老哥带你学数模

灰色关联系分析与预测模型

微信公众号：科研交流

他日若遂凌云志，敢笑黄巢不丈夫

灰色系统分析方法在建模中的应用

- **CUMCM2003A SARS的传播问题**
- **CUMCM2005A 长江水质的评价和预测**
- **CUMCM2006A 出版社的资源配置**
- **CUMCM2006B 艾滋病疗法的评价及疗效的预测问题**
- **CUMCM2007A 中国人口增长预测**

CUMCM2003A SARS的传播

- SARS（Severe Acute Respiratory Syndrome，严重急性呼吸道综合症, 俗称：非典型肺炎）是21世纪第一个在世界范围内传播的传染病。SARS的爆发和蔓延给我国的经济发展和人民生活带来了很大影响，我们从中得到许多重要的经验和教训，认识到定量地研究传染病的传播规律、为预测和控制传染病蔓延创造条件的重要性。请你们对SARS 的传播建立数学模型，具体要求如下：
- （1）对附件1所提供的一个早期的模型，评价其合理性和实用性。

CUMCM2003A SARS的传播

- （2）建立你们自己的模型，说明为什么优于附件1中的模型；特别要说明怎样才能建立一个真正能够预测以及能为预防和控制提供可靠、足够的信息的模型，这样做的困难在哪里？对于卫生部门所采取的措施做出评论，如：提前或延后5天采取严格的隔离措施，对疫情传播所造成的影响做出估计。附件2提供的数据供参考。
- （3）收集SARS对经济某个方面影响的数据，建立相应的数学模型并进行预测。附件3提供的数据供参考

CUMCM2005A 长江水质的评价和预测

1、问题

水是人类赖以生存的资源，保护水资源就是保护我们自己，对于我国大江大河水资源的保护和治理应是重中之重。专家们呼吁：“以人为本，建设文明和谐社会，改善人与自然的环境，减少污染。”

长江是我国第一、世界第三大河流，长江水质的污染程度日趋严重，已引起了相关政府部门和专家们的高度重视。2004年10月，由全国政协与中国发展研究院联合组成“保护长江万里行”考察团，从长江上游宜宾到下游上海，对沿线21个重点城市做了实地考察，揭示了一幅长江污染的真实画面，其污染程度让人触目惊心。为此，专家们提出“若不及时拯救，长江生态10年内将濒临崩溃”（附件１），并发出了“拿什么拯救癌变长江”的呼唤（附件2）。

CUMCM2005A 长江水质的评价和预测

附件3给出了长江沿线17个观测站（地区）近两年多主要水质指标的检测数据，以及干流上 7 个观测站近一年多的基本数据（站点距离、水流量和水流速）。通常认为一个观测站（地区）的水质污染主要来自于本地区的排污和上游的污水。

一般说来，江河自身对污染物都有一定的自然净化能力，即污染物在水环境中通过物理降解、化学降解和生物降解等使水中污染物的浓度降低。反映江河自然净化能力的指标称为降解系数。

事实上，长江干流的自然净化能力可以认为是近似均匀的，根据检测可知，主要污染物高锰酸盐指数和氨氮的降解系数通常介于0.1~0.5之间，比如可以考虑取0.2 (单位：1/天)。附件4是“1995~2004年长江流域水质报告”给出的主要统计数据。下面的附表是国标(GB3838-2002) 给出的《地表水环境质量标准》中4个主要项目标准限值，其中I、II、III类为可饮用水。

CUMCM2005A 长江水质的评价和预测

- 请你们研究下列问题：

（1）对长江近两年多的水质情况做出定量的综合评价，并分析各地区水质的污染状况。

（2）研究、分析长江干流近一年多主要污染物高锰酸盐指数和氨氮的污染源主要在哪些地区？

（3）假如不采取更有效的治理措施，依照过去10年的主要统计数据，对长江未来水质污染的发展趋势做出预测分析，比如研究未来10年的情况。

（4）根据你的预测分析，如果未来10年内每年都要求长江干流的IV类和V类水的比例控制在20%以内，且没有劣V类水,那么每年需要处理多少污水？

（5）你对解决长江水质污染问题有什么切实可行的建议和意见。

主要内容

- 一灰色预测的概念；
- 二　灰色生成数列；
- 三　灰色关联度分析；
- 四　灰色模型GM；
- 五　灰色预测实例

- 1982我国学者**邓聚龙教授**发表第一篇中文论文《灰色控制系统》标志着灰色系统这一学科诞生。
- 1985灰色系统研究会成立，灰色系统相关研究迅速发展。

- 1989海洋出版社出版英文版《灰色系统论文集》，同年，英文版国际刊物《灰色系统》杂志正式创刊。目前，国际、国内200多种期刊发表灰色系统论文，许多国际会议把灰色系统列为讨论专题。国际著名检索网站已检索我国学者的灰色系统论著500多次。灰色系统理论应用范围已拓展到工业、农业、社会、经济、能源、地质、石油等众多科学领域，成功地解决了生产、生活和科学研究中的大量实际问题，取得了显著成果。

灰色系统的应用范畴大致分为以下几方面：

- （1）灰色关联分析。
- （2）灰色预测：人口预测；灾变预测….等等。
- （3）灰色决策。
- （4）灰色预测控制。

灰色系统理论是人们认识客观系统改造客观系统的一个新型的理论工具。

一、灰色预测的概念

（1）灰色系统、白色系统和黑色系统

- 白色系统是指一个系统的内部特征是完全已知的，即系统的信息是完全充分的。

黑色系统是指一个系统的内部信息对外界来说是一无所知的，只能通过它与外界的联系来加以观测研究。

灰色系统内的一部分信息是已知的，另一部分信息是未知的，系统内各因素间有不确定的关系。

（2）灰色预测法

- 灰色预测法是一种对含有不确定因素的系统进行预测的方法。
- 灰色预测是对既含有已知信息又含有不确定信息的系统进行预则，就是对在一定范围内变化的、与时间有关的灰色过程进行预测。

灰色预测通过鉴别系统因素之间发展趋势的相异程度，即进行关联分析，并可对原始数据进行生成处理来寻找系统变动的规律，生成有较强规律性的数据序列,然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

灰色预测法用等时距观测到的反映预测对象特征的一系列数量值构造灰色预测模型，预测未来某一时刻的特征量，或达到某一 特征量的时间。

（3）灰色预测的四种常见类型

- 灰色时间序列预测

即用观察到的反映预测对象特征的时间序列来构造灰色预测模型，预测未来某一时刻的特征量，或达到某一特征量的时间。

- 畸变预测

即通过灰色模型预测异常值出现的时刻，预测异常值什么时候出现在特定时区内。

- 系统预测

通过对系统行为特征指标建立一组相互关联的灰色预测模型，预测系统中众多变量间的相互协调关系的变化。

- 拓扑预测

将原始数据做曲线，在曲线上按定值寻找该定值发生的所有时点，并以该定值为框架构成时点数列，然后建立模型预测该定值所发生的时点。

二　灰色关联度与优势分析

大千世界里的客观事物往往现象复杂，因素繁多。我们经常要对系统进行因素分析，这些因素中哪些对系统来讲是主要的，哪些是次要的，哪些需要发展，哪些需要抑制，哪些是潜在的，哪些是明显的。一般来讲，这些都是我们极为关心的问题。事实上，因素间关联性如何、关联程度如何量化等问题是系统分析的关键。

例如人们关心的人口问题构成一个系统，影响人口发展变化的因素有社会方面的诸如计划生育、社会治安、社会生活方式等；有经济方面的诸如国民收入、社会福利、社会保险等；还有医疗方面的诸如医疗条件、医疗水平等……也就是说，人口是多种因素互相关联、互相制约的系统，对这些因素进行分析将有助于人们对人口的未来预测及人口控制工作。

因素分析的基本方法过去主要是采用回归分析等办法，但回归分析的办法有很多欠缺，如要求大量数据、计算量大以及可能出现反常情况等。为克服以上弊病，本节采用灰色关联度分析的办法来做系统分析。

灰色关联度一定是分析向量与向量之间以及矩阵与矩阵之间的关联度。既然计算关联度，一定是计算某一个待比较的数列与参照物（参考数列）之间的相关程度。

2.1 灰色关联度

选取参考数列

$$X_0=\{X_0(k)\mid k=1,2,\cdots,n\}=(X_0(1),X_0(2),\cdots,X_0(n))$$

其中 k 表示时刻。

假设有 m 个比较数列

$$X_i=\{X_i(k)\mid k=1,2,\cdots,n\}=(X_i(1),X_i(2),\cdots,X_i(n))$$

$$(i=1,2,\cdots,m)$$

则称

$$\zeta_i(k)=\frac{\min_i\min_k\left|X_0(k)-X_i(k)\right|+\rho\max_i\max_k\left|X_0(k)-X_i(k)\right|}{\left|X_0(k)-X_i(k)\right|+\rho\max_i\max_k\left|X_0(k)-X_i(k)\right|}\qquad(1)$$

为比较数列 X_i 对参考数列 X_0 在 k 时刻的关联系数，其中 $\rho\in[0,+\infty)$ 为分辨系数。一般来讲，分辨系数 $\rho\in[0,1]$，由（1）容易看出，ρ 越大，分辨率越大；ρ 越小，分辨率越小。

式（1）定义的关联系数是描述比较数列与参考数列在某时刻关联程度的一种指标，由于各个时刻都有一个关联系数，因此信息显得过于分散，不便于比较，为此我们给出

$$r_i=\frac{1}{n}\sum_{k=1}^{n}\zeta_i(k) \qquad (2)$$

为比较数列 X_i 对参考数列 X_0 的关联度。

应该指出的是，公式（1）中的$|X_0(k)-X_i(k)|$不能区别因素关联是正关联还是负关联，可采取下面办法解决这个问题。记

$$\sigma_i=\sum_{k=1}^{n}kX_i(k)-\sum_{k=1}^{n}X_i(k)\sum_{k=1}^{n}\frac{k}{n}$$

$$\sigma_n=\sum_{k=1}^{n}k^2-(\sum_{k=1}^{n}k)2/n$$

则当$sign\left(\frac{\sigma_i}{\sigma_n}\right)=sign\left(\frac{\sigma_j}{\sigma_n}\right)$，则$X_i$和$X_j$为正关联；

当$sign\left(\frac{\sigma_i}{\sigma_n}\right)=-sign\left(\frac{\sigma_j}{\sigma_n}\right)$，则$X_i$和$X_j$为负关联。

2.2 案例分析

利用灰色关联分析对6位教师工作状况进行综合分析

1 分析指标包括：专业素质、外语水平、教学工作量、科研成果、论文、著作与出勤.

2 对原始数据经处理后得到以下数值，见下表

编号	专业	外语	教学量	科研	论文	著作	出勤
1	8	9	8	7	5	2	9
2	7	8	7	5	7	3	8
3	9	7	9	6	6	4	7
4	6	8	8	8	4	3	6
5	8	6	6	9	8	3	8
6	8	9	5	7	6	4	8

2.2 案例分析

利用灰色关联分析对6位教师工作状况进行综合分析

3 确定参考数据列

$$\{x_0\} = \{9, 9, 9, 9, 8, 9, 9\}$$

4 计算$|x_0(k) - x_j(k)|$，见下表

编号	专业	外语	教学量	科研	论文	著作	出勤
1	1	0	1	2	3	7	0
2	2	1	2	4	1	6	1
3	0	2	0	3	2	5	2
4	3	1	1	1	4	6	3
5	1	3	3	0	0	6	1
6	1	0	4	2	2	5	1

2.2 案例分析

利用灰色关联分析对6位教师工作状况进行综合分析

5 求最值

$$\min_{i=1}^{n}\min_{k=1}^{m}\left|x_0(k)-x_i(k)\right| = \min(0,1,0,1,0,0) = 0$$

$$\max_{i=1}^{n}\max_{k=1}^{m}\left|x_0(k)-x_i(k)\right| = \max(7,6,5,6,6,5) = 7$$

4 依据式（2），取 ρ 为0.5得到

$$\zeta_1(1) = \frac{0+0.5\times 7}{1+0.5\times 7} = 0.778,\ \zeta_1(2) = \frac{0+0.5\times 7}{0+0.5\times 7} = 1.000$$

$$\zeta_1(3) = 0.778,\ \zeta_1(4) = 0.636,\ \zeta_1(5) = 0.467,\ \zeta_1(6) = 0.333$$

$$\zeta_1(7) = 1.000$$

2.2 案例分析

利用灰色关联分析对6位教师工作状况进行综合分析

5 同理得出其它各值，见下表

编号	专业	外语	教学量	科研	论文	著作	出勤
1	0.778	1.000	0.778	0.636	0.467	0.333	1.000
2	0.636	0.778	0.636	0.467	0.636	0.368	0.778
3	1.000	0.636	1.000	0.538	0.538	0.412	0.636
4	0.538	0.778	0.778	0.778	0.412	0.368	0.538
5	0.778	0.538	0.538	1.000	0.778	0.368	0.778
6	0.778	1.000	0.467	0.636	0.538	0.412	0.778

6 分别计算每个人各指标关联系数的均值（关联序）：

$$r_{01}=\frac{0.778+1.000+0.778+0.636+0.467+0.333+1.000}{7}=0.703$$

$$r_2=0.614,\ r_{03}=0.680,\ r_{04}=0.599,\ r_{05}=0.683,\ r_{01}=0.658$$

三、灰色生成数列

灰色系统理论认为，尽管客观表象复杂，但总是有整体功能的，因此必然蕴含某种内在规律。关键在于如何选择适当的方式去挖掘和利用它。灰色系统是通过对原始数据的整理来寻求其变化规律的，这是一种就数据寻求数据的现实规律的途径，即为灰色序列的生成。一切灰色序列都能通过某种生成弱化其随机性，显现其规律性。数据生成的常用方式有累加生成、累减生成和加权累加生成。

（1）累加生成

把数列各项（时刻）数据依次累加的过程称为累加生成过程

由累加生成过程所得的数列称为

累加生成数列。设原始数列为 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n,$$

令

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

称所得到的新数列为数列 $x^{(0)}$ 的1次累加生成数列。类似

地有

$$x^{(r)}(k)=\sum_{i=1}^{k}x^{(r-1)}(i),k=1,2,\cdots,n,r\geq 1$$

称为 $x^{(0)}$ 的r次累加生成数列。

（2）累减生成

对于原始数据列依次做前后相邻的两个数据相减的运算过程称为累减生成过程**IAGO**。如果原始数据列为 $x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n))$

令 $x^{(0)}(k)=x^{(1)}(k)-x^{(1)}(k-1),k=2,3,\cdots,n,$

称所得到的数列 $x^{(0)}$ 为 $x^{(1)}$ 的1次累减生成数列。

注：从这里的记号也可以看到，从原始数列 $x^{(0)}$，得到新数列 $x^{(1)}$，再通过累减生成可以还原出原始数列。实际运用中在数列 $x^{(1)}$ 的基础上预测出 $\hat{x}^{(1)}$，通过累减生成得到预测数列 $\hat{x}^{(0)}$。

（3）加权邻值生成

设原始数列为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$

称 $x^{(0)}(k-1), x^{(0)}(k)$ 为数列 $x^{(0)}$ 的任意一对邻值。

$x^{(0)}(k-1)$ 为后邻值，$x^{(0)}(k)$ 为前邻值，对于常数$\alpha \in [0,1]$ ，令 $z^{(0)}(k) = \alpha x^{(0)}(k) + (1-\alpha)x^{(0)}(k-1), k = 2,3,\cdots,n,$

由此得到的数列$z^{(0)}$称为数列$x^{(0)}$在权α下的邻值生成数，权α也称为生成系数。

特别地，当生成系数$\alpha = 0.5$ 时，则称

$$z^{(0)}(k) = 0.5x^{(0)}(k) + 0.5x^{(0)}(k-1), k = 2,3,\cdots,n,$$

为均值生成数，也称等权邻值生成数。

累加生成计算示例

例：$x^{(0)}=(x^{(0)}(k) \mid k=1，2，3，4，5)$

$=x^{(0)}(1)，x^{(0)}(2)，x^{(0)}(3)，x^{(0)}(4)，x^{(0)}(5)$

$=(3.2，3.3，3.4，3.6，3.8)$

求 $x^{(1)}(k)$

解：

$$k=1, x^{(1)}(1)=x^{(0)}(1)=3.2$$

$$k=2, x^{(1)}(2)=\sum_{i=1}^{2} x^{(0)}(i)=x^{(0)}(1)+x^{(0)}(2)=3.2+3.3=6.5$$

$$k=3,x^{(1)}(3)=\sum_{i=1}^{3}x^{(0)}(i)=x^{(1)}(2)+x^{(0)}(3)=6.5+3.4=9.9$$

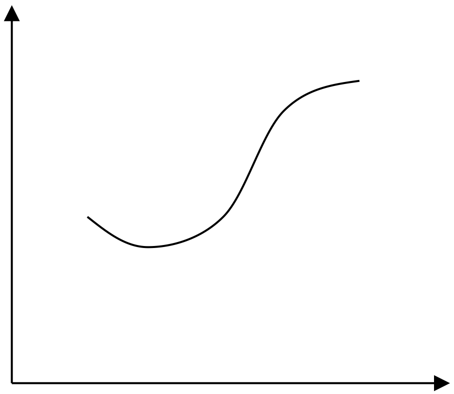
$$k=4,x^{(1)}(4)=\sum_{i=1}^{4}x^{(0)}(i)=x^{(1)}(3)+x^{(0)}(4)=9.9+3.6=13.5$$

$$k=5,x^{(1)}(5)=\sum_{i=1}^{5}x^{(0)}(i)=x^{(1)}(4)+x^{(0)}(5)=13.5+3.8=17.3$$

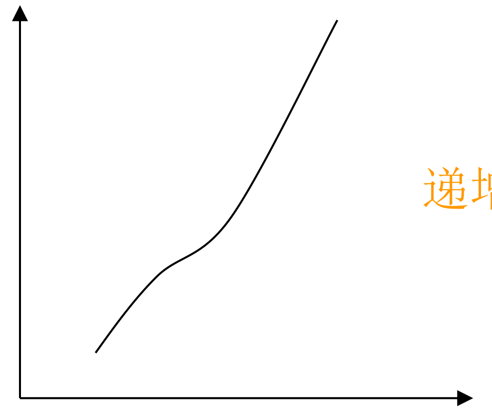
累加生成的特点

一般经济数列都是非负数列。累加生成能使任意非负数列、摆动的与非摆动的，转化为非减的、递增的。

原始数列作图



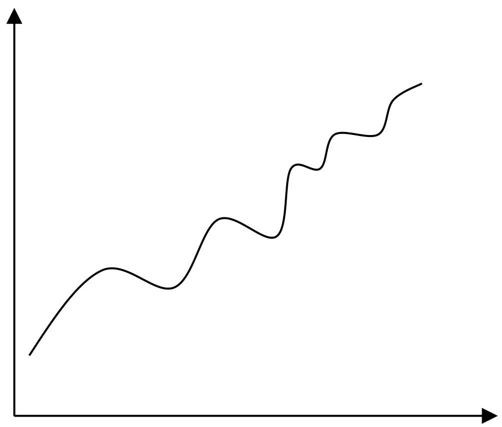
1—AGO作图



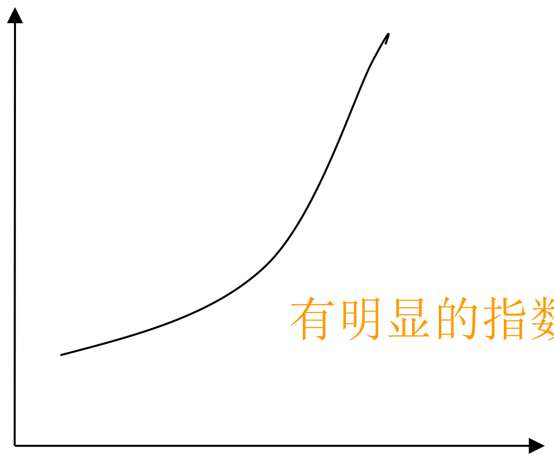
递增的规律

某市的汽车销售量

原始数列作图

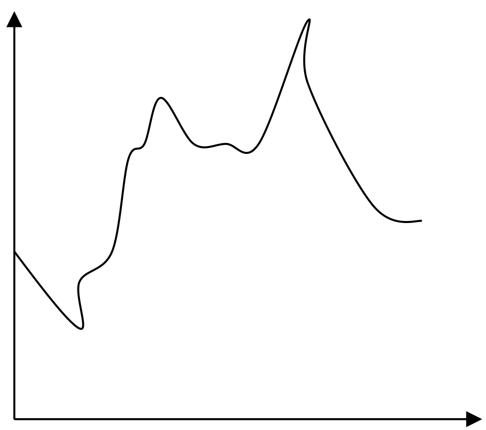


1—AGO作图

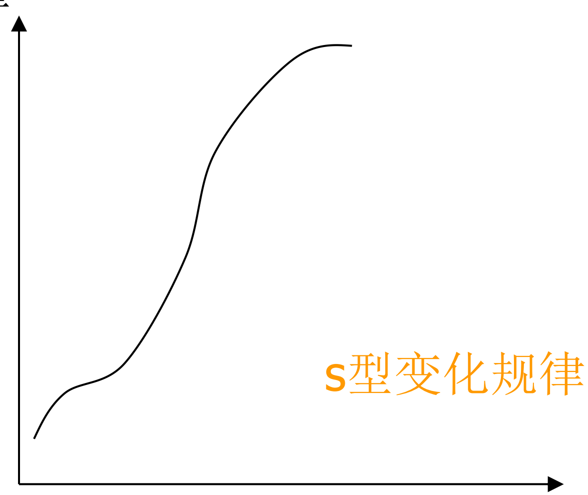


有明显的指数关系的规律

某地区作物产量



某钢厂产量



累减生成计算示例

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), x^{(1)}(4), x^{(1)}(5), x^{(1)}(6)) = (5,9,14,24,35,46)$$

解： $x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$

若 $k = 0, x^{(1)}(0) = 0$

$k = 1, x^{(0)}(1) = x^{(1)}(1) - x^{(1)}(0) = x^{(1)}(1) = 5$

$k = 2, x^{(0)}(2) = x^{(1)}(2) - x^{(1)}(1) = 4$

$k = 3, x^{(0)}(3) = x^{(1)}(3) - x^{(1)}(2) = 5$

$k = 4, x^{(0)}(4) = x^{(1)}(4) - x^{(1)}(3) = 10$

$k = 5, x^{(0)}(5) = x^{(1)}(5) - x^{(1)}(4) = 11$

$k = 6, x^{(0)}(6) = x^{(1)}(6) - x^{(1)}(5) = 11$

从而有： IGAO （$x^{(0)}$）$= (5,4,5,10,11,11)$

不难看出，累减生成具　有求导性质，这是因为

$$\frac{dx(k)}{dt} = \lim_{\Delta t \to 0} \frac{x(k) - x(k - \Delta t)}{\Delta t}$$

而 $\alpha^{(1)}(x(k)) = x(k) - x(k-1)$, 相当于 $\Delta t = 1$

四、灰色模型G M(1,1)

灰色系统理论是基于关联空间、光滑离散函数等概念定义灰导数与灰微分方程，进而用离散数据列建立微分方程形式的动态模型，即灰色模型是利用离散随机数经过生成变为随机性被显著削弱而且较有规律的生成数，建立起的微分方程形式的模型，这样便于对其变化过程进行研究和描述。

G表示grey（灰色），M表示model（模型）

- 设 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ 为原始数列，其1次累
- 加生成数列为 $x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n))$ ，其中

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),k=1,2,\cdots,n,$$

- 定义 $x^{(1)}$ 的灰导数为

$$d(k)=x^{(0)}(k)=x^{(1)}(k)-x^{(1)}(k-1).$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的邻值生成数列，即

$$z^{(1)}(k)=\alpha x^{(1)}(k)+(1-\alpha)x^{(1)}(k-1),$$

于是定义GM（1，1）的灰微分方程模型为

$$d(k)+az^{(1)}(k)=b,$$

即或 $x^{(0)}(k)+az^{(1)}(k)=b$，（1）

在式（1）中，$x^{(0)}(k)$称为灰导数，a称为发展系数，$z^{(1)}(k)$称为白化背景值，b称为灰作用量。

将时刻表 $k=2,3,\cdots,n$ 代入（1）式有

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

引入矩阵向量记号：

$$u=\begin{bmatrix} a \\ b \end{bmatrix} \quad Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \quad \mathbf{B}=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

于是GM（1，1）模型可表示为 $Y=\mathbf{B}u.$

现在问题归结为求a,b在值。用一元线性回归，即最小二乘法求它们的估计值为

$$\hat{u}=\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix}=(B^TB)^{-1}B^TY.$$

注：实际上回归分析中求估计值是用软件计算的，有标准程序求解，如matlab等。

GM（1，1）的白化型

对于GM（1，1）的灰微分方程（1），如果将灰导数 $x^{(0)}(k)$ 的时刻 $k=2,3,\cdots,n$ 视为连续变量t，则 $x^{(1)}$ 视为时间t函数 $x^{(1)}(t)$，于是 $x^{(0)}(k)$ 对应于导数量级 $\frac{dx^{(1)}(t)}{dt}$，白化背景值 $z^{(1)}(k)$ 对应于导数 $x^{(1)}(t)$。于是GM（1，1）的灰微分方程对应于的白微分方程为

$$\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(t)=b, \qquad （2）$$

$$\int_{k-1}^{k}(dx^1/dt)dt+\int_{k-1}^{k}ax^1dt=\int_{k-1}^{k}bdt$$

上式左边第二个式子 $\int_{k-1}^{k}ax^1dt\approx a/2(x^1(k)+x^1(k-1))=az^1$

白化方程化为 $x^0+az^1=b$

GM（1，1）灰色预测的步骤

1.数据的检验与处理

为了保证GM（1，1）建模方法的可行性，需要对已知数据做必要的检验处理。

设原始数据列为了 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ ，计算数列的级比

$$\lambda(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},k=2,3,\cdots,n.$$

如果所有的级比都落在可容覆盖区间 $X=(e^{\frac{-2}{n+1}},e^{\frac{2}{n+1}})$ 内，则数据列 $x^{(0)}$ 可以建立GM（1，1）模型且可以进行灰色预测。

否则对数据做适当的变换处理，如平移变换：

取C使得数据列

$$y^{(0)}(k)=x^{(0)}(k)+c,k=1,2,\cdots,n,$$

的级比都落在可容覆盖内。

2. 建立GM（1，1）模型

不妨设 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$ 满足上面的要求，以它为数据列建立GM（1，1）模型

$$x^{(0)}(k)+az^{(1)}(k)=b,$$

用回归分析求得a,b的估计值，于是相应的白化模型为

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

解为

$$x^{(1)}(t)=(x^{(0)}(1)-\frac{b}{a})e^{-a(t-1)}+\frac{b}{a}. \qquad (3)$$

于是得到预测值

$$\hat{x}^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a},k=1,2,\cdots,n-1,$$

从而相应地得到预测值：

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k),k=1,2,\cdots,n-1,$$

3. 检验预测值

（1）残差检验：计算相对残差

$$\varepsilon(k)=\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)},k=1,2,\cdots,n,$$

如果对所有的 $|\varepsilon(k)|<0.1$ ，则认为达到较高的要求；否则，若对所有的 $|\varepsilon(k)|<0.2$ ，则认为达到一般要求。

（2）级比偏差值检验：计算

$$\rho(k)=1-\frac{1-0.5a}{1+0.5a}\lambda(k),$$

如果对所有的 $|\rho(k)|<0.1$ ，则认为达到较高的要求；否则

若对所有的 $|\rho(k)|<0.2$ ，则认为达到一般要求。

五、灰色预测计算实例一

银行有各种投资理财产品，客户可根据自己的资金实力和投资偏好来自由选择，并且一般会有“10天犹豫期”，在这10天里如果对自己购得的理财产品不放心或者不满意通常情况下是可以退买的，这时候是不收手续费的。否则逾期退买将收取一定的手续费。

通过对客户退买行为数据的分析，发现客户购得理财产品后的每一天继续持有的客户比例依次是[92.810 97.660 98.800 99.281 99.537 99.537 99.817 0.00](单位%)，从这组数列可以看出退买高发期是在前几天，后续退买的可能性持续衰减。建立GM(1,1)模型对以上数据进行分析。

灰色预测计算实例二

- 例 北方某城市1986～1992 年道路交通噪声平均声级数据见表6
- 表6 市近年来交通噪声数据[dB(A)]

序号	年份	L_{eq}
1	1986	71.1
2	1987	72.4
3	1988	72.4
4	1989	72.1
5	1990	71.4
6	1991	72.0
7	1992	71.6

第一步: 级比检验

建立交通噪声平均声级数据时间序列如下：

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots x^{(0)}(7))$$

$$=(71.1, 72.4, 72.4, 72.1, 71.4, 72.0, 71.6)$$

- （1）求级比λ (k) $\lambda(k)=\dfrac{x^{(0)}(k-1)}{x^{(0)}(k)}$
- $\lambda=(\lambda(2),\lambda(3),\cdots,\lambda(7))$

=(0.982,1,1.0042,1.0098, 0.9917,1.0056)

（2）级比判断

由于所有的λ (k)∈[0.982,1.0098]，k = 2,3,…,7，故可以用x(0)作满意的GM（1，1）建模。

第二步: GM（1，1）建模

（1）对原始数据$x^{(0)}$作一次累加，即

$x^{(1)}$ = (71.1,143.5,215.9,288,359.4,431.4,503）

（2）构造数据矩阵B 及数据向量Y

$$B=\begin{bmatrix} -\frac{1}{2}(x^{(1)}(1)+x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2)+x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(6)+x^{(1)}(7)) & 1 \end{bmatrix},\ Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(7) \end{bmatrix}$$

（3）计算u^

$$\hat{u}=(\hat{a},\hat{b})^T=(B^T\cdot B)^{-1}B^TY=\begin{pmatrix}0.0023\\72.6573\end{pmatrix}$$

于是得到a = 0.0023，b = 72.6573

（4）建立模型

$$\frac{dx^{(1)}}{dt}+0.0023x^{(1)}=72.6573$$

求解得

$$x^{(1)}(k+1)=(x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a}=-30929e^{-0.0023k}+31000$$

（5）求生成数列值 $\hat{x}^{(1)}(k+1)$ 及模型还原值 $\hat{x}^{(0)}(k+1)$：

$$\hat{x}^{(1)}(1)=\hat{x}^{(0)}(1)=x^{(0)}(1)=71.1,$$

令k＝1,2,3,4,5,6,由上面的时间响应函数可算得 $\hat{x}^{(1)}$

由 $\hat{x}^{(0)}(k)=\hat{x}^{(1)}(k)-\hat{x}^{(1)}(k-1)$，取k＝2,3,4,...,7，得

$$\hat{x}^{(0)}=(\hat{x}^{(0)}(1),\ \hat{x}^{(0)}(2),\cdots,\ \hat{x}^{(0)}(7))=$$

(71.1, 72.4, 72.2, 72.1, 71.9, 71.7, 71.6)

- 第三步: 模型检验
- 模型的各种检验指标值的计算结果见表 7.
- 表7 GM(1,1)模型检验表

序号	年份	原始值	模型值	残差	相对误差	级比偏差
1	1986	71.1	71.1	0	0	
2	1987	72.4	72.4	-0.0057	0.01%	0.0023
3	1988	72.4	72.2	0.1638	0.23%	0.0203
4	1989	72.1	72.1	0.0329	0.05%	-0.0018
5	1990	71.4	71.9	-0.4984	0.7%	-0.0074
6	1991	72.0	71.7	0.2699	0.37%	0.0107
7	1992	71.6	71.6	0.0378	0.05%	-0.0032

- 经验证，该模型的精度较高，可进行预测和预报。

灰色预测实例三

已知某公司1999-2008年的利润为(单位：元/年)：[89677,99215,109655,120333,135823,159878,182321,209407,246619,300670],现在要预测该公司未来几年的利润情况。

步骤:

(1)对原始数据进行累加;

(2)构造累加矩阵B与常数向量;

(3)求解灰参数;

(4)将参数代入预测模型进行数据预测。

灰色预测实例四

长江水质的预测(2005A)
根据以下数据进行10年的预测

年份	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
污水	174	179	183	189	207	234	220.5	256	270	285

案例：SARS疫情对某些经济指标影响

- **1 问题的提出**
- **2003年的SARS疫情对中国部分行业的经济发展产生了一定的影响，特别是对帮分疫情较严重的省市的相关行业所造成的影响是明显的，经济影响主要分为直接经济影响和间接影响．直接经济影响涉及到商品零售业、旅游业、综合服务等行业．很多方面难以进行定量地评估，现仅就SARS疫情较重的某市商品零售业、旅游业和综合服务业的影响进行定量的评估分析．**
- **究竟SARS疫情对商品零售业、旅游业和综合服务业的影响有多大，已知该市从1997年1月到2003年10月的商品零售额、接待旅游人数和综合服务收入的统计数据如下表1、表2、表3．**

表1 商品的零售额(单位：亿元)

- 年代 1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 11月 12月
- 1997 83.0 79.8 78.1 85.1 86.6 88.2 90.3 86.7 93.3 92.5 90.9 96.9
- 1998 101.7 85.1 87.8 91.6 93.4 94.5 97.4 99.5 104.2 102.3 101.0 123.5
- 1999 92.2 114.0 93.3 101.0 103.5 105.2 109.5 109.2 109.6 111.2 121.7 131.3
- 2000 105.0 125.7 106.6 116.0 117.6 118.0 121.7 118.7 120.2 127.8 121.8 121.9
- 2001 139.3 129.5 122.5 124.5 135.7 130.8 138.7 133.7 136.8 138.9 129.6 133.7
- 2002 137.5 135.3 133.0 133.4 142.8 141.6 142.9 147.3 159.6 162.1 153.5 155.9
- 2003 163.2 159.7 158.4 145.2 124 144.1 157 162.6 171.8 180.7 173.5 176.5

表2 接待海外旅游人数(单位：万人)

- 年代 1月 2月 3月 4月 5月 6月 7月 8月 9月 10月 l1月 12月
- 1997 9.4 11.3 16.8 19.8 20.3 18.8 20.9 24.9 24.7 24.3 19.4 18.6
- 1998 9.6 11.7 15.8 19.9 19.5 17.8 17.8 23.3 21.4 24.5 20.1 15.9
- 1999 10.1 12.9 17.7 21.0 21.0 20.4 21.9 25.8 29.3 29.8 23.6 16.5
- 2000 11.4 26.0 19.6 25.9 27.6 24.3 23.0 27.8 27.3 28.5 32.8 18.5
- 2001 11.5 26.4 20.4 26.1 28.9 28.0 25.2 30.8 28.7 28.1 22.2 20.7
- 2002 13.7 29.7 23.1 28.9 29.0 27.4 26.0 32.2 31.4 32.6 29.2 22.9
- 2003 15.4 17.1 23.5 l1.6 1.78 2.61 8.8 16.2 20.1 24.9 26.5 21.8

表3 综合服务业累计数据(单位：亿元)

年代	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
1997	96	144	194	276	383	466	554	652	747	832	972
1998	111	169	235	400	459	565	695	805	881	1011	1139
1999	151	238	335	425	541	641	739	866	975	1087	1238
2000	164	263	376	531	600	711	913	1038	1173	1296	1497
2001	182	318	445	576	708	856	1000	1 145	1292	1435	1667
2002	216	361	504	642	818	979	1142	1305	1479	1644	1920
2003	241	404	584	741	923	1114	1298	1492	1684	1885	2218

- 试根据这些历史数据建立预测评估模型，评估2003年SARS疫情给该市的商品零售业、旅游业和综合服务业所造成的影响.

2 模型的分析与假设

- 根据所掌握的历史统计数据可以看出，在正常情况下，全年的平均值较好地反映了相关指标的变化规律，这样可以把预测评估分成两部分：
- (i)利用灰色理论建立灰微分方程模型，由1997~2002年的平均值预测2003年平均值；
- (ii)通过历史数据计算每个月的指标值与全年总值的关系，从而可预测出正常情况下2003年每个月的指标值，再与实际值比较可以估算出SARS疫情实际造成的影响.

2 模型的分析与假设

- 给出下面两条假设：
- (1)假设该市的统计数据都是可靠准确的；
- (2)假设该市在SARS疫情流行期间和结束之后，数据的变化只与SARS疫情的影响有关，不考虑其他随机因素的影响.

3 建立灰色预测模型GM(1,1)

- 由已知数据，对于1997~2002年某项指标记为矩阵 $A=(a_{ij})_{6\times12}$ 计算每年的年平均值，记为
- $$x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(6)) \qquad (3)$$
- 并要求级比 $\sigma(i)=x^{(0)}(i-1)/x^{(0)}(i)\in(0.7515,1.3307)(i=2,3,\cdots,6)$.
- 对 $x^{(0)}$ 作一次累加，则
- $x^{(1)}(1)=x^{(0)}(1),x^{(1)}(i)=\sum_{k=1}^{i}x^{(0)}(k)(i=2,3,\cdots,6)$, 记
- $$x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(6)) \qquad (4)$$
- 取 $x^{(1)}$ 的加权均值，则 $z^{(1)}(k)=\alpha x^{(1)}(k)+(1-\alpha)x^{(1)}(k-1)(k=2,3,\cdots,6),\alpha$ 为确定参数 ,于是GM(1,1)的白化微分方程模型为

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b, \qquad (5)$$

其中a是发展灰度,b是内生控制灰度.

3 建立灰色预测模型GM(1,1)

- 由于 $x^{(1)}(k)-x^{(1)}(k-1)=x^{(0)}(k)$ 取 $x^{(0)}(k)$ 为灰导数， 为 $z^{(1)}(k)$ 背景值，则建立灰微分方程为:
-

$$x^{(0)}(k)+az^{(1)}(k)=b(k=2,3,\cdots,6)$$

或

$$x^{(0)}(k)=-az^{(1)}(k)+b(k=2,3,\cdots,6)$$

- 其矩阵形式为 $Y^{(0)}=B\cdot(a,b)^T$ ，
- 其中

$$Y^{(0)}=(x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(6))^T,B=\begin{pmatrix}-z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(6)\\ 1 & 1 & \cdots & 1\end{pmatrix}^T$$

用最小二乘法求得参数的估计值为

$$(\hat{a},\hat{b})^T=(B^T\cdot B)^{-1}\cdot B^T\cdot Y^{(0)}. \quad \textbf{(6)}$$

3 建立灰色预测模型GM(1,1)

则灰微分方程模型(4)的解为 $\hat{x}^{(1)}(t+1)=(x^{(0)}(1)-\frac{b}{a})\cdot e^{-at}+\frac{b}{a}$,

则

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=(x^{(0)}(1)-\frac{b}{a})\cdot(e^{-ak}-e^{-a(k-1)}) \quad (7)$$

由(7)式可以得到2003年的平均值为 $\bar{x}$ ，则预测2003年的总值为 $X=12\cdot\bar{x}$. 根据历史数据，可以统计计算出2003年第i个月的指标值占全年总值的比例为 u_i ，即

$$u_i=\sum_{j=1}^{6}a_{ij}\Big/\sum_{i=1}^{12}\sum_{j=1}^{6}a_{ij}\quad(i=1,2,\cdots,12), \quad (8)$$

则 $u=(u_1,u_2,\cdots,u_{12})$, 于是可得2003年每一个月的指标值为Y=X ·u.

4 模型的求解

(i)商品零售额

- 由数据表1，计算可得每年月平均值、一次累加值分别为

$x^{(0)}$ =(87．6167，98．5000，108．4750，118．4167，132．8083，145．4083)，

$x^{(1)}$ =(87．6167，186．1167，294．5917，413．0083，545．8167，691．2250)．

- 显然 $x^{(0)}$ 的所有级比都在可行域内．经检验，在这里取参数 $\alpha=0.4$ 比较合适，则有 $z^{(1)}=$ (127．0167，229．5067，341．9583，466．1317，603．9800)．

(i)商品零售额

- 由最小二乘法求得a=-0．0993，b=85．5985．可得2003年的月平均值为

$\overline{x}$ =162．8826亿元；年总值为 $X = 12 \cdot \overline{x}$.

=1954．6亿元．　由(8)式得每月的比例为

u=(0．0794，0．0807，0．0749，0．0786，0．0819，
0．0818，0．0845，0．0838，0．0872，0．0886，
0．0866，0．0920)

- 故2003年1—12月的预测值为
- Y=u·X =(155．2，157．8，146．4，153．6，160．1，
159．9，165．2，163．8，17o．5，173．2，
169．3，179．9)(亿元)

将预测值与实际统计值进行比较如下表4所示.

表4 2003年商品的零售额(单位：亿元)

月份	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
预测值	155.2	157.8	146.4	153.6	160.1	159.9	165.2	163.8	170.5	173.2	169.3	179.9
实际值	163.2	159.7	158.4	145.2	124.0	144.1	157.0	162.6	171.8	180.7	173.5	176.5

```
clc,clear
han1=[83.0 79.8 78.1 85.1 86.6 88.2 90.3 86.7 93.3 92.5 90.9 96.9
101.7 85.1 87.8 91.6 93.4 94.5 97.4 99.5 104.2 102.3 101.0 123.5
92.2 114.0 93.3 101.0 103.5 105.2 109.5 109.2 109.6 111.2 121.7 131.3
105.0 125.7 106.6 116.0 117.6 118.0 121.7 118.7 120.2 127.8 121.8 121.9
139.3 129.5 122.5 124.5 135.7 130.8 138.7 133.7 136.8 138.9 129.6 133.7
137.5 135.3 133.0 133.4 142.8 141.6 142.9 147.3 159.6 162.1 153.5 155.9
163.2 159.7 158.4 145.2 124.0 144.1 157.0 162.6 171.8 180.7 173.5 176.5];
han1(end,:)=[];%相当于han1=han1(1:6,:);
m=size(han1,2);%把月份提取出来，12个月
x0=mean(han1,2);%返回x矩阵每行的平均值，其中的2代表返回行
x1=cumsum(x0);%一次累加
alpha=0.4;
n=length(x0);%长度，数据的维度，n=6
z1=alpha*x1(2:n)+(1-alpha)*x1(1:n-1)%q求邻域生成数
Y=x0(2:n);B=[-z1,ones(n-1,1)];
ab=B\Y%求出a,b
k=6;
x7hat=(x0(1)-ab(2)/ab(1))*(exp(-ab(1)*k)-exp(-ab(1)*(k-1)))%预测结果
z=m*x7hat
u=sum(han1)/sum(sum(han1))
v=z*u
```

- x1 = [87.6167 186.1167 294.5917 413.0083 545.8167 691.2250]
- z1 = [127.0167 229.5067 341.9583 466.1317 603.9800]
- ab = [-0.0993 85.5985]
- x7hat = 162.8793 z = 1.9546e+003

- u = [0.0794 0.0807 0.0749 0.0786 0.0819 0.0818 0.0845 0.0838 0.0872 0.0886 0.0866 0.0920]

- v =[155.2152 157.7365 146.4023 153.5421
- 160.1400 159.8337 165.0649 163.7924
- 170.5317 173.1473 169.3064 179.8394]

(ii)接待海外旅游人数

由数据表2，计算每年月平均值 $x^{(0)}$

和一次累加值 $x^{(1)}$. 取参数 $\alpha=0.5$,可得加权平均值 $z^{(1)}$.

进而求得a=-0.0938，b=16.2670，$\overline{x}$=30.2649, X=363.1788，以及

u=(0.0407,0.0732,0.0703,0.0878,0.0907,0.0848,0.0836,0.1022,0.1010,0.1041,0.0914,0.0701).

于是可得到2003年的接待海外旅游人数的预测值，并与实际值比较如下表5所示.

- 表5 2003年接待海外旅游人数(单位：万人)
- 月份 1月 2月 3月 4月 5月 6月 7月8月 9月 10月 11月 12月
- 预测值 14.8 26.6 25.5 31.9 33.0 30.8 30.4 37.1 36.7 37.8 33.2 25.5
- 实际值 15.4 17.1 23.5 11.6 1.78 2.61 8.8 16.2 20.1 24.9 26.5 21.8

(iii)综合服务收入类似处理(略)

5 模型的结果分析

- 根据该市的统计报告显示，2003年4、5、6三个月的实际商品零售额分别为145．2，124、144．1亿元．在这之前，根据统计部门的估计4、5、6三个月份SARS疫情对该市的商品零售业的影响最为严重，这三个月估计大约损失62亿元左右．从我们的模型预测结果来计算，4、5、6三个月的损失为60．3亿元，这个数据基本与专家的估算值相符，8月份基本恢复正常，这也说明了模型的正确性和可靠性．
- 对于旅游业来说是受影响最严重的行业之一，最严重的4、5、6、7四个月就损失100多万人，按最新统计数据，平均每人消费1002美元计算，大约损失1O亿美元．全年大约损失160万人，约合16亿美元，到年底基本恢复正常．

- 对于综合服务业中的部分行业影响较大，如航空交通运输、宾馆餐饮等，但有些行业影响不大，如电信、通讯等，总平均来看，影响还不算太大，5、6、7、8四个月大约损失70亿元.
- 从预测结果可以看出，虽然下半年没有发生疫情，但人们一直担心SARS会卷土重来，所以，对这些行业还是有一定的影响，即SARS影响的延续性的作用.
- 该模型虽是就某经济指标的发展规律进行评估预测而建立的，但类似的也适用于其他方面的一些数据规律的评估预测问题，即该模型具有很广泛的应用性.

课件代码下载地址

更多模型、代码、优秀论文等请加QQ群：1077734962，更多资料请关注微信公众号：科研交流

THE END