

# The Internal Connection between Grey GM(1,1) Model and Simple Linear Regression Model

Wen Liu, Yong Wang

College of Information Science and Engineering, Ocean University of China, Qingdao Shandong  
Email: 297619178@qq.com, markwy@126.com

Received: Apr. 25<sup>th</sup>, 2018; accepted: May 10<sup>th</sup>, 2018; published: May 17<sup>th</sup>, 2018

---

## Abstract

**Purpose:** The analog sequence of GM(1,1) model for the original sequence after translation transformation, is very similar to simple linear regression line. So, we will explore the relationship between the two models. **Method:** Get some original sequences from real world, construct GM(1,1) model after translation transformation and simple linear regression model, and use T test and difference measure to compare the two models. **Result:** The GM(1,1) model constructed by the original sequence after translation transformation, and the simple linear regression model constructed by the sequence except the first data, are the same. **Conclusion:** After translation transformation, the grey model becomes a special linear regression model.

## Keywords

GM(1,1) Model, Simple Linear Regression Model, Translation Transformation

---

# 灰色GM(1,1)模型与线性回归模型的内在联系

刘 雯, 王 勇

中国海洋大学信息科学与工程学院, 山东 青岛  
Email: 297619178@qq.com, markwy@126.com

收稿日期: 2018年4月25日; 录用日期: 2018年5月10日; 发布日期: 2018年5月17日

---

## 摘 要

**目的:** 对经过平移变换后的原始序列建立GM(1,1)模型, 其模拟序列与一元线性回归直线非常相似, 为

此探究两种模型间的联系。方法：在真实世界中取多组原始序列，进行平移变换后构建GM(1,1)模型，再与一元线性回归模型进行比较，使用T检验和差异度量方法对结果进行量化。结果：去掉原始序列第一个数据后用其余数据建立的一元线性回归模型和经过较大平移变换后基于全部原始序列建立的GM(1,1)模型相同。结论：经过平移变换后，灰色模型表现为一种特殊的线性回归模型。

## 关键词

GM(1,1)模型, 一元线性回归模型, 平移变换

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

软件成本预测是软件开发中重要的步骤之一，灰色模型和一元线性回归模型是软件成本预测中常用的预测模型。灰色模型是邓聚龙教授提出的灰色系统理论中的重要内容，灰色系统理论是针对小样本、贫信息的不确定系统进行研究，在医学、教育、工业、军事等领域有着广泛的应用[1]。线性回归是最基本和应用最广泛的预测方法。灰色理论在被提出后的三十多年来经过了快速成长与发展，线性回归模型是受到普遍认可的成熟模型，如果能在这两个模型上找到共通点，能够增强对灰色模型内在机理的了解，有助于灰色模型的普及。

在使用灰色模型建模时，为了提高预测模型的精度，人们一直在致力于研究新的建模技术和改进方法，研究表明，原始数据的模式及光滑性是影响灰色模型精度的两个主要因素[2]，所以为了提高模型的模拟和预测精度，会对模型采取一些处理措施，如文献[3]。对原始序列进行平移变换是其中的处理方法之一。平移变换即在建模之前，对原始序列的每一个数据加上一个数，构造一个新的序列，再对新的序列进行建模[4]。对单调递增数列进行平移变换，会使变换后数列的光滑性增强[5]，进而影响模型的预测误差，这是平移变换的核心作用。文献[4] [5] [6]中探讨了平移变换对数据建模的作用，得出平移变换可以提升模型的性能的结论。

作为灰色模型和线性回归模型中最典型的模型，近年来将 GM(1,1)模型和一元线性回归模型一起讨论的学者也很多。包括将两种模型结合起来进行数据的预测，弥补各自的缺点，能得到比较好的效果[7]，或者单纯比较两种模型的建模效果。但是将平移变换后的 GM(1,1)模型与一元线性回归模型进行比较并研究两者内在联系的研究还未发现。

本文是对灰色模型中的 GM(1,1)模型和一元线性回归模型相关性的研究，揭示了应用平移变换后的 GM(1,1)模型与一元线性回归模型有着高度的一致性，换句话说，平移变换可以将 GM(1,1)模型从某种意义上转化为一元线性回归模型，从而把两种模型在一定程度上统一起来，使灰色模型更易于让人理解和接受，对灰色系统理论的推广有积极的意义。本文将在第二节中阐述本文研究的问题、实验步骤和所用的检验方法。接下来通过实验得出结果，分析结果并得出最后的结论。

## 2. 研究问题与方法

### 2.1. 研究问题

实验发现，对经过平移变换的原始序列建立 GM(1,1)模型，随着平移量的增加，模拟序列曲线会趋

于平直, 最后基本成为一条直线。因此我们想要研究, 与图形为一条直线的一元线性回归模型相比较, 两种模型所得的建模结果有什么内在联系, 随着平移量的增加, 两者之间的联系会有什么变化。

## 2.2. 实验步骤

研究平移变换后 GM(1,1)模型和一元线性回归模型相似性的步骤如下:

- 1) 准备实验数据集, 对数据进行分组, 整理成所需的原始序列集。
- 2) 对原始序列进行平移变换。
- 3) 对原始序列构建一元线性回归模型, 对每次平移后得到的新序列分别构建 GM(1,1)模型。
- 4) 对每一组 GM(1,1)模型的模拟值与对应的一元线性回归模型的模拟值进行比较。
- 5) 分析随平移量的变化, 两种模型的相似性的变化趋势, 并得出结论。

## 2.3. 检验方法

本文在对原始序列进行平移变换后建立 GM(1,1)模型的建模结果和一元线性回归模型建模结果进行比较时, 采用两种方法进行研究:

1) 对平移变换后的 GM(1,1)模型所得数据与一元线性回归模型所得数据进行配对样本 T 检验, 每平移一次进行一次检验, 如果验证每次平移后, 配对样本 T 检验所得的 p 值均大于 0.05, 说明两种模型的结果不存在显著差异, 来自一个总体, 具有相似性。

2) 对两组实验数据进行差异度量。两组数据之间的差异可以用偏差和来表示。将 GM(1,1)模型每次平移后所得的结果分别与一元线性回归模型所得的结果作差并取绝对值, 将每组结果的差值绝对值相加, 若差值绝对值和越小, 则说明两组数据越接近, 反之, 两组数据的差距越大。

## 3. 实验与分析

### 3.1. 平移变换对 GM(1,1)模型的作用

我们首先用一组数据来验证原始序列进行平移变换后, 对 GM(1,1)模型建模结果的影响。

设  $X(k)$  为原始序列,  $X = [1, 1.5, 2.5, 4, 8.9, 17.7]$ 。在建模之前, 首先对原始序列进行光滑性检验。经检验, 所得的光滑比为:

$$\rho(3) = 1.0000, \rho(4) = 0.8000, \rho(5) = 0.9889, \rho(6) = 0.9888$$

当  $k > 3$  时, 光滑比均大于 0.5, 准光滑性条件不满足。

对原始序列进行平移变换是使其变光滑的一种手段。对原始序列平移 10 次, 每次平移的量分别为 0, 2, 4, 8, 16, 32, 64, 128, 256, 512。第一次平移量为 0, 即平移后序列  $X_{A1}$  与原始序列  $X$  相同, 第二次平移量为 2, 依次类推。得到 10 组新的序列为:

$$X_{A1} = \{1, 1.5, 2.5, 4, 8.9, 17.7\}$$

$$X_{A2} = \{3, 3.5, 4.5, 6, 10.9, 19.7\}$$

...

$$X_{A10} = \{513, 513.5, 514.5, 516, 520.9, 529.7\}$$

对平移后的 10 组新序列进行光滑性检验, 发现前 3 组数据, 即平移量为 0, 2, 4 时的  $X_{A1}, X_{A2}, X_{A3}$ , 仍不满足准光滑性条件; 当进行第 4 次平移, 即平移量达到 8 时, 序列  $X_{A4}$  所得的光滑比在  $k > 3$  时, 光滑比均小于 0.5, 准光滑性条件满足。此后的平移数组准光滑性也均满足。

尽管前三个序列不满足光滑性条件, 仍可应用 GM(1,1)模型得到建模结果, 为了更清楚地看到平移

变换后序列建模的变化趋势, 我们保留前三组, 用全部 10 组序列构建 GM(1,1)模型。

对  $X_{.1}$  到  $X_{.10}$  10 组序列进行 GM(1,1)模型的构建, 得到的 10 次建模的曲线如图 1 所示。

由图 1 首先可以看出, GM(1,1)模型建模后的第一个点不变, 一直都与原始序列的第一个点相同。从曲线的形状来看, 平移量为 0 时所得的 GM(1,1)模型原始曲线比较弯曲, 随着平移量的增加, 所得曲线弯曲的弧度变小, 逐渐变得平直, 进行到最后一次平移时, 第二到第六个点的曲线几乎成为一条直线。那我们就将这最后一次平移的曲线与一元线性回归模型曲线进行对比, 探究两种模型之间的联系。

序列  $X$  构建一元线性回归模型所得模拟值为  $f = [-1.7238, 1.3390, 4.4019, 7.4648, 10.5276, 13.5905]$ 。GM(1,1)模型最后一次平移所得结果为  $X'_{.10} = [1, -0.8400, 3.0400, 6.9200, 10.8000, 14.6800]$ 。两组数据所得图像如图 2。

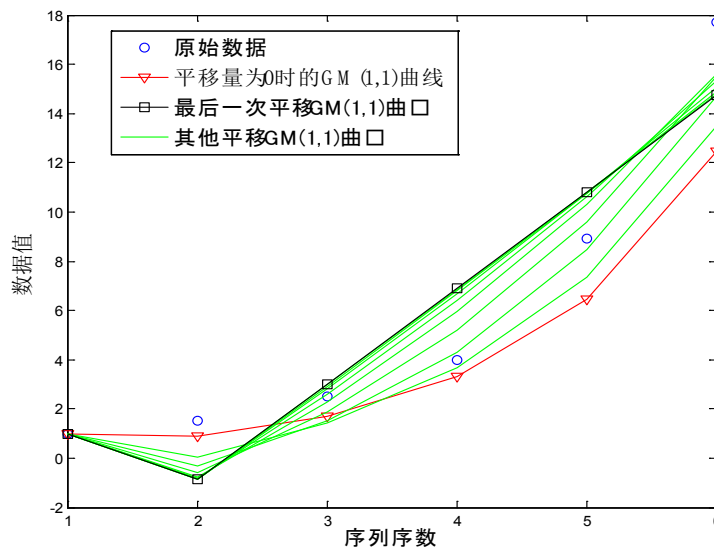


Figure 1. Curve: 10 translations of original sequence of GM(1,1) model  
图 1. 原始序列 10 次平移后建立 GM(1,1)模型曲线

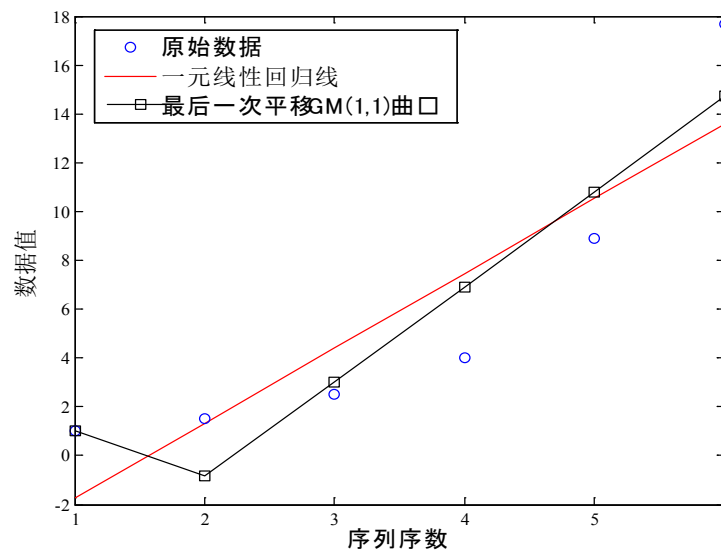


Figure 2. Curve: The last translational GM (1,1) model and one linear regression model  
图 2. GM(1,1)模型与一元线性回归模型比较曲线

如图 2 所示, 虽然经过平移变换后的 GM(1,1)模型曲线第二到第六个点部分几乎为一条直线, 但与一元线性回归模型的曲线并不重合。

这是因为, 在 GM(1,1)模型中, 原始序列第一个点虽然在建模中也进行了模拟, 但是从图 1 中发现, 在模拟序列中, 第一个点始终和原始序列中的点重合, 并没有发生变化, 而一元线性回归模型的所有点都发生了变化。既然 GM(1,1)模型中原始序列的第一个点对模拟序列没有影响, 那么构建一元线性回归模型时, 我们可以去掉原始序列  $X$  的第一个点, 用其余各点构建模型。因此, 在本实验中, 构建一元线性回归模型时使用原始序列的第二至六个点, 构建 GM(1,1)模型时使用原始序列的全部数据。此时得到的一元线性回归建模结果为  $f' = [-0.8416, 3.0087, 6.8881, 10.7967, 14.7347]$ , 与最后一次平移的 GM(1,1)模型图像如图 3 所示。

如图所示, 当去掉原始序列的第一个点再构建一元线性回归模型时, 一元线性回归模型的曲线与 GM(1,1)模型后五个点组成的曲线从观察上完全重合了。

但是因为本实验只使用了一组原始序列数据, 无法确定这个结果是否是普遍规律, 即无论对什么类型的原始序列, 经过足够大的平移变换后, 用 GM(1,1)模型建模, 得到的模型曲线就会与一元线性回归曲线重合? 为此, 我们选用更多的数据, 进行系统的实验来证明这个发现。

### 3.2. 原煤产量实验

为了确定上节构造序列进行的实验是否具有普遍性, 本节我们使用现实世界的真实数据进行探究。在国家统计局网站年度数据中, 选取了 1998 年至 2014 年 17 年的原煤年产量数据, 由此得到由 17 个原始数据组成的原始序列为:

$$x_0 = \{13.32, 13.64, 13.84, 14.72, 15.5, 18.35, 21.23, 23.65, 25.7, 27.6, 29.03, 31.15, 34.28, 37.64, 39.45, 39.74, 38.74\}。$$

上节实验原始序列长度为 6, 事实上, 我们可以在原始序列上建立其他长度的灰色模型, 同样, 我们也关心在不同长度的序列上, 前述发现是否同样存在。在本实验中, 我们将建立长度分别为 5、6、8、10 的序列。由于原始数据数量较少, 为得到更多的数据进行实验, 我们采用重叠分组的方式: 以原始序列长度等于 6 为例, 把第 1 至 6 个数据分为一组, 第 2 至 7 个数据为一组, 以此类推, 直到第 12 至 17 个

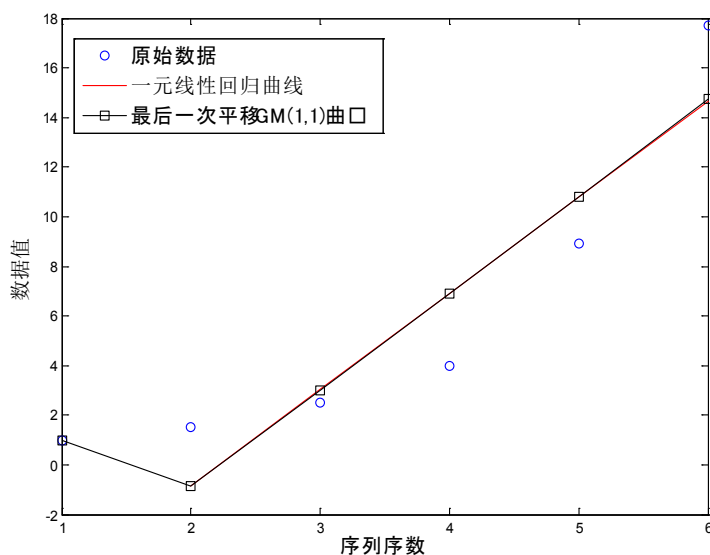


Figure 3. Curve: GM(1,1) model and simple linear regression model for no first point

图 3. 无第一个点时 GM(1,1)模型与一元线性回归模型曲线

数据为一组, 共分成了 12 组, 每组 6 个数据。

验证数据符合光滑性检验后, 按照 3.1 节中的方法, 对每组数据进行平移, 平移量依然为 0, 2, 4, 8, 16, 32, 64, 128, 256, 512。

因为本次实验数据量相对较多, 可以使用配对样本 T 检验和差异度量的方法检验两种模型模拟值的关系, 以此来定量评估。

使用 SPSS 软件对每组数据每次平移所得的 GM(1,1)模型模拟数据与一元线性回归模型模拟数据进行配对样本 T 检验, 4 次分组分别进行配对样本 T 检验所得 P 值如表 1 所示。

由表 1 得知, 10 次平移的 GM(1,1)模型模拟值与一元线性回归模型模拟值进行配对样本 T 检验, 所得的 P 值均大于 0.05, 此时不能拒绝原假设, 即两种模型的数据不存在显著差异, 说明原始序列经过平移变换后建立的 GM(1,1)模型与一元线性回归模型具有相似性。

下面用差异度量方法进行检验。得出的两个模型模拟值差值绝对值如下表 2 所示。

**Table 1.** P value of paired sample T test

**表 1.** 配对样本 T 检验 P 值

平移量	每组数据量			
	5	6	8	10
0	0.246	0.420	0.718	0.942
2	0.251	0.425	0.705	0.996
4	0.255	0.429	0.696	0.964
8	0.263	0.436	0.686	0.911
16	0.273	0.447	0.679	0.858
32	0.287	0.461	0.678	0.822
64	0.300	0.474	0.684	0.807
128	0.310	0.486	0.691	0.805
256	0.318	0.493	0.697	0.808
512	0.322	0.498	0.701	0.811

**Table 2.** The difference measure of 4 subgroups of the output of raw coal

**表 2.** 原煤产量 4 次分组两模型差值绝对值和

平移量	每组数据量			
	5	6	8	10
0	4.503	8.784	19.551	32.021
2	4.154	8.104	18.074	29.789
4	3.857	7.525	16.805	27.849
8	3.380	6.593	14.746	24.635
16	2.716	5.297	11.861	20.004
32	1.957	3.816	8.563	14.531
64	1.260	2.455	5.512	9.388
128	0.738	1.436	3.224	5.496
256	0.404	0.786	1.763	3.005
512	0.212	0.412	0.925	1.576

由上表可以看出, 不论对原始数据如何分组, 所得的实验规律都是相同的, 即随着平移量的增加, GM(1,1)模型和一元线性回归模型模拟值的差值均减小。

根据表 2 的差值绝对值之和, 按平移量作折线图, 如图 4 所示。

折线图直观地显示出了原始数据进行平移变换后构建 GM(1,1)模型的模拟值与一元线性回归模型模拟值差值绝对值之和的变化趋势。曲线呈下降趋势, 说明随平移量的增加, 差值越来越小, 逐渐接近于 0, 即 GM(1,1)模型愈加接近一元线性回归模型。除此之外, 也使用了其他多组真实数据进行了实验, 所得结果与以上实验相同。因此, 我们可以说, 对原始序列进行平移变换, 随平移量的增加, 构建 GM(1,1)模型所得的模拟值, 与在第二个点后构建一元线性回归模型的模拟值会越来越接近, 这一结论对不同类型和不同长度的序列均适用。

#### 4. 总结

本文发现了原始序列经过平移变换后构建的 GM(1,1)模型以及去掉第一个数据后构建的一元线性回归模型, 两种模型所得的结果会随着平移量的增加而逐渐接近乃至重合, 并通过 1998 年至 2014 年 17 年

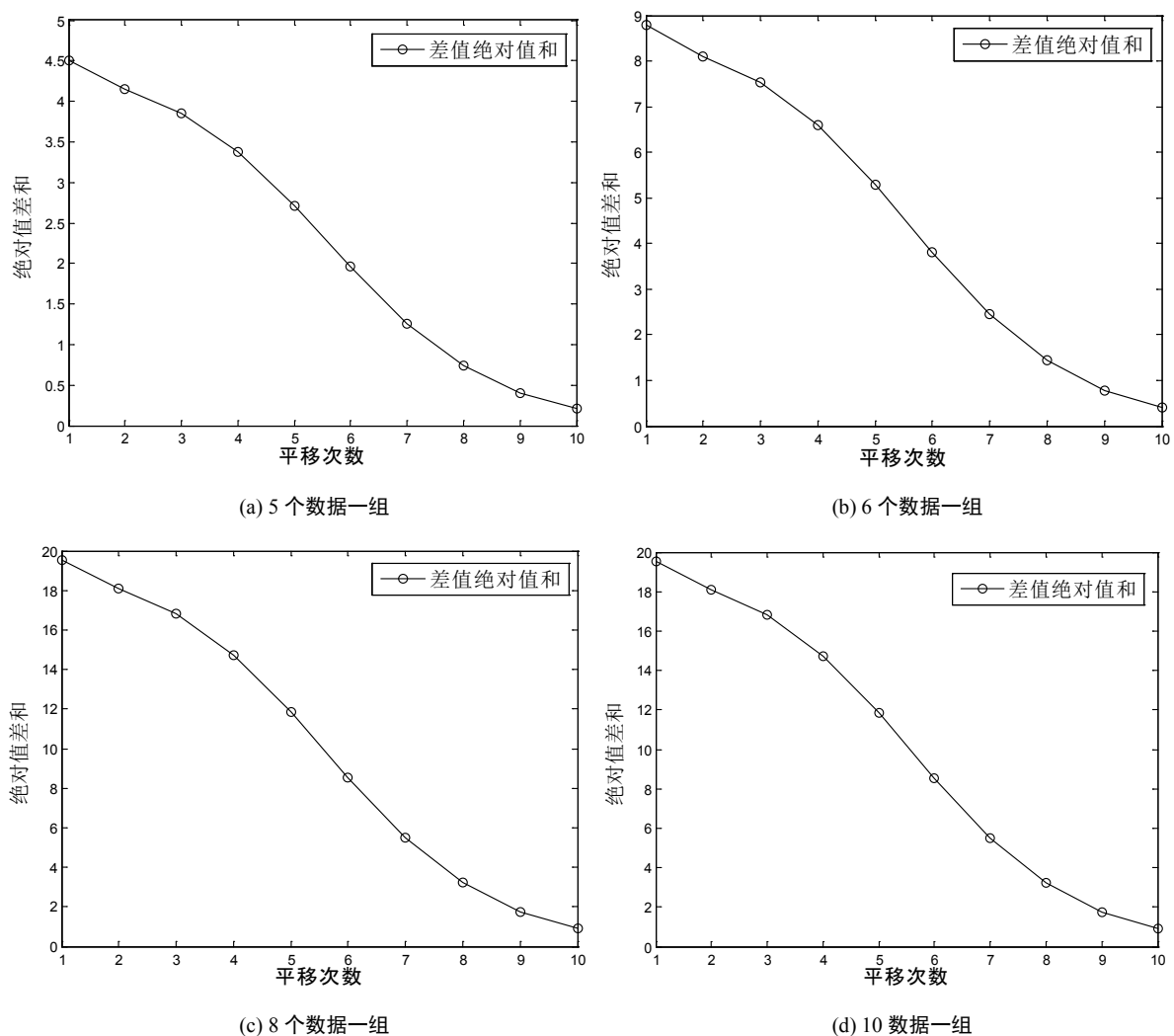


Figure 4. The difference measure of 4 subgroups of the output of raw coal

图 4. 原煤产量 4 次分组两模型差值绝对值和

的原煤年产量数据进行实验证明这一观点。这说明平移变换可以使 GM(1,1)模型转化为一元线性回归模型。因为线性回归模型的简单性,这两种模型间联系的发现有利于帮助进一步认识 GM(1,1)模型的本质,对灰色模型的理论发展和推广传播有积极意义。

## 参考文献

- [1] Liu, S.F. (2007) The Current Developing Status on Grey System Theory. *Journal of Grey System*, **19**, 111-123.
- [2] 肖新平. 灰技术基础及其应用[M]. 北京: 科学出版社, 2005.
- [3] Qian, W.-Y. (2009) A New Type of Data Transformation and Its Application in GM (1, 1) Model. *IEEE International Conference on Grey Systems and Intelligent Services*, **5**, 335-343. <https://doi.org/10.1109/GSIS.2009.5408271>
- [4] Wang, Y., Song, Q., Zeng, B., et al. (2013) Simulation Error Characteristics of Grey Model GM(1,1) under Translation Transformation. *Fifth International Conference on Computational Intelligence, Modelling and Simulation*, Seoul, 137-141.
- [5] 沈继红. 灰色系统理论预测方法研究及其在舰船运动预报中的应用[D]: [博士学位论文]. 哈尔滨: 哈尔滨工程大学, 2002.
- [6] 王子亮. 灰色建模技术理论[D]: [博士学位论文]. 武汉: 华中科技大学, 1998.
- [7] Li, B.J. and He, C.H. (2007) The Combined Forecasting Method of GM(1,1) with Linear Regression and Its Application. *IEEE International Conference on Grey Systems and Intelligent Services*, Nanjing, 394-398.

### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)