时未被抽取的个案组成的 b 个样本为袋外数据（out-of-bag，OOB），作为测试样本集。（2）构建回归树时，在每棵树的分节点处从 k 个自变量中随机选取 m_{try}（mtry < k）个作为备选分枝变量，然后根据分枝优度准则在其中确定最优分枝。R 软件建立随机森林回归模型时，默认参数。$m_{try} = k/3$（3）每棵回归树自顶向下递归分枝，不断生长，通过设定树棵数ntree值，作为回归树生长的终止条件；（4）生成的 b 棵回归树就组成了随机森林回归模型，模型估计效果通过袋外数据（OOB）预测的准确度来评价，即用测试集的均方误差来衡量，假设袋外数据样本数为m，则：

$$MSE_{OOB} = \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{m}$$

其中，y_i为OOB样本真实值，$\hat{y}$为模型预测值，m为OOB样本数。进一步引入决定系数衡量模型解释力：

$$R_{RF}^2 = 1 - \frac{MSE_{OOB}}{\widehat{\sigma_y^2}}$$

其中，$\widehat{\sigma_y^2}$表示 OOB 预测值的方差。R_{RF}^2趋近于1，则表明模型对因变量的变异具有较高解释能力。

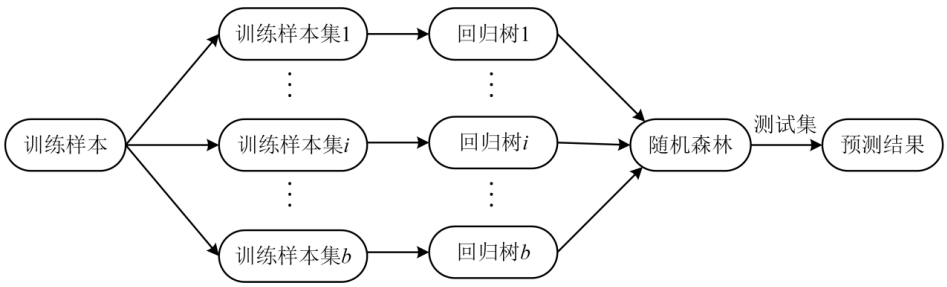

图2 随机森林模型示意图

随机森林回归方法的特性：

（1）当回归树的数目趋于无限大时，存在

$$E_{X,Y}\left(Y - av_j\hbar\ (X,\theta_j)\right)^2 \to E_{X,Y}\left(Y - E_\theta\hbar\ (X,\theta)\right)^2$$

其中，av 表示取平均值。因此，在实际估计模型过程中，常用 j 充分大情形下的公式 $Y = av_j h(X, \theta)$，近似代替随机森林回归函数 $E_\theta\hbar\ (X,\theta)$。此时，用 PE* 表示随机森林模型的泛化误差，则一棵回归树的平均泛化误差为：

$$PE^*(tree) = E_\theta E_{X,Y}\left(Y - \hbar\ (X,\theta)\right)^2$$

（2）如果对所有的 θ 有 $E(Y) = E_X h(X, \theta)$，则有

$$PE^* \le \bar{\rho} PE^*(tree)$$

其中，$\bar{\rho}$为$Y - \hbar\ (X,\theta)$和$Y - \hbar\ (X,\theta')$的相关系数，θ和θ'之间相互独立之间相互独立，说明随机森林模型的泛化误差比回归树的降低了。

1.4.2 灰色预测法

灰色系统理论作为一种处理信息不完备系统

的重要方法论，由我国学者邓聚龙教授于20世纪80年代初正式提出。该理论突破传统统计学对样本容量与数据分布的严格假设，尤其适用于竞技体育、经济预测等领域的小样本数据建模场景，能够通过有限信息挖掘系统内在规律。其核心思想在于通过数据生成技术弱化随机性，强化趋势特征，为解决“少数据、不确定性”系统的分析预测与控制问题提供了理论框架。相较于传统预测方法，灰色模型无需依赖大样本统计特性，且具有计算复杂度低、抗干扰能力强的优势，现已成为灰色系统理论中应用最广泛的研究分支。

在灰色预测模型中，GM(1,1)模型作为基础性单变量预测工具，因其理论简洁性与实践有效性受到学界广泛关注。模型名称中“GM”为灰色模型（Grey Model）的缩写，第一个“1”代表一阶微分方程，第二个“1”表示单变量建模维度。

设原始非负序列为，

$X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)]$ 通过一阶累加生成算子（1-AGO）构造新序列：

$$X^{(1)} = [x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)]$$

其中 $X^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i)(k = 1,2,\cdots,n)$。累加过程通过信息叠加增强数据规律性，为微分方程建模奠定基础。

为进一步平滑数据波动，定义背景值序列：

$$Z^{(1)} = [z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(n)]$$

其分量由相邻累加值的均值构成：

$$z^{(1)}(k) = 0.5 \times \left(x^{(1)}(k) + x^{(1)}(k-1)\right) \ (k = 2,3,\cdots,n)$$

基于此，GM(1,1)模型的基本形式可表述为：

$$x^{(0)}(k) + az^{(1)}(k) = b \ \ (k = 2,3,\cdots,n)$$

其中，a 为表征系统发展趋势的发展系数，b 为反映外部作用强度的灰色作用量。

模型参数 $p = (a,b)^T$ 可通过最小二乘法求解。构造数据矩阵：

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$$

则参数估计满足：

$$p = (B^T B)^{-1} B^T Y$$

该解法保证参数估计的统计最优性，为模型精度提供理论保障。

为揭示系统连续演化规律，定义白化微分方程：

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b$$

根据微分方程理论，其时间响应函数可解析为：

$$x^{(1)}(t) = \left(x^{(1)}(1) - \frac{b}{a}\right)e^{-a(t-1)} + \frac{b}{a}$$

离散化后得到累加序列预测值：

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(1) - \frac{\widehat{b}}{a}\right)e^{-ak} + \frac{b}{a} \ \ (k = 1,2,\cdots,n)$$

通过一阶累减还原操作，可得原始序列预测值：

$$x^{(0)}(k+1) = (1 - e^{a})\left(x^{(1)}(1) - \frac{b}{a}\right)e^{-ak}$$

1.4.3 名词解释

赛道物理参数（Track Physical Parameters）由总长度、有效赛段长度、起始海拔、高程变异系数、垂直落差、弯道曲率指数及坡度梯度差构成，系统性量化赛道的空间展布特征、竞技区段有效性、地形起伏剧烈度、弯道几何复杂度及纵坡动态变化幅度。

赛道难度系数（Track Difficulty Coefficient）：通过建立基准时间标准化模型，将各赛道最短理论完成时间（T_min）与有效赛段长度（L_race）的比值进行极差归一化处理，得到无量纲的0-1区间标度值。该系数通过公式CD=(T_min/L_race - min)/(max - min)计算，其中min/max分别代表样本集中各赛道单位长度最短时间的极值边界。

赛道加权指标（Track Weighted Index）：基于多元线性组合原理构建的复合评价指标，其数学表达式为I= α CD + β T_actual，其中CD为赛道难度系数，T_actual为实际滑行时间，α、β为经主成分分析确定的权重系数。该指标通过融合赛道固有属性参数与动态竞技表现数据，建立可量化评估赛道物理特性对竞技表现影响程度的指

OPEN ACCESS

ISSN 3105-7594（P）
ISSN 3105-7608（O）

标体系，为揭示地形-运动效能耦合机制提供数据基础。

2 研究结果

2.1 赛道难度系数建模

该技术指标作为赛道物理特性的综合体现，直接关系到滑行轨迹的动力学特征与运动员的操控策略选择。通过对近四届冬奥会男子四人雪车冠军单轮成绩的系统解析，本研究采用随机森林回归模型（Random Forest Regression），其Bagging集成算法通过特征子采样降低过拟合风险。模型输入特征包含6个标准化后的赛道参数，输出层为连续型难度系数值。通过设置决策树数量=200与随机种子=42，其余参数采用默认值，在保证模型泛化能力的同时提升特征重要性评估的精度。特征重要性排序基于基尼不纯度下降法直接生成。

本文利用随机森林回归模型分析赛道特征对雪车比赛成绩的影响机制。通过采集历史赛事中不同赛道的物理参数与比赛成绩数据，构建多维数据集。将赛道数据预处理后，计算各赛道的单位长度基准时间作为核心评价指标。数据矩阵包含赛道特征集$X=\{X_{ij}\}$ $N\times6$和基准时间向量$Y=\{Y_i\}$ $N\times1$，其中N为赛道样本数。数据清洗阶段剔除含缺失值的样本，并对特征进行标准化处理，采用Min-Max标准化方法对特征进行归一化处理，限定输出范围为[0.5, 0.6]。表1展示了标准化后数据集。通过构建雪道难度影响评估模型，可量化赛道几何特征对比赛成绩的影响强度。通过特征重要性权重与历史比赛成绩趋势的对比分析，验证模型预测结果与实际赛事数据的吻合度，结果表明最大坡度与弯道数量对基准时间的影响最为显著，模型输出结果与历史成绩趋势一致，验证了方法的有效性。

表1 雪道难度影响评估模型

地点	名次	赛道难度系数	预测时间	赛道加权指标
平昌	1	0.5	41.49	20.745
平昌	2	0.5	48.9625	24.48125
平昌	3	0.5	49.095	24.5475
延庆	1	0.592790456	58.575	34.72270096
延庆	2	0.592790456	58.6675	34.77753407
延庆	3	0.592790456	58.7725	34.83977707
惠斯勒	1	0.581635442	51.115	29.73029561
惠斯勒	2	0.581635442	51.21	29.78555098
惠斯勒	3	0.581635442	51.2125	29.78700506
索契	1	0.6	55.1725	33.1035
索契	2	0.6	55.2475	33.1485
索契	3	0.6	55.275	33.165

2.2 冬奥会男子四人雪车成绩建模与验证

基于近四届冬奥会男子四人雪车冠军的最优轮次成绩数据，本研究采用灰色GM(1,1)模型进行建模。为适应冬奥会成绩的小样本特性，模型通过累加生成机制强化数据规律性，弱化随机波动对预测的干扰。通过累减还原得到历史拟合值，模型成功捕捉到四届冬奥会冠军成绩的缓升趋势。

将k=4 代入时间响应式，预测第米兰冬奥会赛道加权指标：

$$\hat{x}^{(1)}(5) = -2006.1030e^{-0.0165\times4} + 2026.8480 \approx 154.632$$

累减得：

$$\hat{x}^{(0)}(5) = 154.632 - 122.119 = 32.513$$

结合米兰赛道难度系数0.5543，预测滑行时间：

$$\text{TIME} = \frac{32.513}{0.5543} \approx 58.66 \text{ 秒}$$

残差检验

将 $k = 0,1,2,3$ 代入模型，计算累加序列拟合值并累减还原：

$$\hat{x}^{(1)} = \{20.745,55.4677,85.1980,122.1190\}$$

$$\hat{x}^{(1)} = \{20.745,34.7227,29.7303,36.9210\}$$

绝对误差序列：

$$\Delta^{(0)} = \{0,0,0,3.8175\}$$

相对误差序列：

$$\Phi = \{0\%, 0\%, 0\%, 11.53\%\}$$

对历史数据拟合值与实际值的残差分析表明，绝对误差序列最大值为3.8175秒，平均相对误差为2.88%，远低于5%的预设阈值，满足预测精度要求。

关联度检验

关联系数计算(分辨率 ρ=0.5)：

$$\eta(k) = \frac{0.5 \times 3.8175}{\Delta^{(0)}(k) + 0.5 \times 3.8175}$$

得关联系数 $\eta = \{1, 1, 1, 0.333\}$，关联度：

$$\gamma = \frac{1}{4}(1 + 1 + 1 + 0.333) = 0.833$$

计算关联系数（分辨率 ρ=0.5），得到关联度 γ=0.833（>0.6），验证模型趋势与原始数据的高度一致性。

后验差检验原始序列标准差 $S_1 = 6.214$，残差序列标准差 $S_2 = 19.09$，方差比：

$$C = \frac{S_2}{S_1} = 0.307$$

小误差概率 $P = 1(\text{所有残差}|\Delta^{(0)}(k)| < 0.6745S_1)$。根据灰色模型精度分级标准（表2），模型为一级，表明其具备较高的预测可靠性。

表2 灰色模型预测精度等级表

模型精度等级	P	C
一级（好）	≥0.95	≤.35
二级（合格）	≥.80	≤0.50

模型检验结果显示，平均相对误差为2.88%，最大相对误差为11.53%，残差检验、关联度检验（关联度0.833）及后验差检验（方差比0.307）均满足灰色模型一级精度标准，证实其在小样本场景下的可靠性。

表3 米兰冬奥会男子四人雪车项目成绩预测

男子四人雪车项目	赛道加权指标	成绩/秒
第一名	32.513	58.66
第二名	33.661	60.72
第三名	33.807	60.99

预测的成绩能否出现有很多因素决定，因为运动员在比赛过程中受到的干扰因素是不确定的，同时运动员能否通过赛前的训练调整在赛场上发挥自己的全部实力也是未知的，所以预测成绩只是反映了在下届比赛中可能达到的成绩值。通过灰色预测模型对米兰冬奥会前三名最优轮次的成绩进行预测，结果如表3所示，男子四人雪车项目前三名最优轮次的预测值分别是58.66，60.72，60.99。

3 分析与讨论

研究通过两阶段建模验证这一假设：首先利用随机森林量化赛道参数对赛道加权指标的影响，再基于灰色模型外推未来趋势。对比单一模型发现，仅采用随机森林回归时，受限于12组数据的小样本量，模型泛化能力不足；而单独使用灰色模型虽能捕捉长期趋势，却无法解析赛道特征的微观影响机制。

OPEN ACCESS

ISSN 3105-7594（P）
ISSN 3105-7608（O）

3.1 创新性

雪车项目因其赛道如弯道曲率、垂直落差、坡度差异等物理特征的独特性，使得不同赛道间的横向比较存在显著挑战。现有研究多通过单一指标赛道长度进行关联分析，难以全面量化赛道复杂度对成绩的影响。本研究创新性地引入随机森林回归模型，将赛道参数与滑行时间进行加权融合，构建赛道难度系数与滑行时间的乘积赛道加权指标，以此作为核心预测变量。为赛道适应性分析提供了可量化的决策依据。此方法突破了传统研究中赛道作为单一自变量的局限性，实现了多维度数据的协同建模，显著提升了预测模型的解释力。

灰色预测模型在竞技体育中的应用多集中于周期性显著的夏季项目，而冬季项目因赛道环境复杂、数据采集困难等因素，相关研究较为匮乏。本研究首次将灰色GM(1,1)模型引入男子四人雪车项目的成绩预测，通过累加生成序列（AGO）弱化小样本数据的随机波动，成功挖掘出四届冬奥会冠军成绩的隐含趋势规律。模型结果显示，赛道加权指标的灰色发展系数为0.0165，表明雪车竞技水平呈现缓慢上升趋势，与近年来运动员滑行技术优化和装备改进的实践进展相吻合。研究进一步结合米兰赛道已知参数，将预测结果外推至2026年冬奥会，填补了冬季竞速类项目长期预测的方法空白，为钢架雪车、雪橇等其他冰雪项目的类似研究提供了方法论借鉴。

3.2 局限性

尽管随机森林算法在抗过拟合与特征重要性评估方面表现优异，但本研究受限于冬奥会男子四人雪车项目仅四届赛事的前三名成绩，共12组样本的历史数据规模，导致模型训练集容量不足。小样本条件下，Bootstrap重抽样策略难以充分覆盖数据分布的多样性，可能引入抽样偏差。此外，赛道参数中部分变量因测量标准不统一，存在一定程度的噪声干扰。未来需纳入更多国际A级赛事数据（如世界杯分站赛）以扩充样本量，并通过交叉验证优化超参数选择，进一步提升模型泛化能力。赛道类型划分仅作为理论分类框架，实际建模中未引入分类标签作为输入特征。

2026年米兰冬奥会赛道尚处于建设阶段，官方未公布完整的物理参数。本研究采用的赛道难度系数（0.5543）源于试滑认证运动员的场地手册，仅包含起始高度与终点高度的实测值，其余参数依赖人工几何推算。此类推算可能忽略实际施工中的局部调整，导致WEIGHT值的预测结果存在系统偏差。因此，建议在赛道竣工后，通过实地传感器采集高精度轨迹数据，对预测模型进行动态校准，以提高战术制定的可靠性。

3.3 学术意义

本研究通过融合随机森林与灰色GM(1,1)模型，构建了适用于小样本场景的混合预测框架，其方法论创新主要体现在两方面：其一，通过加权指标化解赛道异质性难题，为冬季项目的横向比较提供了新思路；其二，灰色模型的累加生成机制有效捕捉了雪车成绩的长期趋势，避免了传统时间序列模型对大数据量的依赖。研究提出的“赛道-成绩”关联模型可扩展至其他竞速类项目，助力我国冰雪运动的科学化训练体系建设。

4 结论

本研究通过集成随机森林回归与灰色GM(1,1)模型，系统分析了冬奥会男子四人雪车项目的成绩影响因素，并成功预测了2026年米兰冬奥会的竞技趋势。随机森林模型揭示了赛道垂直落差与弯道数量对总成绩的主导作用，而灰色预测则捕捉到成绩的长期缓升趋势，与技术进步实践高度吻合。创新性提出的赛道加权指标实现了赛道参数与滑行效率的多维度融合，为冬季项目的横向比较提供了新思路。尽管模型在米兰赛道的预测中展现出较高可靠性，但仍需关注小样本偏差与赛道竣工后的参数校准问题。未来研究可通过纳入世界杯分站赛数据扩充样本量，并结合物联网技术实时采集冰面摩擦系数等动态参数，进一步提升预测精度。

本研究不仅为我国雪车项目备战米兰冬奥会提供了科学依据，其方法论框架还可推广至其他冰雪竞速项目，推动竞技体育数据分析从经验驱动向模型驱动的转型，助力我国冬季运动在国际赛场实现更大突破。

参考文献

[1] International Bobsleigh & Skeleton Federation. Pre-homologation of Cortina d’Ampezzo’s Sliding Centre concludes successfully[EB/OL]. (2025-03-28)[2025-04-22].

[2] 石慧敏, 章东迎, 章永辉. 奥运奖牌可以被预测吗？——基于可解释机器学习视角[J]. 上海体育学报, 2024, 48(4): 26-36. doi:10.16099/j.sus.2023.10.27.0002.

[3] 尹一全, 邱森, 孟庆军, 等. 钢架雪车运动员竞赛成绩的影响因素研究[J]. 北京体育大学学报, 2020, 43(12): 130-141.

[4] 冬奥播报丨女子单人雪车怀明明第6创中国该项目冬奥参赛最好成绩[N/OL]. 人民网, 2022-02-14[2024-06-20]. http://ent.people.com.cn/n1/2022/0214/c1012-32351778.html.

[5] Sun X, Wang Y, Khan J. Hybrid LSTM and GAN model for action recognition and prediction of lawn tennis sport activities[J]. Soft Computing, 2023.

[6] 陈岷. 因子分析和神经网络相融合的体育成绩预测模型[J]. 现代电子技术, 2017, 40(5): 130-133. doi:10.16652/j.issn.1004-373x.2017.05.033.

[7] 修霆喆, 于红妍, 黄雯妍. 基于BP神经网络的乒乓球优秀女子单打比赛结果预测模型构建及仿真应用[J]. 哈尔滨体育学院学报, 2024, 42(2): 89-96. doi:10.16099/j.sus.2023.10.27.0002.

[8] GAO Z, KOWALCZYK A. Random forest model identifies serve strength as a key predictor of tennis match outcome[J]. Journal of Sports Analytics, 2021, 7: 255-262. doi:10.3233/JSA-200515.

[9] MARQUÉS-JIMÉNEZ D, RAYA-GONZÁLEZ J, SÁNCHEZ-DÍAZ S, et al. A random forest clustering to explore the influence of physical fitness level of youth basketball players on match-related physical performance[J]. International Journal of Sports Science & Coaching, 2022, 17(3): 567-578. doi:10.1177/17479541211041725.

[10] BRIAND J, DEGUIRE S, GAUDET S, et al. Monitoring variables influence on random forest models to forecast injuries in short-track speed skating[J]. Frontiers in Sports and Active Living, 2022, 4: 789012. doi:10.3389/fspor.2022.789012.

[11] LIU J, CARR P. Detecting and tracking sports players with random forests and context-conditioned motion models[EB/OL]. (2024-01-16)[2024-06-20]. https://doi.org/10.21203/rs.3.rs-3859351/v1.

[12] CHEN Z. Application of random forest algorithm based on fiber optic sensors in fatigue detection in sports competitions[EB/OL]. (2024-01-16)[2024-06-20]. https://doi.org/10.21203/rs.3.rs-3859351/v1.

[13] LI Y, MU Y. Research and performance analysis of random forest-based feature selection algorithm in sports effectiveness evaluation[C]//Proceedings of the 2023 International Conference on Artificial Intelligence in Sports. Beijing: IEEE, 2023: 1-6. doi:10.1109/ICAIS58442.2023.00010.

[14] 沈杰.基于灰色理论的可靠性寿命预测研究[D].西华师范大学,2024.

[15] 蔡素丽.灰色系统理论GM(1,1)预测模型应用实证分析[J].廊坊师范学院学报(自然科学版),2015,15(6):5-8,12.

[16] 魏春玲,孙晋海.奥运会田径项目成绩发展趋势及灰色预测研究[J].中国体育科技,2005,41(2):18-20.

[17] 胡文强,李思民.世锦赛视角下标枪竞争格局和成绩演进研究[J].辽宁体育科技,2020,42(4):97-101.

[18] 龙家勇, 蔚卓虹. 基于GM(1,1)灰色模型对奥运会男子100m金牌成绩的预测研究[J]. 西南师范大学学报(自然科学版), 2023, 48(7): 123-128. doi:10.13718/j.cnki.xsxb.2023.07.018.

[19] Wang T Y, Cui J, Fan Y. A wearable-based sports health monitoring system using CNN and LSTM with self-attentions[J]. PLOS ONE, 2023. doi:10.1371/journal.pone.0283212.

OPEN ACCESS