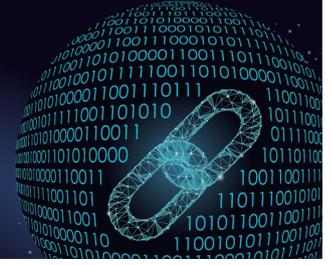




中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 基于面板数据的灰色差值关联模型构建及其应用

孙婧, 党耀国, 杨圣祥, 王俊杰, 杨绍闻

引用本文:

孙婧, 党耀国, 杨圣祥, 王俊杰, 杨绍闻. 基于面板数据的灰色差值关联模型构建及其应用[J]. 控制与决策, 2024, 39(11): 3839–3847.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0354>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 区间条件下基于GRA和TOPSIS的辐射源威胁评估

Combining TOPSIS and GRA for emitter threat evaluation with interval number

控制与决策. 2021, 36(6): 1516–1522 <https://doi.org/10.13195/j.kzyjc.2019.1559>

#### 考虑主体心理行为的三边单向非循环稳定匹配

Three-sided stable matching with one-way acyclic preference considering agent's psychological behavior

控制与决策. 2021, 36(3): 741–746 <https://doi.org/10.13195/j.kzyjc.2019.1047>

#### 基于群决策考虑属性效用一致性的DEA他评交叉效率公共权重排序法

A common-weight ranking method for DEA peer-efficiency based on group decision-making and considering the consistency of attribute utility

控制与决策. 2021, 36(9): 2279–2289 <https://doi.org/10.13195/j.kzyjc.2019.1719>

#### 带有输出约束的柔性关节机械臂预设性能自适应控制

Prescribed performance adaptive control of flexible-joint manipulators with output constraints

控制与决策. 2021, 36(2): 387–394 <https://doi.org/10.13195/j.kzyjc.2019.0974>

#### 基于改进GNG算法的燃煤锅炉数据动态特征分析与控制

Dynamic characteristics analysis and control of coal-fired boiler based on improved GNG algorithm

控制与决策. 2021, 36(8): 1855–1861 <https://doi.org/10.13195/j.kzyjc.2019.1343>

# 基于面板数据的灰色差值关联模型构建及其应用

孙 婧<sup>1</sup>, 党耀国<sup>1</sup>, 杨圣祥<sup>2</sup>, 王俊杰<sup>1†</sup>, 杨绍闻<sup>1</sup>

(1. 南京航空航天大学 经济与管理学院, 南京 210016;

2. 德蒙福特大学 计算机科学与信息学院, 莱斯特 LE19BH)

**摘要:** 针对现有面板数据灰色关联模型在保序性、稳定性和正负性等方面存在的问题, 利用差值特征和指数函数构建灰色关联系数, 提出方向判断因子来表征正负相关性, 设计波动调节因子以刻画序列间的相对波动程度, 并利用熵权法确定面板数据的指标权重, 提出适用于面板数据的灰色差值关联模型. 此外, 给出灰色关联分析中反对称性和排列保序性的定义, 并证明新模型具有反对称性和排列保序性等性质. 最后, 将该模型应用于京津冀地区 2013~2021 年大气污染的趋势分析, 表明模型结果的实用性; 通过两组蒙特卡洛实验与传统模型进行对比, 验证新模型具有稳定性和排列保序性, 更适用于对象排列具有不固定性的面板数据.

**关键词:** 灰色关联分析; 面板数据; 差值; 保序性; 稳定性

中图分类号: N941.5

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0354

**引用格式:** 孙婧, 党耀国, 杨圣祥, 等. 基于面板数据的灰色差值关联模型构建及其应用 [J]. 控制与决策, 2024, 39(11): 3839-3847.

## Grey difference incidence model of panel data and its application

SUN Jing<sup>1</sup>, DANG Yao-guo<sup>1</sup>, YANG Sheng-xiang<sup>2</sup>, WANG Jun-jie<sup>1†</sup>, YANG Shao-wen<sup>1</sup>

(1. College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. School of Computer Science and Informatics, De Montfort University, Leicester LE19BH, United Kingdom)

**Abstract:** There are three problems such as order preservation, stability and positive and negative in the grey incidence model applicable to panel data. To solve these problems, this paper has the following efforts. Firstly, the grey incidence coefficient is constructed based on the difference characteristics and exponential function, and the direction judgment factor is proposed to represent the positive and negative correlation. Secondly, the fluctuation adjustment factor is designed to describe the relative fluctuation degree between sequences. Thirdly, indicator weights are determined using the entropy method. A grey difference incidence model for panel data is proposed. In addition, antisymmetry and order-preserving property are defined in grey incidence analysis, and the new model is proved to have antisymmetry and order-preserving property. Finally, the model is applied to analyze the trend of air pollution in the Beijing-Tianjin-Hebei region from 2013 to 2021, which shows the practicability of the model. Through two sets of Monte Carlo experiments, the robustness and order-preserving property of the model are verified. The model is more suitable for the panel data with non-fixed object arrangement.

**Keywords:** grey incidence analysis; panel data; difference characteristics; order-preserving property; robustness

## 0 引言

大气污染严重影响人体健康<sup>[1]</sup>、气候环境<sup>[2]</sup>和经济社会发展<sup>[3]</sup>, 已成为世界城市居民生活中难以逃避的现实问题. 京津冀地区作为我国的“首都经济圈”, 2022 年经济总量和人口数量分别突破 10 万亿元和 1.1 亿人, 在社会经济和人力资源方面展现出极强的

发展潜力, 拥有举世瞩目的发展成就, 但同时京津冀地区的大气污染问题也同样备受关注<sup>[4]</sup>. 大气污染数据具有突变性和波动性等不确定性特征, 数据信息含量较少. 针对此类数据的相关性分析, 传统统计方法效果往往不尽人意. 灰色关联分析作为灰色系统理论的重要分支, 已被广大学者认为是获取“贫信息”

收稿日期: 2023-03-26; 录用日期: 2024-01-18.

基金项目: 国家自然科学基金项目 (72271120, 72001107, 71771119); 中国博士后科学基金项目 (2020T130297, 2019M660119).

责任编辑: 梁樑.

†通讯作者. E-mail: wangjj@nuaa.edu.cn.

\*本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

数据间关联性的一种有效方法。

灰色关联模型具有应用背景多元化、模型性质多样化和构建视角多维化的特征。在应用背景方面,涉及了大气污染、碳排放、经济社会发展等领域。其中:贺祥等<sup>[5]</sup>运用灰色关联模型判别江苏省PM<sub>2.5</sub>浓度的影响因素;Huang等<sup>[6]</sup>利用灰色关联模型确定影响碳排放的主要因素;蒋珂等<sup>[7]</sup>根据时延参数重构后的时序关联矩阵,采用灰色关联分析量化其多重相关性;黄仁全<sup>[8]</sup>以灰色关联模型为工具,分析黄河流域在经济发展、生态环境和科技创新之间存在的问题。在模型性质方面,大多聚焦于正向关联模型的性质研究。其中:韦保磊等<sup>[9]</sup>定义了三类关联算子的幂等、线性和仿射变换性质;魏勇等<sup>[10]</sup>指出关联度的整体性和偶对称性无必要;彭张林等<sup>[11]</sup>提出应注重对评价方法稳健性的研究。在模型构建视角方面,诸多学者从斜率<sup>[12]</sup>、振幅<sup>[13]</sup>、凸度<sup>[14]</sup>、高度差<sup>[15]</sup>、面积<sup>[16]</sup>和向量模<sup>[17]</sup>等不同的几何视角构建灰色关联模型,极大地丰富了灰色关联分析的建模思想。

近年来,考虑到区域大气污染程度分析、能源消费情况解析和经济发展水平评判等诸多现实问题,均需要以面板数据为代表的高维数据参与,广大学者对适用于面板数据的灰色关联模型进行了积极探索。张可等<sup>[18]</sup>将面板数据展示为三维空间中的点,利用曲面簇体积差构建拓展灰色绝对关联度;吴鸿华等<sup>[19]</sup>将面板数据在时间和指标维度分别投射为离散曲线,构建灰色曲率关联模型。上述两个模型拓展了面板数据灰色关联模型的构建视角,但当研究对象或指标排列顺序改变时,关联序可能发生改变。刘震等<sup>[20]</sup>利用三维空间中的网格描述面板数据,并以网格中独立线段的斜率为基础,构建可以表征正负相关性的灰色网格关联度模型;李雪梅等<sup>[21]</sup>设计平均生成速率序列,构建灰色指标关联聚类模型;Wang等<sup>[22]</sup>将三角函数引入模型构建,利用发展趋势关联测度因子和趋势正负判断因子构建灰色变趋势关联模型;党耀国等<sup>[23]</sup>利用增量和离差分别表征时间维度上的差异特征和对象维度上的分布特征,构建了灰色指标关联模型。上述4个面板数据灰色关联模型,不仅拓宽了模型的构建视角,而且在灰色关联的正负性表征方面做了积极探索,但模型存在关联度稳定性较低的问题。

现有面板数据灰色关联模型在评价空气质量时,主要存在以下问题:1)在大气污染研究中,研究对象多为城市或检测站点,其在面板数据中的排列顺序具有不固定性。而现有面板数据灰色关联模型大多

存在城市排列顺序改变将引起关联度排序结果改变的问题,即不具备排列保序性。2)灰色关联度的正负性对判断污染程度具有极强的现实意义,若关联度为正,则表示所选城市(或站点)的空气质量优于对标城市(或站点);反之亦然。但由于现有模型未针对空气质量的优劣判断特点进行建模,使其关联度符号无法作为空气质量优劣性的判断依据。3)灰色关联系数和灰色关联度分别表示空气质量在序列局部和整体的关联情况,但现有模型在计算灰色关联度时忽视了关联系数的正负抵消问题,易出现数据轻微变动就会导致关联度结果显著变化的情况,即模型稳定性较低。

基于以上问题,本文以面板数据为例,构建灰色差值关联模型,更新灰色关联模型构建步骤,提出反对称性和排列保序性在灰色关联分析中的定义,推动现有灰色关联模型的改进和模型性质的完善。在模型构建过程中,以差值绝对值为差异度量基本单元确定关联系数大小,并提出方向判断因子判断关联方向;离散提取差值特征,消除由于对象排列顺序改变对关联度的影响。此外,引入波动调节因子对关联度进行修正,强化灰色关联模型对波动序列的刻画能力;最后,分别计算关联度的符号和大小,解决正负灰色关联系数相互抵消导致关联度稳定性较低的问题。将新模型应用于京津冀地区大气污染的整体趋势分析,并与传统模型进行对比,通过两类重复实验验证模型的排列保序性和稳定性。

## 1 面板数据灰色差值关联模型的构建及性质

### 1.1 面板数据表征及规范化

**定义1** 设数据集 $X_m$ 在 $k$ 时刻关于对象 $m$ 的第 $n$ 个指标值为 $x'_m(n, k)$ ,  $k = 1, 2, \dots, K$ ,  $n = 1, 2, \dots, N$ ,  $m = 1, 2, \dots, M$ , 则称

$$X_m = \begin{bmatrix} x'_m(1, 1) & x'_m(1, 2) & \dots & x'_m(1, K) \\ x'_m(2, 1) & x'_m(2, 2) & \dots & x'_m(2, K) \\ \vdots & \vdots & \ddots & \vdots \\ x'_m(N, 1) & x'_m(N, 2) & \dots & x'_m(N, K) \end{bmatrix}$$

为第 $m$ 个对象的原始面板数据。称 $X'_m(n) = (x'_m(n, 1), x'_m(n, 2), \dots, x'_m(n, K))$ 为对象 $m$ 关于指标 $n$ 的原始时间序列; $X'(n, k) = (x'_1(n, k), x'_2(n, k), \dots, x'_M(n, k))$ 为指标 $n$ 在时刻 $k$ 的原始截面数据。

大气污染浓度等成本型指标可按下面式(1)或式(2)对原始数据进行规范化处理,人均GDP等效益型指标可按下面式(3)对原始数据进行规范化处理:

$$x_m(n, k) = \frac{x'_m(n, k)}{\max_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k)}, \tag{1}$$

$$x_m(n, k) = \frac{\max_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k) - x'_m(n, k)}{\max_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k) - \min_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k)}, \tag{2}$$

$$x_m(n, k) = \frac{x'_m(n, k) - \min_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k)}{\max_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k) - \min_{k \in \{1, 2, \dots, K\}, m \in \{1, 2, \dots, M\}} x'_m(n, k)}. \tag{3}$$

将规范化后的面板数据、时间序列和截面数据分别表示为  $X_m$ 、 $X_m(n)$  和  $X(n, k)$ , 以下模型构建过程中使用的数据均为规范化后的数据.

### 1.2 面板数据灰色差值关联模型的构建

#### 1.2.1 提取差值特征

面板数据差值特征的提取如图 1 所示. 其中:  $A$ 、 $B$  和  $C$  均代表在时刻  $k$  对象  $m_1$ 、 $m_2$  和  $m_3$  关于指标  $n$  的指标值, “+” 和 “-” 表示  $B$  和  $C$  两点分别从大于和小于  $A$  点的方向判断与  $A$  点的差异性特征.

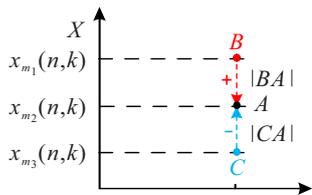


图 1 差值特征提取示意

**定义 2** 令  $\Delta_{m_1, m_2}(n, k) = x_{m_1}(n, k) - x_{m_2}(n, k)$  和  $|\Delta_{m_1, m_2}(n, k)| = |-\Delta_{m_2, m_1}(n, k)| = |x_{m_1}(n, k) - x_{m_2}(n, k)|$  是对象  $m_1$  和  $m_2$  关于指标  $n$  在  $k$  时刻的差值和差值绝对值, 则面板数据中时间序列的差值和差值绝对值可表示为  $\Delta_{m_1, m_2}(n) = (\Delta_{m_1, m_2}(n, 1), \Delta_{m_1, m_2}(n, 2), \dots, \Delta_{m_1, m_2}(n, K))$  和  $|\Delta_{m_1, m_2}(n)| = (|\Delta_{m_1, m_2}(n, 1)|, |\Delta_{m_1, m_2}(n, 2)|, \dots, |\Delta_{m_1, m_2}(n, K)|)$ ; 两个面板数据  $X_{m_1}$  与  $X_{m_2}$  的差值以及差值绝对值可以表示为  $\Delta_{m_1, m_2}$  和  $|\Delta_{m_1, m_2}|$ .

需要注意的是, 虽然用来评判差异特征的  $\Delta_{m_1, m_2}(n, k)$  和  $\Delta_{m_2, m_1}(n, k)$  绝对值大小相同, 但二者互为相反数, 表明关联方向不同.

**定义 3** 设  $x_{m_2}(n, k)$  为对标值, 若  $T_{m_1, m_2} = 1$ , 则表示  $x_{m_1}(n, k)$  从大于  $x_{m_2}(n, k)$  的方向判断与  $x_{m_2}(n, k)$  的差异特征; 若  $T_{m_1, m_2} = -1$ , 则表示  $x_{m_1}(n, k)$  从小于  $x_{m_2}(n, k)$  的方向判断与  $x_{m_2}(n, k)$  的差异特征. 因此称

$$T_{m_1, m_2}(n, k) = \begin{cases} 1, & \Delta_{n_1, n_2}(n, k) \geq 0; \\ -1, & \Delta_{n_1, n_2}(n, k) < 0 \end{cases} \tag{4}$$

为  $x_{m_1}(n, k)$  与  $x_{m_2}(n, k)$  间的方向判断因子.

在上述的差值特征提取中, 差值绝对值作为差异度量基本单元, 用以量化灰色关联系数大小, 方向判断因子则可判断关联方向. 在确保关联度排序结果不受影响的前提下, 为提高关联系数和关联度的分辨性<sup>[18]</sup>, 增强排序结果的可信度, 本文以指数函数为基础构建新模型.

#### 1.2.2 构建灰色关联系数

**定义 4** 设对象  $m_1$  和  $m_2$  关于  $N$  个指标和  $K$  个时刻的面板数据为  $X_{m_1}$  和  $X_{m_2}$ , 称

$$\gamma_{m_1, m_2}(n, k) = T_{m_1, m_2}(n, k) \cdot e^{-\alpha \cdot |\Delta_{m_1, m_2}(n, k)|} \tag{5}$$

为对象  $m_1$  和  $m_2$  关于指标  $n$  在  $k$  时刻的关联系数. 其中  $\alpha$  ( $\alpha \in (0, 1]$ ) 为距离度量调节因子, 可对差值在关联系数中的作用程度进行调节.

传统关联模型直接对关联系数加权平均计算关联度, 此计算方式易导致关联度稳定性较低. 本文对关联度的计算方式进行了如下改进.

#### 1.2.3 计算灰色关联度

**定义 5** 设对象  $m_1$  和  $m_2$  关于指标  $n$  时间序列间的方向判断因子为

$$T_{m_1, m_2}(n) = \begin{cases} 1, & \sum_{t=1}^K \Delta_{m_1, m_2}(n, k) \geq 0; \\ -1, & \sum_{t=1}^K \Delta_{m_1, m_2}(n, k) < 0. \end{cases} \tag{6}$$

依据相同的构建思想, 面板数据关于对象  $m_1$  和  $m_2$  的方向判断因子可表示为

$$T_{m_1, m_2} = \begin{cases} 1, & \sum_{t=1}^K \sum_{n=1}^N \Delta_{m_1, m_2}(n, k) \geq 0; \\ -1, & \sum_{t=1}^K \sum_{n=1}^N \Delta_{m_1, m_2}(n, k) < 0. \end{cases} \tag{7}$$

**定义 6** 灰色差值关联度由方向判断因子和关联度大小构成. 在面板数据中, 对象  $m_1$  和  $m_2$  关于指标  $n$  时间序列间的关联度可定义为

$$\phi_{m_1, m_2}(n) = T_{m_1, m_2}(n) \cdot \overline{|\gamma_{m_1, m_2}(n)|} \cdot e^{-\beta \cdot \sigma(\Delta_{m_1, m_2}(n))}. \tag{8}$$

其中:  $|\overline{\gamma_{m_1, m_2}(n)}| = \sum_{t=1}^K |\gamma_{m_1, m_2}(n, k)| / K$ ;  $\beta$  ( $\beta \in [0, 1]$ ) 为时间序列的波动调节因子;  $\sigma(\Delta_{m_1, m_2}(n))$  表示对象  $m_1$  和  $m_2$  关于指标  $n$  时间序列差值的标准差, 其主要作用是表征时间序列间的相对波动程度, 并通过  $e^{-\beta \cdot \sigma(\Delta_{m_1, m_2}(n))}$  增强关联度对时间序列相对波动的刻画能力;  $|\overline{\gamma_{m_1, m_2}(n)}| \cdot e^{-\beta \cdot \sigma(\Delta_{m_1, m_2}(n))}$  为对象  $m_1$  和  $m_2$  关于指标  $n$  时间序列关联度大小.

在计算面板数据关联度时, 采用熵权法获取各指标权重, 消除由于指标权重主观判定对关联结果产生的影响. 首先, 根据面板数据的规范化结果求各指标的信息熵为

$$E(n) = -\ln(K \cdot M)^{-1} \sum_{k=1}^K \sum_{m=1}^M p_m(n, k) \ln(p_m(n, k)),$$

其中  $p_m(n, k) = x_m(n, k) / \sum_{k=1}^K \sum_{m=1}^M x_m(n, k)$ . 若  $p_m(n, k) = 0$ , 则令  $\lim_{p_m(n, k) \rightarrow 0} p_m(n, k) \ln(p_m(n, k)) = 0$ . 然后, 获取指标权重序列  $w = (w_1, w_2, \dots, w_N)$ ,  $w_n = (1 - E(n)) / (N - \sum_{n=1}^N E(n))$ .

在考虑指标权重后, 对象  $m_1$  与  $m_2$  间的关联度可定义为

$$\phi_{m_1, m_2} = T_{m_1, m_2} \cdot |\overline{\gamma_{m_1, m_2}}| \cdot e^{-\delta \cdot \sigma(\Delta_{m_1, m_2})}. \quad (9)$$

其中:  $|\overline{\gamma_{m_1, m_2}}| = \sum_{k=1}^K \sum_{n=1}^N w_n \cdot |\gamma_{m_1, m_2}(n, k)| / K = \sum_{n=1}^N w_n \cdot |\overline{\gamma_{m_1, m_2}(n)}|$ ;  $\delta$  ( $\delta \in [0, 1]$ ) 为面板数据的波动调节因子;  $\sigma(\Delta_{m_1, m_2})$  表示面板数据关于对象  $m_1$  与  $m_2$  之间差值的标准差, 表征两个面板数据间的相对波动程度, 然后通过  $e^{-\delta \cdot \sigma(\Delta_{m_1, m_2})}$  增强关联度对面板数据间相对波动的刻画能力;  $|\overline{\gamma_{m_1, m_2}}| \cdot e^{-\delta \cdot \sigma(\Delta_{m_1, m_2})}$  为对象  $m_1$  与  $m_2$  间的关联度大小.

本文主要通过离散形式获取差异特征, 所以新模型不仅适用于两个对象间的面板数据关联研究, 而且适用于不同时刻间的数据关联情况测度. 模型中波动调节因子 ( $\beta$  和  $\delta$ ) 的取值主要根据评价者对数据集间相对波动的重视程度来选取. 若不考虑序列间相对波动对关联度的影响, 则可能出现差值绝对值相同, 但差值符号不完全相同的两数据集间灰色关联度相同的情况, 影响模型分析效果.

### 1.3 面板数据灰色差值关联模型的性质

本节先给出反对称性和排列保序性在灰色关联分析中的定义, 然后对新模型具备性质进行证明.

**定义7** 设任意等长度序列  $X_i$  和  $X_j$ , 若任意灰色关联系数均满足  $\gamma(x_i, x_j) = -\gamma(x_j, x_i)$ , 灰色关联度满足  $\phi(X_i, X_j) = -\phi(X_j, X_i)$ , 则称此灰色关联模型满足反对称性.

**定义8** 设  $X_l$ ,  $X_i$  和  $X_j$  为对象(或指标)排列顺序可任意变化的等长序列, 其中  $X_l$  为参考序列,  $X_i$  和  $X_j$  为相关序列, 改变对象(或指标)排列顺序后的序列分别为  $X'_l$ ,  $X'_i$  和  $X'_j$ . 若对于任意对象(或指标)排列顺序均有  $\phi(X_i, X_l) > \phi(X_j, X_l) \Rightarrow \phi(X'_i, X'_l) > \phi(X'_j, X'_l)$ , 则称此灰色关联模型满足排列保序性.

#### 性质1

1)  $\gamma_{m_1, m_2}(n, k) \in (-1, 0) \cup (0, 1]$ ,  $\gamma_{m_1, m_2}(n) \in (-1, 0) \cup (0, 1]$ ,  $\gamma_{m_1, m_2}(k) \in (-1, 0) \cup (0, 1]$ ,  $\phi_{m_1, m_2} \in (-1, 0) \cup (0, 1]$ ;

2)  $\gamma_{m_1, m_2}(n, k) = -\gamma_{m_2, m_1}(n, k)$ ,  $\gamma_{m_1, m_2}(n) = -\gamma_{m_2, m_1}(n)$ ,  $\gamma_{m_1, m_2}(k) = -\gamma_{m_2, m_1}(k)$ ,  $\phi_{m_1, m_2} = -\phi_{m_2, m_1}$ ;

3)  $\gamma_{m_1, m_2}(n, k) = 1 \Leftrightarrow x_{m_1}(n, k) = x_{m_2}(n, k)$ ,  $\phi_{m_1, m_2} = 1 \Leftrightarrow X_{m_1} = X_{m_2}$ .

**证明** 1) 因为  $\Delta_{n_1, n_2}(n, k) \in (-\infty, +\infty)$ , 所以  $|\Delta_{n_1, n_2}(n, k)| \in [0, +\infty)$ . 又因为  $\alpha \in (0, 1]$ , 当  $T_{m_1, m_2}(n, k) = 1$  时,  $\gamma_{m_1, m_2}(n, k) \in (0, 1]$ ; 当  $T_{m_1, m_2}(n, k) = -1$  时,  $\gamma_{m_1, m_2}(n, k) \in (-1, 0)$ . 所以  $\gamma_{m_1, m_2}(n, k) \in (-1, 0) \cup (0, 1]$ .

性质1的1)中的其他性质与  $\gamma_{m_1, m_2}(n, k) \in (-1, 0) \cup (0, 1]$  中的证明过程类似, 此处不展开.

2) 因为  $\Delta_{m_1, m_2}(n, t) = x_{m_1}(n, t) - x_{m_2}(n, k) = -(x_{m_2}(n, k) - x_{m_1}(n, k)) = -\Delta_{m_2, m_1}(n, k)$ , 所以  $T_{m_1, m_2}(n, k) = -T_{m_2, m_1}(n, k)$ ,  $e^{-\alpha \cdot |\Delta_{m_1, m_2}(n, k)|} = e^{-\alpha \cdot |\Delta_{m_2, m_1}(n, k)|}$ , 因此  $\gamma_{m_1, m_2}(n, k) = -\gamma_{m_2, m_1}(n, k)$ . 性质1的2)中的性质主要体现模型的反对称性,  $\gamma_{m_1, m_2}(n) = -\gamma_{m_2, m_1}(n)$ ,  $\gamma_{m_1, m_2}(k) = -\gamma_{m_2, m_1}(k)$  和  $\phi_{m_1, m_2} = -\phi_{m_2, m_1}$  的证明过程与之类似, 此处不展开.

3) 容易证明, 当且仅当两数据相等时,  $\gamma_{m_1, m_2}(n, k) = 1 \Leftrightarrow x_{m_1}(n, k) = x_{m_2}(n, k)$ .  $\phi_{m_1, m_2} = 1 \Leftrightarrow X_{m_1} = X_{m_2}$  同理可证.  $\square$

**性质2** 当变换对象(或指标)排列顺序时, 不影响灰色关联度, 即模型满足排列保序性.

从新模型的构建过程可以看出, 关联系数和关联度均通过离散数据计算得出, 研究数据在对象(或指标)上的排列顺序变动不会影响关联度.

**性质3** 波动调节因子将在一定程度上降低灰

色关联度.

以时间序列中的灰色关联度为例, 因为  $\beta \in [0, 1], \sigma(\Delta_{m_1, m_2}(n)) \geq 0$ , 所以  $|\phi_{m_1, m_2}(n)|$  是  $\sigma(\Delta_{m_1, m_2}(n))$  的非增函数. 即在  $\beta \neq 0$  且差值绝对值相同的情况下, 数据间的相对波动越大, 关联度越低. 该性质表明数据间的相对波动将在一定程度上降低灰色关联程度.

## 2 京津冀大气污染变化特征分析

### 2.1 研究问题及研究数据

随着“大气十条”和“蓝天保卫战三年行动计划”等大气污染防治政策的实施, 我国大气污染防治工作取得了一定成效, 但主要污染物的改善程度有所差异. 如何判断主要污染物的变化趋势和特征, 确定区域治理的重点污染物, 已成为深入细化大气污染防治措施的重要问题. 此外, 区域内城市排列顺序的不固定性、不同城市在不同时间大气污染优劣性的不唯一性, 给区域大气污染变化特征分析带来了严峻挑战. 本文选取在全国范围内污染相对严重的京津冀地区为研究对象, 利用构建的面板数据灰色差值关联模型判断京津冀地区不同时段间主要污染物的变化特点, 分析京津冀地区各污染物在2013~2021年间的污染趋势,

在验证新模型的可操作性和适用性的同时, 为区域内重点治理污染物的识别提供参考, 为大气污染联防联控提供依据. 本文还通过两组蒙特卡洛重复实验验证新模型的排列保序性和稳定性.

在研究数据及参数选取方面, 依据2012年颁布的《环境空气质量标准》, 选取细颗粒物(PM<sub>2.5</sub>)、可吸入颗粒物(PM<sub>10</sub>)、二氧化硫(SO<sub>2</sub>)、二氧化氮(NO<sub>2</sub>)、一氧化碳(CO)和臭氧(O<sub>3</sub>)六种污染物为主要分析指标, 以京津冀地区13个城市(北京、天津、石家庄、唐山、秦皇岛、邯郸、保定、张家口、承德、沧州、廊坊、衡水和邢台)在2013~2021年的6种污染物浓度年度数据为基础. 在数据来源方面, 2013~2021年京津冀地区6种污染物浓度数据来源于各地环境公报. 此外, 原始数据均按照式(1)进行规范化处理. 在模型参数选取方面, 本文采取折中思想, 调节因子均取0.50, 即  $\alpha = \beta = \delta = 0.50$ .

### 2.2 空气质量变化趋势判断

本节以2013年京津冀地区6种污染物数据为参照, 分析2013~2021年京津冀地区大气污染的年度变化情况. 不同年份污染物与2013年对应污染物的关联度如图2所示.

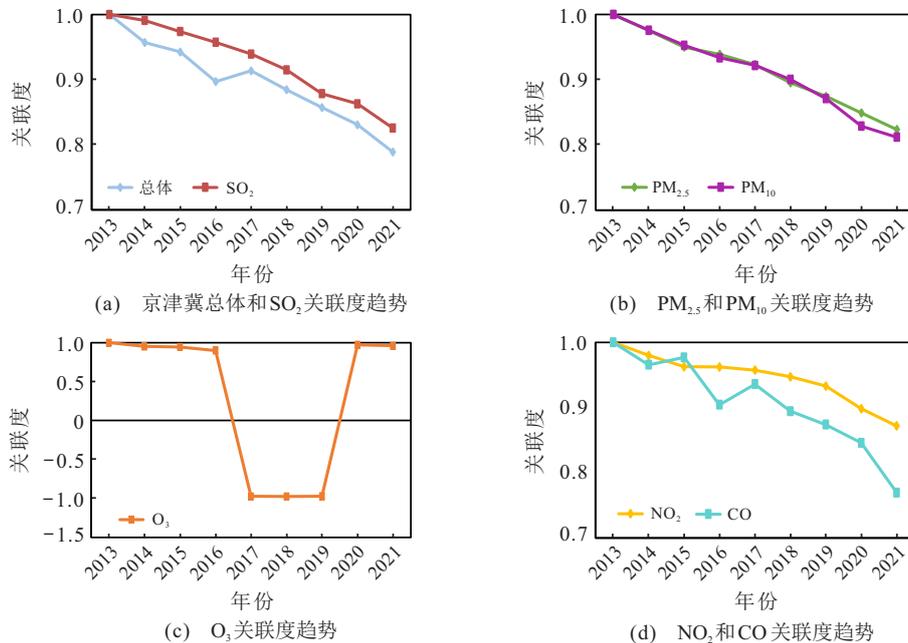


图2 以2013年数据为基准的关联度

首先, 从图2(a)中京津冀地区总体空气质量状况来看, 2013~2021年间空气质量状况明显好转. 由于污染物浓度指标已按照“成本型”指标进行处理, 经过转换的数据值越大代表空气质量越好. 从图2可知, 2014~2021年京津冀地区大部分污染物与2013年的关联度为正数, 且有逐渐降低的趋势. 关

联度符号为正, 表明经过处理的其他年份污染物浓度值整体上高于2013年, 则2014~2021年原始污染物浓度低于2013年, 所以其他年份污染情况较2013年更低. 关联度数值大小有逐渐降低的趋势, 表明2014~2021年间污染情况与2013年相比变化有逐渐增大的倾向. 结合关联度符号为正数, 则可以判断

出2014~2021年间京津冀地区空气质量状况相较于2013年明显好转,与诸多学者<sup>[24-25]</sup>研究结论相同.其次,从不同指标污染情况来看,除O<sub>3</sub>外的其他指标在2013~2021年间均呈现逐渐好转的趋势.从图2(c)可以看出,2017~2019年O<sub>3</sub>与2013年的关联度为负数,表明2017~2019年京津冀地区O<sub>3</sub>污染情况劣于2013年,其他年份优于2013年.从图2(d)可知,CO在2013~2021年间污染情况呈现波动下降态势,其他污染物(SO<sub>2</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>和NO<sub>2</sub>)的情况逐年好转.京津冀地区在持续改善PM<sub>2.5</sub>和PM<sub>10</sub>等污染的同时,应继续加强对O<sub>3</sub>污染防治的重视.

据生态环境局发布信息,2021年京津冀地区PM<sub>2.5</sub>年均浓度首次全面步入“30+”阶段,优良天数比例持续上升,严重污染天数大幅消减,区域生态环境质量持续改善,与本文模型分析结果相符.此外,图3为2021年与2013年污染物浓度原始数据差值的散点图,其中横轴1~13分别代表按北京、天津、石家

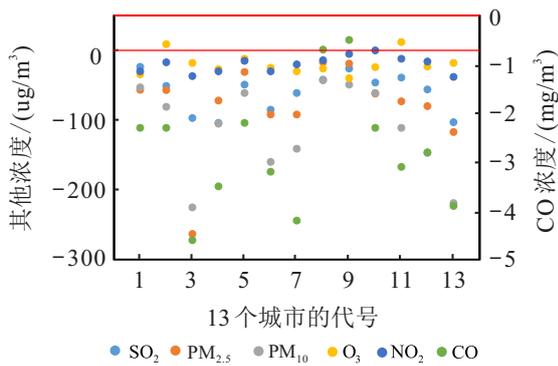


图3 2021年与2013年污染物原始浓度值差值的散点图

庄、唐山、秦皇岛、邯郸、保定、张家口、承德、沧州、廊坊、衡水和邢台顺序排列的13个城市.从图3中可以看出,2021年绝大部分城市的6种污染物浓度均低于2013年,从而验证了上述分析结果的有效性,表明文本模型不仅可以测度关联程度大小,其关联度符号也具有极强的现实意义.

### 2.3 模型排列保序性和稳定性检验

#### 2.3.1 模型排列保序性检验

本节将通过观察对象改变排列顺序前后关联序是否一致,来判断模型结果是否受到对象排列顺序的影响.其中,判断关联序是否一致的指标为Kendall相关系数( $R$ ).当对象排列顺序变动不影响排序结果时, $R = 1$ ;当对象排列顺序变动导致排序结果完全相反时, $R = -1$ ;当对象排列顺序变动前后排序结果完全独立时, $R = 0$ ;其他情况下 $R \in (-1, 0) \cup (0, 1)$ .选取的对比模型为邓氏关联度、接近性关联模型<sup>[18]</sup>、拓展灰色关联模型<sup>[20]</sup>、网格关联模型<sup>[22]</sup>和相似性关联模型<sup>[18]</sup>.

本文共有13个研究对象(城市),因此共有13!种排列顺序可供验证.为节约资源和时间,随机选取500种无重复的对象排列,计算对象顺序改变前后关联序之间的 $R$ ,结果如图4所示.其中,对照的对象排列顺序为“邢台、邯郸、保定、廊坊、天津、北京、石家庄、唐山、沧州、张家口、承德、秦皇岛和衡水”.

从图4可以看出,6个模型中仅有本文模型和邓氏关联度不受对象排列顺序的影响,更适用于对象排列具有不固定性特征的面板数据.其他4种对比模型

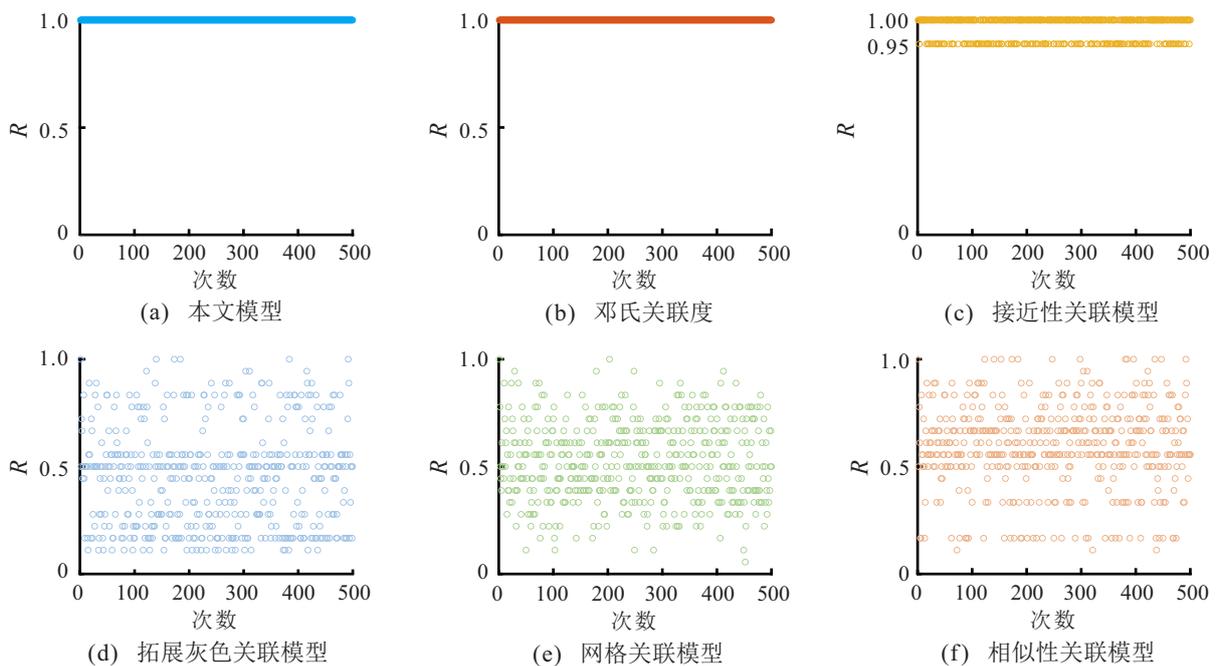


图4 各模型改变对象排列顺序前后关联序之间的R散点图

均出现了对象排列顺序改变导致关联序发生改变的情况,表明这4种模型更适用于研究对象排列顺序相对固定的案例. 模型不满足排列保序性的原因是: 在考虑对象维度/截面数据中的关联度时, 仅将某种对象排列中相邻对象间的特征作为关联程度的判断依据, 未将其他对象间的相关特征加入关联度计算过程中. 解决此问题的方法是: 离散或遍历提取面板数据对象维度的关联特征, 然后计算关联系数和关联度.

### 2.3.2 模型稳定性检验

对于能体现正负相关性的灰色关联模型, 关联度能抵抗数据变动的影响是模型稳定性高的重要体现. 为说明新模型具有较高的稳定性, 选取2个灰色关联模型(灰色斜率关联模型<sup>[12]</sup>和灰色网格关联模型<sup>[22]</sup>)以及2个非灰色模型(皮尔逊相关系数和马氏距离)作为对比模型, 从研究数据中截取2013年和

2016年北京、天津、石家庄、唐山、秦皇岛、邯郸和保定规范化后的PM<sub>2.5</sub>浓度数据作为实验数据. 在与传统灰色关联模型的对比方面, 分别计算不同城市2016年与2013年的关联系数, 并获取关联度, 结果如表1所示.

从表1中关联系数来看, 3个模型关联系数绝对值均较大, 表明3个模型均认为两序列关联程度较高. 但通过对比关联度和关联系数可以发现, 网格关联模型和斜率关联模型的关联度均表现出与关联系数呈现的关联程度不匹配的情况, 具体表现为: 两模型关联系数大小均在0.750以上, 但是关联度分别为0.322和0.305. 而本文模型关联系数和关联度均在0.900以上, 对关联程度的展示效果一致. 因此, 仅通过一次对比实验来看, 本文模型的稳定性明显优于传统灰色关联模型.

表1 关联系数及关联度计算结果

模型	关联系数						关联度	
网格关联模型	-0.967	0.952	0.962	0.974	0.978	-0.965	-	0.322
斜率关联模型	-0.904	0.957	0.783	0.989	0.958	-0.953	-	0.305
本文模型	0.971	0.954	0.924	0.946	0.930	0.944	0.962	0.947

注: “-”表示关联系数不存在(因网格关联模型和斜率关联模型均以斜率差为基础构建模型, 故7组数据只存在6个关联系数).

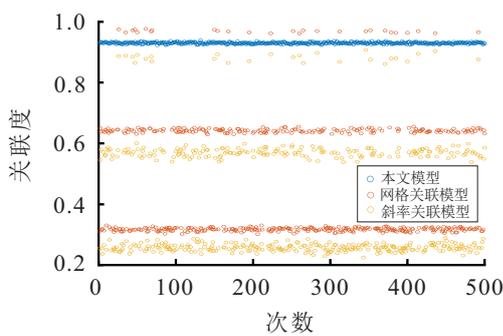
为进一步验证模型稳定性, 实验数据范围设定为原始实验数据上下浮动5%的区间数据, 实验数据序列由区间数据内随机取值数据构成. 在500次重复实

验中, 5个模型分别根据随机选取的序列进行计算, 结果如图5所示.

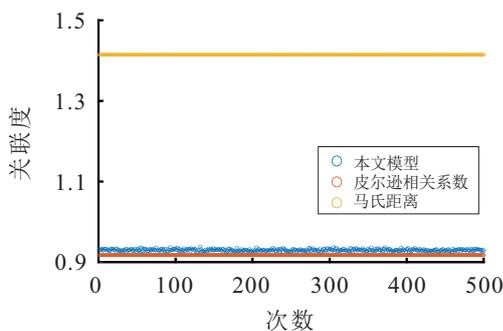
从图5(a)可以看出, 在500次的重复实验中, 当数据仅在原序列正负5%的小区间内变动时, 斜率关联度存在3个区间: 0.20~0.30, 0.50~0.60和0.85~0.90, 网格关联度也存在3个区间: 0.30~0.35, 0.60~0.70和0.90~1.00. 这说明斜率关联模型和网格关联模型受数据轻微变动的影响较大, 模型稳定性较低; 本文模型结果集中在0.90~0.95范围内, 受数据波动影响较小, 稳定性更强, 与表1中单次实验结果相符. 传统灰色关联模型出现稳定性低的主要原因是: 忽视了关联系数之间的正负抵消会导致关联度稳定性降低的问题. 解决此问题的方法是: 厘清灰色关联度符号和大小的现实意义, 关联度符号代表关联方向, 关联度大小表示关联程度. 因此, 在计算关联度时, 应分别计算关联度的符号和大小.

从图5(b)可以看出, 在500次的重复实验中, 3个模型均展示出较强的稳定性. 但是当数据集发生轻微变动时, 皮尔逊相关系数和马氏距离结果几乎未发生改变, 表明两模型的灵敏性较低.

通过上述模型对比可以得出: 本文模型比传统灰色关联模型更稳定, 其灵敏性也强于皮尔逊相关系



(a) 灰色关联模型



(b) 非灰色模型

图5 随机取值序列间的所选模型计算结果的散点图

数和马氏距离,更适用于数据量较少的相关性分析.

### 3 结论

本文剖析了现有灰色关联模型在保序性、稳定性和正负性等方面存在的问题,构建了适用于面板数据的灰色差值关联模型. 首先,将差值特征的现实意义纳入新模型构建中,使得新模型既可以测度关联程度大小,又可以判断关联方向,具有极强的现实意义;其次,新模型改进了传统灰色关联度计算方式,还考虑了数据相对波动对关联度的影响,使得新模型不仅消除了对象排列顺序变动对关联度结果的影响,而且解决了正负关联系数相互抵消导致关联度稳定性较低的问题,还增强了灰色关联模型对序列间相对波动的刻画能力. 在模型性质方面,提出了反对称性和排列保序性在灰色关联分析中的定义,并对新模型具备的性质进行了证明. 在实际应用方面,新模型检测出2013~2021年京津冀地区大气污染情况持续向好,但仍需加强对O<sub>3</sub>污染的治理. 在与所选模型的重复实验对比中,体现了模型结果的合理性、稳定性和排列保序性,展示了本文模型更加完善的体系结构. 此外,新模型的构建对现有灰色关联模型的改进具有极强的参考价值,为能体现正负相关性的灰色关联模型构建提供了清晰的研究思路.

虽然新模型的提出解决了灰色关联理论的诸多问题,但模型仍然存在以下问题值得深入研究:1)模型仅适用于数据集为实数的情况,将其拓展至灰数和模糊数应用领域是值得深入研究的方向;2)模型在理论上可以运用到时空数据等更高维度数据,但仍需在实际案例中进行验证;3)与数据漂移检测方法结合,判断大气污染相关数据分布变化特点,为大气污染物浓度预测提供数据支持是值得拓展的方向.

### 参考文献(References)

- [1] Liu C, Chen R J, Sera F, et al. Ambient particulate air pollution and daily mortality in 652 cities[J]. *The New England Journal of Medicine*, 2019, 381(8): 705-715.
- [2] Mehmood K, Bao Y S, Abrar M M, et al. Spatiotemporal variability of COVID-19 pandemic in relation to air pollution, climate and socioeconomic factors in Pakistan[J]. *Chemosphere*, 2021, 271: 129584.
- [3] 张可, 汪东芳. 经济集聚与环境污染的交互影响及空间溢出[J]. *中国工业经济*, 2014(6): 70-82.  
(Zhang K, Wang D F. The interaction and spatial spillover between agglomeration and pollution[J]. *China Industrial Economics*, 2014(6): 70-82.)
- [4] Song Y, Li Z R, Yang T T, et al. Does the expansion of the joint prevention and control area improve the air quality? — Evidence from China's Jing-Jin-Ji region and surrounding areas[J]. *The Science of the Total Environment*, 2020, 706: 136034.
- [5] 贺祥, 林振山, 刘会玉, 等. 基于灰色关联模型对江苏省PM<sub>2.5</sub>浓度影响因素的分析[J]. *地理学报*, 2016, 71(7): 1119-1129.  
(He X, Lin Z S, Liu H Y, et al. Analysis of the driving factors of PM<sub>2.5</sub> in Jiangsu Province based on grey correlation model[J]. *Acta Geographica Sinica*, 2016, 71(7): 1119-1129.)
- [6] Huang Y S, Shen L, Liu H. Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China[J]. *Journal of Cleaner Production*, 2019, 209: 415-423.
- [7] 蒋珂, 蒋朝辉, 谢永芳, 等. 基于时序关联矩阵的高炉冶炼过程多重关联时延估计方法[J]. *自动化学报*, 2023, 49(2): 329-342.  
(Jiang K, Jiang Z, Xie Y F, et al. A multi-correlated time-delay estimation method in the blast furnace ironmaking process based on time-series correlation matrix[J]. *Acta Automatica Sinica*, 2023, 49(2): 329-342.)
- [8] 黄仁全. 黄河流域高质量发展水平时空演变与灰色关联分析——基于2000~2018年的实证[J]. *生态经济*, 2022, 38(9): 62-70.  
(Huang R Q. The spatiotemporal changes and grey relational analysis of high-quality development level in the Yellow River basin: An empirical research based on the data from 2000-2018 [J]. *Ecological Economy*, 2022, 38(9): 62-70.)
- [9] 韦保磊, 谢乃明. 广义灰色关联分析模型的统一表述及性质[J]. *系统工程理论与实践*, 2019, 39(1): 226-235.  
(Wei B L, Xie N M. Unified representation and properties of generalized grey relational analysis models[J]. *Systems Engineering—Theory & Practice*, 2019, 39(1): 226-235.)
- [10] 魏勇, 曾柯方. 关联度公理的简化与特殊关联度的公理化定义[J]. *系统工程理论与实践*, 2015, 35(6): 1528-1534.  
(Wei Y, Zeng K F. The simplified relational axioms and the axiomatic definition of special incidence degrees[J]. *Systems Engineering—Theory & Practice*, 2015, 35(6): 1528-1534.)
- [11] 彭张林, 张强, 杨善林. 综合评价理论与方法研究综述[J]. *中国管理科学*, 2015, 23(S1): 245-256.  
(Peng Z L, Zhang Q, Yang S L. Review on the theory and method of comprehensive evaluation[J]. *Chinese Journal of Management Science*, 2015, 23(S1): 245-256.)
- [12] 党耀国, 刘思峰, 刘斌, 等. 灰色斜率关联度的改进[J]. *中国工程科学*, 2004, 6(3): 41-44.

- (Dang Y G, Liu S F, Liu B, et al. Improvement on degree of grey slope incidence[J]. Engineering Science, 2004, 6(3): 41-44.)
- [13] 施红星, 刘思峰, 方志耕, 等. 灰色振幅关联度模型[J]. 系统工程理论与实践, 2010, 30(10): 1828-1833.  
(Shi H X, Liu S F, Fang Z G, et al. Grey amplitude incidence model[J]. Systems Engineering—Theory & Practice, 2010, 30(10): 1828-1833.)
- [14] 吴利丰, 王义闹, 刘思峰. 灰色凸关联及其性质[J]. 系统工程理论与实践, 2012, 32(7): 1501-1505.  
(Wu L F, Wang Y N, Liu S F. Grey convex relation and its properties[J]. Systems Engineering—Theory & Practice, 2012, 32(7): 1501-1505.)
- [15] 翟艳丽, 罗格格. 基于面板数据的灰色点集关联分析模型[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2022.2044.  
(Zhai Y L, Luo G G. Grey point set association analysis model based on panel data [J]. Control and Decision, DOI: 10.13195/j.kzyjc.2022.2044.)
- [16] 刘思峰, 谢乃明, FORREST Jeffery. 基于相似性和接近性视角的新型灰色关联分析模型[J]. 系统工程理论与实践, 2010, 30(5): 881-887.  
(Liu S F, Xie N M, Forrest J. On new models of grey incidence analysis based on visual angle of similarity and nearness[J]. Systems Engineering—Theory & Practice, 2010, 30(5): 881-887.)
- [17] 党耀国, 王一博, 王俊杰, 等. 基于面板数据的灰色向量模关联模型及其应用[J]. 控制与决策, DOI: 10.13195/j.kzyjc.2023.1020.  
(Dang Y G, Wang Y B, Wang J J, et al. Grey vector modulus incidence model based on panel data and its application[J]. Control and Decision, DOI: 10.13195/j.kzyjc.2023.1020.)
- [18] 张可, 刘思峰. 灰色关联聚类在面板数据中的扩展及应用[J]. 系统工程理论与实践, 2010, 30(7): 1253-1259.  
(Zhang K, Liu S F. Extended clusters of grey incidences for panel data and its application[J]. Systems Engineering—Theory & Practice, 2010, 30(7): 1253-1259.)
- [19] 吴鸿华, 屈忠锋. 基于面板数据的灰色曲率关联模型[J]. 控制与决策, 2020, 35(5): 1072-1076.  
(Wu H H, Qu Z F. The grey curvature incidence model based on panel data[J]. Control and Decision, 2020, 35(5): 1072-1076.)
- [20] 刘震, 党耀国, 钱吴永, 等. 基于面板数据的灰色网格关联度模型[J]. 系统工程理论与实践, 2014, 34(4): 991-996.  
(Liu Z, Dang Y G, Qian W Y, et al. Grey grid incidence model based on panel data[J]. Systems Engineering—Theory & Practice, 2014, 34(4): 991-996.)
- [21] 李雪梅, 党耀国, 金镛. 基于灰色变化率关联度的灰色趋势分析模型的构建及应用[J]. 中国管理科学, 2015, 23(9): 132-138.  
(Li X M, Dang Y G, Jin L. Grey trend analysis based on grey RelationalDegree with rate of change and its application[J]. Chinese Journal of Management Science, 2015, 23(9): 132-138.)
- [22] Wang J J, Hipel K W, Dang Y G. An improved grey dynamic trend incidence model with application to factors causing smog weather[J]. Expert Systems with Applications, 2017, 87: 240-251.
- [23] 党耀国, 尚中举, 王俊杰, 等. 基于面板数据的灰色指标关联模型构建及其应用[J]. 控制与决策, 2019, 34(5): 1077-1084.  
(Dang Y G, Shang Z J, Wang J J, et al. Grey incidence model for relationship between indicators of panel data and its application[J]. Control and Decision, 2019, 34(5): 1077-1084.)
- [24] Zhang L, Wu P C, Niu M C, et al. A systematic assessment of city-level climate change mitigation and air quality improvement in China[J]. The Science of the Total Environment, 2022, 839: 156274.
- [25] Cai S Y, Wang Y J, Zhao B, et al. The impact of the “Air pollution prevention and control action plan” on PM<sub>2.5</sub> concentrations in Jing-Jin-Ji region during 2012-2020[J]. Science of the Total Environment, 2017, 580: 197-209.

## 作者简介

孙婧(1995—), 女, 博士生, 从事灰色系统理论的研究, E-mail: sunjing\_candice@nuaa.edu.cn;

党耀国(1964—), 男, 教授, 博士生导师, 从事灰色系统理论、数量经济等研究, E-mail: iamdangyg@163.com;

杨圣祥(1972—), 男, 教授, 博士生导师, 从事多目标优化、演化计算等研究, E-mail: syang@dmu.ac.uk;

王俊杰(1988—), 男, 副教授, 博士生导师, 从事灰色系统理论、数量经济等研究, E-mail: wangjj@nuaa.edu.cn;

杨绍闻(1994—), 男, 博士生, 从事灰色系统理论的研究, E-mail: ysw00886@nuaa.edu.cn.