厦门大学
XIAMEN UNIVERSITY

灰色系统分析方法

谭　忠

目录
CONTENTS

Part 1 源头问题与当今应用

12.1源头问题与当今应用

干热风是小麦接近成熟期的一种严重的自然灾害，可引起小麦过早枯死，导致大面积减产2~4成．若能提前预测干热风发生，就可以采取一些补救措施，如种植早熟品种，或提前喷施抗风、催熟农药等，以减少损失．下表是该地区历年干热风发生的日期．如干热风发生在5月29日以前成灾．试根据表格数据进行预测．

年	1975	1976	1977	1978	1979	1980	1981
月	6	5	6	6	5	5	6
日	3	25	7	1	29	26	5

年	1982	1983	1984	1985	1986	1987	1988
月	5	6	5	5	5	5	5
日	27	4	24	31	28	25	25

Part 2 灰色思想与建模方法

知识介绍

01 灰数的概念及其表示法

02 灰色关联度

03 灰色模型GM

04 灰色预测

11.2灰色思想与建模方法

从信息的完备性与模型的构建上看，工程技术等系统具有较充足的信息量，其发展变化规律明显，定量描述比较方便，结构与参数比较具体，人们称之为**白色系统**;

一个系统内部的特性完全未知，则称之为**黑色系统。**

对于另外一类系统诸如社会系统之类的不能建立起客观的数学模型，其作用原理也不明确，内部因素难以辨识或者之间关系隐蔽，人们很难准确了解这类系统的行为特征，这类系统称之为**灰色系统。**

灰色系统理论以“部分信息已知，部分信息未知”的“小样本”、“贫信息”不确定性系统为研究对象. 主要对“部分”已知信息的生成、开发，提取有价值的信息，实现对系统运行行为、演化规律的正确描述和有效监控. 灰色系统模型对实验观测数据没有什么特殊的要求和限制，因此应用领域十分宽广.

一、 灰数的概念及其表示法

灰数是灰色系统的基本“单元”或者“细胞”. 所谓的灰数是指信息不完全的数. 在应用中，灰数实际上是指在某个区间或者某个一般的数集内取值不确定数. 通常用记号⊗表示灰数.

例如，某个小姑娘的身高大约有165cm左右，体重只有40kg左右. 这里的165cm左右和40kg左右都是指灰数，分别记为⊗**(165)**和 ⊗**(40).**

再如，他的体温大约在38度~39度之间，关于体温是灰数，记为⊗**(T)[38,39].**

二、灰色关联度

1. 灰色关联度分析

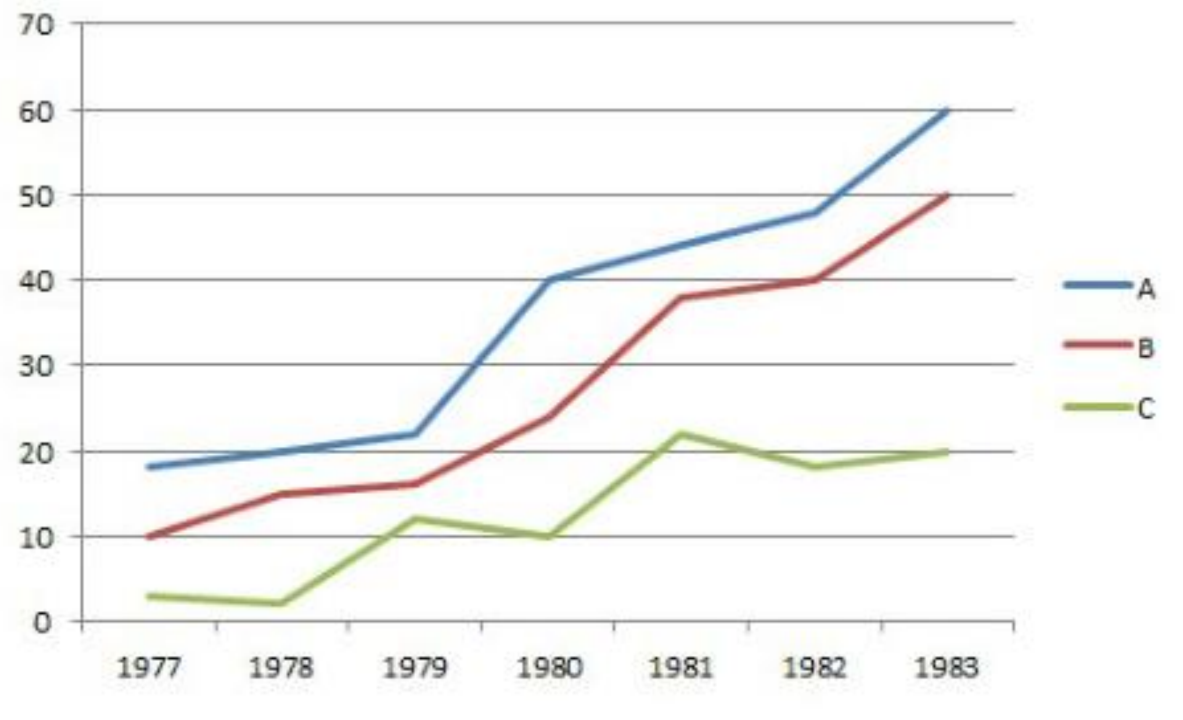
关联度是表征两个事物的关联程度，基本思想是根据序列曲线几何形状的相似程度来判断其联系是否紧密，曲线越接近，相应序列之间的关联度就越大，反之就越小.

具体来说，关联度是因素之间关联性大小的量度，它定量的描述了因素之间相对变化的情况. 如果两者在发展过程中，相对变化基本一致，则认为两者关联度较大，否则认为关联度较小.

例： 某地区1977~1983年的总收入与养猪、养兔收入的资料见表.

	1977	1978	1979	1980	1981	1982	1983
总收入	18	20	22	40	44	48	60
养猪	10	15	16	24	38	40	50
养兔	3	2	12	10	22	18	20

将表中的数据用图像表现出来



从上图易看出，曲线A（总收入）与曲线B（养猪收入）发展趋势比较接近，而与曲线C（养兔收入）相差较大，因此可以判断，该地区对总收入影响较直接的是养猪业，而不是养兔业.

事实上，这种直观的几何形状的判断比较，是比较粗糙的. 并且，如果好几条形状相差不大，或者在某些区间形状比较接近，就很难用直接观察的办法来判断各曲线间的关联词程度.

关联度分析的目的，是在影响某参考数列x_0的诸因素x_i中找出主要因素. 也就是按对x_0的关联程度大小对x_0进行排序. 关联序则反映各评价对象对理想(标准)对象的接近次序，即评价对象的优劣次序，其中，灰色关联度最大的评价对象为最佳. 因此，利用灰色关联度可对评价对象的优劣进行分析比较.

下面介绍最常用的衡量因素间关联度大小的量化方法.

先要制定参考的数据列(母因素时间数列)，参考数列常记为x_0，一般表示为

$$x_0 = (x_0(1), x_0(2), \cdots, x_0(n))$$

关联分析中被比较的数列(子因素时间数列)常记为x_i，一般表示为

$$x_i = (x_i(1), x_i(2), \cdots, x_i(n)), \quad i = 1, 2, \cdots, m$$

$$\xi_i(k)=\frac{\min_i\min_k|x_0(k)-x_i(k)|+\rho\max_i\max_k|x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\rho\max_i\max_k|x_0(k)-x_i(k)|}$$

$\xi_i(k)$: 是第k个时刻比较曲线x_i与参考曲线x_0的相对差值，这种形式的相对差值称为x_i对x_0在k时刻的关联系数.

其中ρ为分辨系数，$\rho\in[0,1]$，引入它是为了减少极值对计算的影响. 在实际使用时，一般选用ρ=0.5.

这里的关联系数只表示各个时刻数据间的关联程度，由于关联系数的数很多，信息过于分散，不便于比较，为此，有以下定义：

$$r_i=\frac{1}{n}\sum_{k=1}^{n}\xi_i(k)$$

为数列xi对参考数列x0的关联度.

XIAMEN UNIVERSITY

这里的关联系数只表示各个时刻数据间的关联程度，由于关联系数的数很多，信息过于分散，不便于比较，为此，有以下定义：

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k)$$

为数列xi对参考数列x0的关联度.

一般来讲，实际问题中，不同数列往往具有不同的量纲，为我们在计算关联数的时候要求量纲要相同. 因此，需要首先对各种数据进行无量纲化. 使其消除量纲和具有可比性.

无量纲化的方法，常用的有初值化、均值化、区间值变换.

设有序列 $x=(x(1),x(2),\cdots,x(n))$，称映射

$$f(x(k))=y(k)\quad k=1,2,\cdots,n$$

为序列x到y的变换.

（1）初值化

$$f(x(k))=\frac{x(k)}{x(1)}=y(k) \qquad x(1)\neq 0$$

（2）均值化

$$f(x(k))=\frac{x(k)}{\bar{x}}=y(k) \qquad \bar{x}=\frac{1}{n}\sum_{k=1}^{n}x(k)$$

（3）区间值变换

$$f(x(k))=\frac{x(k)-\min\limits_{k}x(k)}{\max\limits_{k}x(k)-\min\limits_{k}x(k)}$$

2、灰色关联度分析具体步骤：

(1)确定比较数列(评价对象)和参考数列(评价标准)，

设评价对象为m个，评价指标为n个，比较数列为：

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

参考数列为： $X_0 = \{X_0(k) \mid k = 1, 2, \cdots, n\}$

(2)确定各指标值对应的权重，可利用层次分析法等确定各指标对应的权重： $W = \{W_k \mid k = 1, 2, \cdots, n\}$

其中W_k为第k个评价指标对应的权重.

(3)计算灰色关联系数 $\xi_i(k)$

$$\xi_i(k)=\frac{\min\limits_i \min\limits_k |x_0(k)-x_i(k)|+\rho\max\limits_i \max\limits_k |x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\rho\max\limits_i \max\limits_k |x_0(k)-x_i(k)|}$$

其中常数$\rho\in[0,1]$, 称为分辨率系数.

当ρ越大的时候，分辨率也越大；当ρ越小，分辨率就越小.

一般情况下取 ρ=0.5. $\xi_i(k)$是比较数列x_i与参考数列x_0在第k个评价指标上的相对差值.

(4)计算灰色加权关联度

这里的关联系数只表示各个时刻数据间的关联程度，由于关联系数的数很多，信息过于分散，不便于比较，因此我们有如下定义：

$$r_i = \sum_{k=1}^{n} W_k \xi_i(k)$$

ri为第i个评价对象对理想对象的灰色加权关联度.

(5)评价分析

根据灰色加权关联度的大小，对各评价对象进行排序，即建立评价对象的关联序，关联度越大其评价效果越好.

例：某矿务局有5对矿井，年终考核各矿企业管理情况，5对矿井的各项指标实际数据列于表中.

产量:以评判期计划产量为100,希望实现完成的百分数越大越好;

掘进:以评判期计划掘进进尺为100,希望完成的百分数越大越好;

工效:以评判期计划全员工效为100,希望工效越高越好;

质量:以评判期商品煤计划含矸率为100,希望实际含矸率越低越好;

成本:以评判期计划吨煤成本为100,希望实际成本越低越好;

安全因素:以评判期事故率为100,希望实际事故率越低越好.

对于上述6项评价指标，采用权重按上述指标出现的先后顺序依次为

W=(W1,W2,W3,W4,W5,W6)=(0.2,0.2,0.1,0.15,0.15,0.2)

指标/%	一矿	二矿	三矿	四矿	五矿
产量	123.2	112.2	92.2	118.4	87.5
掘进	90.4	114.4	91.1	120.5	85.5
工效	115.6	108.6	90.4	116.3	96.8
质量	100.5	85.2	100.7	85.7	120.5
成本	80.2	87.3	115.6	80.5	140.1
安全	0.858	0.914	0.946	0.606	0.806

(1)构造理想对象. 把各评价对象中每一项指标的最佳值作为理想对象的指标值. 最佳值从参加比选的被评对象中选取，对不同影响因素而言，有的指标以最大为好，有的指标则以最小为好. 以最佳值为基础，便可构造理想对象的指标值. x0=(123.2,120.5,116.3,85.2,80.2,0.606).

(2) 计算指标的关联度. （$\rho=0.5$）

$$\xi_i(k)=\frac{\min\limits_i\min\limits_k|x_0(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}{|x_0(k)-x_i(k)|+\rho\max\limits_i\max\limits_k|x_0(k)-x_i(k)|}$$

(3)计算加权关联度 $r_i=\sum_{k=1}^{6}W_k\xi_i(k)$

$$r_1=0.768, r_2=0.785, r_3=0.558, r_4=0.973, r_5=0.527$$

根据以上关联度建立关联序：

r4>r2>r1>r3>r5

可见，四矿管理水平最高，二矿次之，一矿第三，三矿第四，五矿最差.

3、优势分析

参考数列不只一个，被比较的数列也不止一个时，需要进行优势分析．假设有m个参考数列，记为y1,y2,…,ym，再假设有l个比较数列，记为x1,x2,…,xl．显然，每个参考数列对l个比较数列有l个关联度，假设r_{ij}表示比较数列xj对参考数列yi的关联度，可以构造关联矩阵$R = (r_{ij})_{mXl}$

根据矩阵R的各个元素大小，可判断出哪些因素起主要影响，哪些因素起次要影响. 起主要影响的称之为优势因素.

例：某地区有6个母因素yi(i=1,2,...,6)，5个子因素x_j(j=1,2,...,5)如下

x_1 :固定资产投资
x_2 :工业投资
x_3 :农业投资
x_4 :科技投资
x_5 :交通投资

y_1 :国民收入
y_2 : 工业收入
y_3 : 农业收入
y_4 : 商业收入
y_5 : 交通收入
y_6 : 建筑业收入

其数据见表:

	1979年	1980年	1981年	1982年	1983年
x_1	308.58	310	295	346	367
x_2	195.4	189.9	187.2	205	222.7
x_3	24.6	21	12.2	15.1	14.57
x_4	20	25.6	23.3	29.2	30
x_5	18.98	19	22.3	23.5	27.655
y_1	170	174	197	216.4	235.8
y_2	57.55	70.74	85.38	99.83	103.4
y_3	88.56	70	85.38	99.83	103.4
y_4	11.19	13.28	16.82	18.9	22.8
y_5	4.03	4.26	4.34	5.06	5.78
y_6	13.7	15.6	13.77	11.98	13.95

通过计算出各个子因素对母因素的关联度，（这里ρ=0.5）从而得到关联矩阵

$$
\boldsymbol{R}=\begin{bmatrix}
0.802 & 0.761 & 0.557 & 0.810 & 0.936\\
0.689 & 0.666 & 0.529 & 0.885 & 0.800\\
0.891 & 0.858 & 0.579 & 0.577 & 0.675\\
0.678 & 0.663 & 0.568 & 0.780 & 0.731\\
0.811 & 0.774 & 0.565 & 0.804 & 0.921\\
0.743 & 0.766 & 0.562 & 0.607 & 0.632
\end{bmatrix}
$$

从关联矩阵R可以看出：

(1) 第四行元素都比较小，表明各种投资对商业收入的影响不大，即商业是一个不需要依赖外资而能自行发展的行业. 从消耗投资上看，这是劣势，但从少投资多收入的效益观点看，商业是优势.

(2) $r_{15}=0.936$最大，表明交通投资的多少对国民收入的影响最大. 也可以从此看出交通的影响.

(3) r_{55}仅次于r_{15}，表明交通收入主要取决于交通投资，这是很自然的.

(4) 在第四列中r24最大，表明科技对工业的影响最大；而r34=0.577是该列中最小的，表明从全面来衡量，还没有使科技投资与农业经济挂上钩，即科技投资针对的不是农村需要的科技.

(5)第三行的前三个元素比较大，表明农业是个综合性行业，需要其他方面的配合，例如：r31=0.891表明固定资产能够较大地促进农业的发展. 另外，r32=0.858表明农业发展与工业投资也是密切相关的.

三、灰色模型GM

如果一个系统具有层次、结构关系的模糊性，动态变化的随机性，指标数据的不完备或不确定性，则称这些特为灰色性. 具有灰色性的系统称为灰色系统. 在灰色系统理论中，利用较少的或不确切的表示灰色系统行为特征的原始数据序列作生成变换后建立的，用以描述灰色系统内部事物连续变化过程的模型，称为灰色模型，简称GM(Grey Model)模型.

1、灰色序列生成

灰色系统是通过对原始数据的挖掘、整理来寻求其变化规律的，这是一种就数据寻找数据的现实规律的途径，我们称之为灰色序列生成. 灰色系统理论认为，尽管客观系统表象复杂，数据离乱，但它总是有整体功能的，因此必然蕴含着某种内在的规律.一切灰色序列都能通过某种生成弱化其随机性，显示其规律性.

序列的生成方式有多种：累加生成、累减生成、加权累加等等.

累加生成是使灰色过程由灰变白的一种方法，它在灰色系统理论中占有极其重要的地位. 通过累加可以看出，灰量积累的发展态势，使离乱的原始数据中蕴含的积分特性或规律充分显示出来.

比如对于一个家庭的支出，若按日计算的，可能没什么明显的规律，若按月计算，支出的规律性就可能体现出来，它大体与月工资收入成某种关系；一种农作物的单粒重，一般来说没有什么规律，人们常用千粒重作为农作物品种的评估标准.

(1)累加生成数列：把数列各个时刻数据一次累加的过程称为累加生成数列，记为AGO. 由累加生成过程所得的新数列称为累加生成数列.

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

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

则称$x^{(1)}(k)$为数列$x^{(0)}$的1-次累加生成. 数列

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

称为$x^{(0)}$的1- 次累加生成数列.

类似地有，

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

称为$x^{(0)}$的r-次累加生成. $x^{(r)}=\left(x^{(r)}(1),x^{(r)}(2),\cdots,x^{(r)}(n)\right)$

称为$x^{(0)}$的r-次累加生成数列.

对于非负序列，累加次数越多，则随机性弱化越多，当累加次数足够大后，可认为时间序列已由随机序列变为非随机序列了.

(2)累减生成数列：对原始数列一次做相邻两个数据的相减运算过程称为累减生成过程，记为IAGO.

设原始数列为

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

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

则称 $x^{(0)}(k)$ 为数列$x^{(1)}$的1-次累减生成. 若$x^{(r)}$为r-AGO数列，则称

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

为r次累减生成.

(3)均值生成:

设原始数列

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

称 $x^{(0)}(k-1)$ 与$x^{(0)}(k)$ 为数列$x^{(0)}$的邻值， $x^{(0)}(k-1)$

为后邻值， $x^{(0)}(k)$ 为前邻值.

对于常数 $\alpha\in[0,1]$，则称 $z^{(0)}(k)=\alpha x^{(0)}(k)+(1-\alpha)x^{(0)}(k-1)$

为由数列$x^{(0)}$的邻值在生成系数α下邻值生成数。

特别地，当生成系数α=0.5时，则称

$$z^{(0)}(k) = 0.5x^{(0)}(k) + 0.5x^{(0)}(k-1)$$

为邻值生成数，即等权邻值生成数.

类似地，可以定义非邻值生成数：

$$z^{(0)}(k) = \alpha x^{(0)}(k+1) + (1-\alpha)x^{(0)}(k-1)$$

和

$$z^{(0)}(k) = 0.5x^{(0)}(k+1) + 0.5x^{(0)}(k-1)$$

而数列

$$z^{(0)} = \left(z^{(0)}(1), z^{(0)}(2), \cdots, z^{(0)}(n)\right)$$

称为均值生成数列.

2、GM(1,1)

2.1 GM（1，1）的定义：

设$x^{(0)}$是n个元素的数列 $x^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)\right)$，

$x^{(0)}$的AGO生成数列 $x^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\right)$

其中 $x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i),\ k = 1, 2, \cdots, n,$ 则定义$x^{(1)}$的灰导数为

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

令$z^{(1)}$为$x^{(1)}$的紧邻均值生成序列，

$$z^{(1)}=\left(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\right)$$

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

于是定义GM（1,1）的灰微分方程模型为

$$\mathrm{d}x^{(1)}(k)+az^{(1)}(k)=b$$

即

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

其中 $x^{(0)}(k)$ 为灰导数，a为发展系统， $z^{(1)}(k)$ 为白化背景值，b称为灰作用量. 将时刻k=2,3,…,n代入上式中有

$$
\left\{
\begin{array}{c}
x^{(0)}(2) + az^{(1)}(2) = b, \\
x^{(0)}(3) + az^{(1)}(3) = b, \\
\cdots\cdots\cdots\cdots \\
x^{(0)}(n) + az^{(1)}(n) = b,
\end{array}
\right.
$$

令 $Y_N = (x^{(0)}(2), x^{(0)}(3), \cdots, x^{(0)}(n))^T, u = (a, b)^T$

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

则YN为数据向量，B为数据矩阵，u为参数向量，则GM（1,1）可以表示为矩阵方程

$$Y_N = Bu.$$

对于参数向量u的确定方法：

由最小二乘法则来计算 $\hat{u} = (B^T B)^{-1} B^T Y_N$

具体地 $\hat{a} = \frac{CD-(n-1)E}{(n-1)F-C^2}, \hat{b} = \frac{DF-CE}{(n-1)F-C^2}$

其中

$$C=\sum_{k=2}^{n} z^{(1)}(k),\ D=\sum_{k=2}^{n} x^{(0)}(k),\ E=\sum_{k=2}^{n} z^{(1)}(k)x^{(0)}(k),$$

$$F=\sum_{k=2}^{n}\left(z^{(1)}(k)\right)^{2}.$$

2.2 GM(1,1)的白化型

对于GM(1,1)的灰微分方程，如果 $x^{(0)}(k)$ 的时刻k=2,3,…,n视为连续的变量t，则数列 $x^{(1)}$ 就可以视为时间t的函数，记为 $x^{(1)} = x^{(1)}(t)$ ，并让灰导数 $x^{(0)}(k)$ 对应于导数 $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}$,背景值 $z^{(1)}(k)$ 对应于 $x^{(1)}(t)$.

于是得到GM（1,1）的灰微分方程对应的白微分方程为

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = b$$

称之为GM（1,1）的白化型.

值得注意的是：GM(1,1)的白化型并不是由GM(1,1)的灰微分方程直接导出的，它仅仅是一种“借用”或“白化默认”.

另一方面，GM(1,1)的白化型是一个真正的微分方程，如果白化型模型精度高，则表明所用数列建立的模型GM(1,1）与真正的微分方程吻合较好，反之亦然.

设B,Y,u如上所述,

$$[a,b]^T=(B^TB)^{-1}B^TY$$

则有:

1. 白化方程 $\frac{dx^{(1)}}{dt}+ax^{(1)}=b$ 的解

$$x^{(1)}(t)=\left(x^{(1)}(t_0)-\frac{b}{a}\right)e^{-a(t-t_0)}+\frac{b}{a}$$

2.取 $t_0=1$，对等间隔取样的离散值的时间相应数列为

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

当k=1,2,...,n-1时，$\hat{x}^{(1)}(k+1)$ 为拟合值，当k>=N时，计算出的 $\hat{x}^{(1)}(k+1)$ 为预测值.

3.还原值，根据 $\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$

可还原 $\hat{x}^{(0)}(k+1)$ 的拟合值和预报值.

四、灰色预测

1. 灰色预测模型的检验

(1) 相对误差检验

现设原始序列：

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

相应的预测模型模拟序列：

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

残差序列:

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

相对应的相对误差序列:

$$\Delta=\left(\left|\frac{\varepsilon(1)}{x^{(0)}(1)}\right|,\left|\frac{\varepsilon(2)}{x^{(0)}(2)}\right|,\cdots,\left|\frac{\varepsilon(n)}{x^{(0)}(n)}\right|\right)=\{\Delta_k\}_1^n$$

则

(i) 对于 $k \leq n$ 称 $\Delta_k = \left|\frac{\varepsilon(k)}{x^{(0)}(k)}\right|$ 为k点模拟误差，称

$\bar{\Delta} = \frac{1}{n}\sum_{k=1}^{n}\Delta_k$ 为平均相对误差.

(ii) 称 $1-\bar{\Delta}$ 为平均相对精度， $1-\Delta_k$ 为k点的模拟精度，k=1,2,…,n；

(iii) 给定α,当$\bar{\Delta} < \alpha$且$\Delta_n < \alpha$成立时，称模型为残差合格模型.

(2) 均方差比和小误差概率检验

设 $x^{(0)}$ 为原始序列， $\hat{x}^{(0)}$ 为相应的模拟序列， $\varepsilon^{(0)}$ 为残差序列，$x^{(0)}$ 的均值、方差分别为

$$\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x^{(0)}(k) \qquad S_1^2 = \frac{1}{n}\sum_{k=1}^{n}(x^{(0)}(k) - \bar{x})$$

残差的均值、方差分别为

$$\bar{\varepsilon} = \frac{1}{n}\sum_{k=1}^{n}\varepsilon(k) \qquad S_2^2 = \frac{1}{n}\sum_{k=1}^{n}(\varepsilon(k) - \bar{\varepsilon})$$

(i) 若其中 $C=\frac{S_2}{S_1}$称为方差比值. 对于给定的C0，当C<C0时，称模型为均方差比合格模型；

(ii) 若其中$p=P(|\varepsilon(k)-\bar{\varepsilon}|<0.6745S_1)$称为小误差概率. 对于给定的p0>0当p>p0时，称模型为小误差概率合格模型.

这两种方法都是对残差的考察来判断模型的精度. 其中平均相对误差和模拟误差都要求越小越好，均方差比值C越小越好（因为C小说明S2小，S1大，即残差方差小，原始数据方差大，说明残差比较集中，摆动幅度小，原始数据

比较分散，摆动幅度的，所以模拟效果好要求S2与S1相比尽可能小）,以及小误差概率p越大越好.

若给定 $\alpha, \varepsilon_0, C_0, p_0$的一组取值，我们就确定了检验模型模拟精度的一个等级. 常用的精度等级见表. 一般情况下，最常用的是相对误差检验指标.

Table 1: 精度检验等级参照表

指标临界值 / 等级精度	相对误差α	均方差比值C_0	小误差概率p_0
一级	0.01	0.35	0.95
二级	0.05	0.5	0.8
三级	0.1	0.65	0.7
四级	0.2	0.8	0.6

XIAMEN UNIVERSITY

2. 数列预测

数列预测就是对系统变量的未来行为进行预测，常用的数列预测模型是GM(1,1)模型. 根据实际情况，也可以考虑采用其他灰色模型.

在定性分析的基础上，定义适当的序列算子，然后建立GM(1,1)模型，通过精确度检验之后，即可用来预测.

灰色预测的步骤:

(1) 数据的检验与处理

首先，为了保证建模方法的可行性，需要对已知数列做必要的检验处理. 设参考数据为 $x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$,

计算数列的级比

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

如果所有的级比$\lambda(k)$都落在可容覆盖$\Theta=(e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$，

则数列$x^{(0)}$可以作为模型GM(1,1)的参考数列和进行数据灰色预测.

否则,需要对数列$x^{(0)}$做必要的变换处理,使其落入可容覆盖内.

取适当的常数c，作平移变换 $y^{(0)}(k)=x^{(0)}(k)+c$ (k=2,3,...n)

则使数列 $y^{(0)}=(y^{(0)}(1), y^{(0)}(2), \cdots, y^{(0)}(n))$ 的级比

$$\lambda_y(k)=\frac{y^{(0)}(k-1)}{y^{(0)}(k)}\in\Theta \qquad (k=2,3,\cdots,n)$$

(2)建立模型GM(1,1)

(3) 检验预测值

(4)预测预报：由模型所得到的指定区间内的预测值，根据实际问题需要，给出相应的预测预报.

3. 灾变预测

灰色灾变预测实质上是对异常值预测. 什么样的值看作异常值，往往是人们根据主观经验和历史值来确定. 灾变预测的任务就是要给出下一个或者下几个异常值出现的时刻，以便人们提前做好准备，采取预防性对策.

现设x为原始序列 $x_\xi = (x[q(1)], x[q(2)], \cdots, x[q(m)])$ 为灾变序列，

则称 $Q^{(0)} = (q(1), q(2), \cdots, q(m))$为灾变日期序列.

灾变预测就是要通过对灾变日期序列的研究，来寻找其规律性，预测以后若干次灾变发生的日期. 灰色系统的灾变预测是通过对灾变日期序列建立GM(1,1)模型来实现的.

设 $Q^{(0)} = (q(1), q(2), \cdots, q(m))$ 为灾变日期序列，其一次累加序列为 $Q^{(1)} = (q^{(1)}(1), q^{(1)}(2), \cdots, q^{(1)}(m))$,

其中 $q^{(1)}(i)=\sum_{j=1}^{i}q(j), \quad i=1,2,\cdots,m$

$Q^{(0)}$ 的紧邻均值生成序列为z(1). 则称 $q(k)+az^{(0)}=b$ 为灾变GM(1,1)模型.

现设原始序列为 $x_\xi=(x[q(1)],x[q(2)],\cdots,x[q(m)])$, n为日期. 给定某一异常值 ξ，相应的灾变日期序列 $Q^{(0)}=(q(1),q(2),\cdots,q(m))$

其中 $q(m)(\leq n)$ 为最近一次灾变发生的日期，则称 $\hat{q}(m+k)$

为未来第k次灾变的预测日期.

4. 季节灾变预测

设 $\Omega=[a,b]$ 为总时区，若 $\omega_i=[a_i,b_i]\subset[a,b], i=1,2,\cdots,s,$

满足

$$\Omega=\cup_{i=1}^{s}\omega_i;\omega_i\cap\omega_j=\emptyset$$

任意的$j\neq i$，则称$\omega_i(i=1,2,\cdots,s)$ 为Ω 中的季节，也称时段或区间.

设 $\omega_i\subset\Omega$为一个季节，原始序列

$$x=(x(1),x(2),\cdots,x(n))$$

对给定的异常值 ξ，称与之对应的灾变序列

$$x_{\xi}=(x[q(1)],x[q(2)],\cdots,x[q(m)])$$

相应地，我们称 $Q^{(0)}=(q(1),q(2),\cdots,q(m))$ 为季节灾变日期序列.

季节灾变预测可以按照以下步骤进行：

第一步：给出原始序列

$$x=(x(1),x(2),\cdots,x(n));$$

第二步：研究原始序列数据的变化范围，确定季节 $\omega_i = [a_i, b_i]$;

第三步：令 $y(k) = x(k) - a_i$, 化原始序列为

$$y = (y(1), y(2), \cdots, y(n)),$$

以提高数据分辨率；

第四步：给定异常值 ξ 找出季节灾变序列 $y_\xi = (y[q(1)], y[q(2)], \cdots, y[q(m)])$

第五步：建立灾变GM(1,1)模型；$q(k) + az^{(1)}(k) = b$;

第六步：检验模拟精度，进行预测.

Part3 案例分析

案例分析

案例一 某城市道路交通噪声预测

案例二 某地区小麦干热风预测

案例1：某城市道路交通噪声预测

北方某城市1986~1992年道路交通噪声平均声级数据见下表:

序号	年份	L_{eq}	序号	年份	L_{eq}
1	1986	71.1	5	1990	71.4
2	1987	72.4	6	1991	72
3	1988	72.4	7	1992	71.6
4	1989	72.1			

解：第一步：级比检验

建立交通噪声平均声级时间序列如下

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(7))$$
$$= (71.1, 72.4, 72.4, 72.1, 71.4, 72, 71.6)$$

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

$$= (0.982, 1, 2.0042, 1.0098, 0.9917, 1.0056)$$

(2)级比判断

由于所有的 $\lambda(k) \in [0.982, 1.0098], k = 2, 3, \cdots, 7,$

故可以用x(0)作满意的GM(1,1)建模。

第二步：GM(1,1)建模

(a)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) (b)构造数据矩阵B及数据向量Y

$$B=\begin{bmatrix} -0.5(x^{(1)}(1)+x^{(1)}(2)) & 1 \\ -0.5(x^{(1)}(2)+x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(x^{(1)}(6)+x^{(1)}(7)) & 1 \end{bmatrix}, Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(7) \end{bmatrix}$$

(c)计算 $\hat{u}$

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

于是得到a=0.0023,b=72.6573。

(d)建立模型

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+0.0023x^{(1)}=72.6573,$$

并求解

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

(e)求生成数列预测值$\hat{x}^{(1)}(k+1)$及模型还原值$\hat{x}^{(0)}(k+1)$:

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

$$\hat{x}^{(0)}(k)=\hat{x}^{(1)}(k)-\hat{x}^{(1)}(k-1), k=2,3,\cdots,7$$

得

$$
\begin{aligned}
\hat{x}^{(0)} &= (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \cdots, \hat{x}^{(0)}(7)) \\
&= (71.1, 72.4, 72.2, 72.1, 71.9, 71.7, 71.6)
\end{aligned}
$$

第三步：模型检验

模型的各种检验指标值得计算结果（见下表）.

经检验，该模型的精度较高，可进行预测和预报.

序号	年份	原始值	模型值	残差	相对误差
1	1986	71.1	71.1	0	0
2	1987	72.4	72.4057	-0.0057	0.01%
3	1988	72.4	72.2362	0.1638	0.23%
4	1989	72.1	72.0671	0.0329	0.05%
5	1990	71.4	71.8984	-0.4984	0.7%
6	1991	72 .0	71.7301	0.2699	0.37%
7	1992	71.6	71.5622	0.0378	0.05%

厦门大学
XIAMEN UNIVERSITY

案例2 某地区小麦干热风预测

问题背景：干热风是小麦接近成熟期的一种严重的自然灾害，可引起小麦过早枯死，导致大面积减产2~4成. 若能提前预测干热风发生，就可以采取一些补救措施，如种植早熟品种，或提前喷施抗风、催熟农药等，以减少损失. 下表是该地区历年干热风发生的日期. 如干热风发生在5月29日以前成灾. 试根据表格数据进行预测.

年	1975	1976	1977	1978	1979	1980	1981
月	6	5	6	6	5	5	6
日	3	25	7	1	29	26	5

年	1982	1983	1984	1985	1986	1987	1988
月	5	6	5	5	5	5	5
日	27	4	24	31	28	25	25

第一步：以1月1日为起点，由表可得原始数据序列：

$$x=(x(1),x(2),\cdots,x(14))$$
$$=(185,176,189,183,180,177,187,178,186,175,182,179,176,176)$$

第二步：取 $\omega=$[5月20日，6月10日]$=[171,192]$，则$x\subset\omega$

第三步：令 $y(k)=x(k)-171;k=1,2,\cdots,14.$

将x化为 $y=\{y(k)\}_1^{14}=(14,5,18,12,9,6,16,7,15,4,11,8,5,5)$

第四步：取下限异常值 $\xi = 9$(5月20日至5月29日)，得季节灾变序列

$$
\begin{aligned}
y_\xi &= (y[q(1)], y[q(2)], \cdots, y[q(8)]) = (y[q(k)] \mid y[q(k)] \le 9) \\
&= (y(2), y(5), y(6), y(8), y(10), y(12), y(13), y(14))
\end{aligned}
$$

于是 $$q^{(0)} = (2, 5, 6, 8, 10, 12, 13, 14)$$

第五步：设 $q(k) + az^{(1)}(k) = b$，则

$$
\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0.1588 \\ 5.017 \end{bmatrix}
$$

由此的灾变GM(1,1)序号相应式

$$\begin{cases} \hat{q}^{(1)}(k+1) = 33.59e^{0.1588k} - 31.59 \\ \hat{q}(k+1) = \hat{q}^{(1)}(k+1) - \hat{q}^{(1)}(k), k \geq 1 \end{cases}$$

即

$$\hat{q}(k+1) = 33.59(1 - e^{-0.15888})e^{0.15888(k-1)} = 4.93e^{0.15888(k-1)}$$

第六步：由 $\hat{q}(k+1) = 4.93e^{0.15888(k-1)}$ 得到Q(0)的模拟矩阵

厦门大学 XIAMEN UNIVERSITY

$$\hat{q}^{(0)} = (\hat{q}(1), \hat{q}(2), \cdots, \hat{q}(8)) \\ = (2, 5.77, 6.77, 7.93, 9.30, 10.91, 12.78, 14.98)$$

由 $\varepsilon(k) = q(k) - \hat{q}(k); k = 1, 2, \cdots, 8$, 得

$$\varepsilon^{(0)} = (0, -0.77, -0.77, 0.07, 0.07, 1.09, 0.22, -0.98)$$

再由

$$\Delta_k = \left|\frac{\varepsilon(k)}{q(k)}\right|; k = 1, 2, 3, \cdots, 8$$

得 $$\Delta = (0, 0.154, 0.128, 0.009, 0.07, 0.091, 0.017, 0.07)$$

平均相对误差

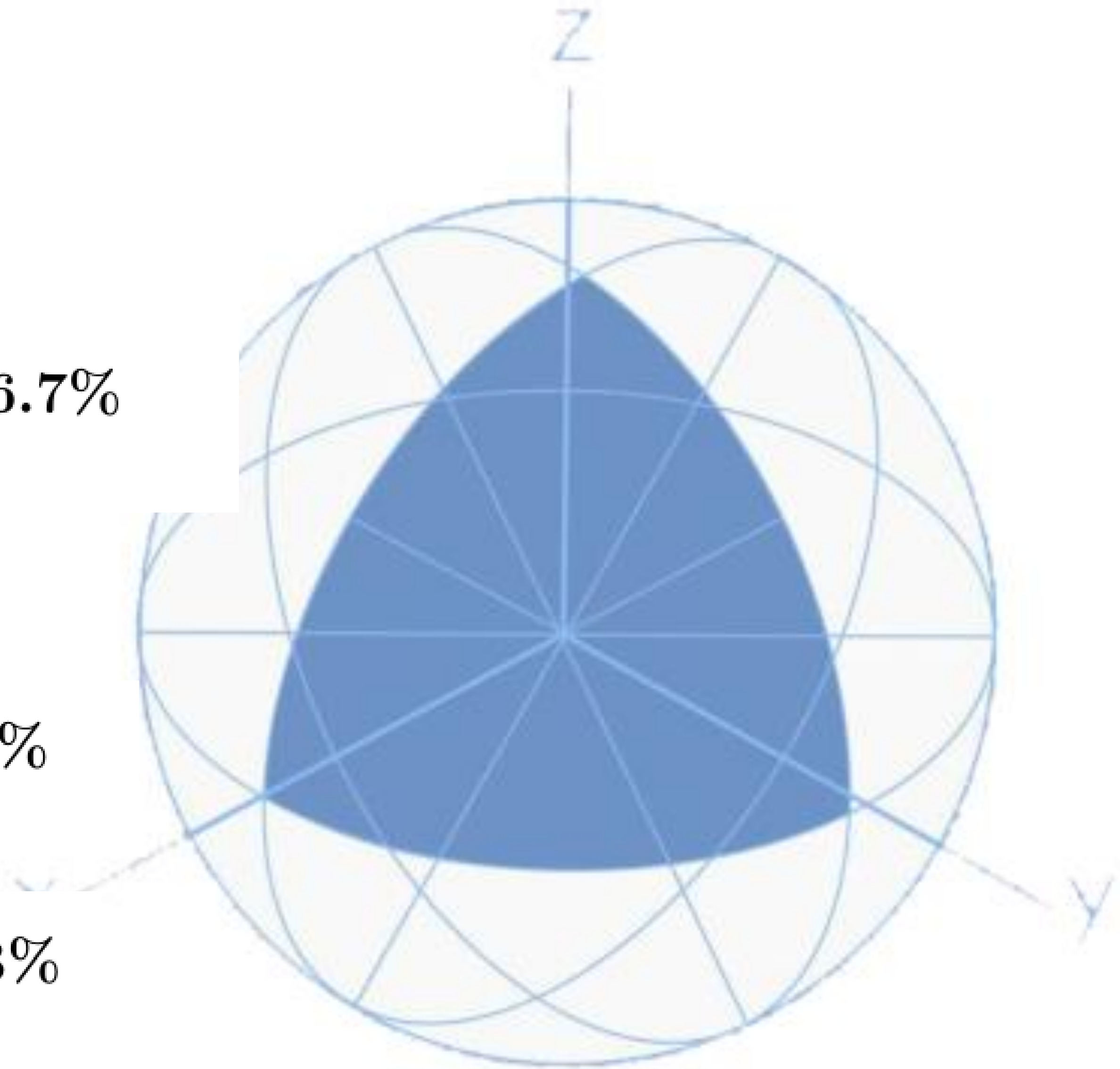
$$\bar{\Delta} = \frac{1}{8}\sum_{k=1}^{8}\Delta_k = 6.7\%$$

平均相对精度

$$1 - \bar{\Delta} = 93.3\%$$

模拟精度

$$1 - \Delta_8 = 93\%$$



精度接近二级，可做预测模型. 由

$$\hat{q}(9) \approx 17, \hat{q}(10) \approx 20$$

得

$$\hat{q}(9) - \hat{q}(8) \approx 17 - 14 = 3, \quad \hat{q}(10) - \hat{q}(9) \approx 20 - 17 = 3$$

即1991年、1994年干热风可能成灾.

THANK YOU