

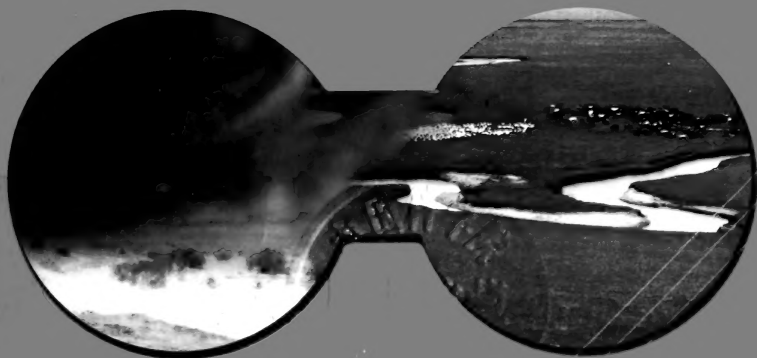
白·图格吉扎布 著

# 趋势分析



其在生态股市中的应用

*Trend Analysis*



 民族出版社  
THE ETHNIC PUBLISHING HOUSE



白·图格吉扎布，草原生态学博士[美]，蒙古族，1945年生于吉林省郭尔罗斯前旗。一九七九年考入内蒙古大学，师从李博教授，硕士论文《锡林河流域羊草草原数量分类的探讨》，提出多元向量夹角余弦值做相似系数。毕业后参加“浅耕翻改良羊草草原课题组”，写出《草场演替趋势的数字预测》，提出状态转移矩阵的通解是对角矩阵。一九八三年主持《草原退化趋势分析及对策研究》，在期中报告中提出草原在退化的论断，并预报了草原地区、北京地区大风天数增加的趋势。一九八五年到美国科罗拉多州立大学草原系，师从数量生态学教授Charles Bonham，一九八九年获博士学位。一九九三年参加美国国防部《土地条件趋势分析》课题，任副研究员。一九九五年发表《卡森堡植被演替趋势分析-超球面模型应用实例》，提出超球面模型。同年，注册版权《股票市场监测器-52》。二〇〇四年注册公司，从事超球面数据分析和投资俱乐部的工作。

捐贈

北京植物所

白. 國格 吉 托 仲

2006.10





58.181  
184

白·图格吉扎布 著

# 趋势分析

# 及

## 其在生态股市中的应用

### *Trend Analysis*

$$3^2 \times 4^2 = 5^2$$



 民族出版社  
THE ETHNIC PUBLISHING HOUSE

中科院植物所图书馆  
  
S0048463

责任编辑：松布尔  
责任校对：乌日娜  
装帧设计：周 方

**图书在版编目 (CIP) 数据**

趋势分析及其在生态股市中的应用/白·图格吉扎布著. —北京：民族出版社，2006.9

(生态丛书)

ISBN 7 - 105 - 07978 - 9

I . 趋... II . 白... III . 数学分析—应用—草原保护—研究  
IV . S912.6

中国版本图书馆 CIP 数据核字 (2006) 第 113154 号

民族出版社出版发行

<http://www.e56.com.cn>

(北京市和平里北街 14 号 邮编 100013)

民族出版社微机照排 北京艺辉印刷有限公司印刷

各地新华书店经销

2006 年 9 月第 1 版 2006 年 9 月北京第 1 次印刷

开本：850 毫米 × 1168 毫米 1/32 印张：9.75 字数：250 千字

印数：0001 - 1000 册 定价：17.80 元

---

该书如有印装质量问题，请与本社发行部联系退换

(总编辑室电话：64212794；发行部电话：64211734)



中国生物数学会  
2000年鞍山会议合影

1999年9月3日  
在美国首都华盛顿参加  
中国欧美同学会  
“二十一世纪中国与  
可持续发展所面临的  
挑战国际研讨会”上  
与“海外生态学者俱  
乐部”部分成员合影  
(左起：伍业钢、白·  
图格吉扎布、林俊达、  
金焰、高放)



2001年“多维球  
面模型在中国股市应  
用研讨会”合影



2000年5月3日与中国科学院应用数学研究所前所长王寿仁先生合影



1995年11月20日在美国丹佛国际机场与中国科学院资深院士阳含熙先生合影



2005年9月11日“草原生态与游牧文明研讨会”合影



## 前 言

感谢您对超球面模型研究和对趋势分析的支持。这是第一本由研发者的科研论文、科研通信、应用实例组成的，从不同视角介绍超球面模型的论文集。超球面模型是一种新的多元分析工具，可以对多元系统进行趋势分析。我们身边的许多事物，都是由多个元素组成的多元系统。在本书中列举的两个比较典型的多元系统，分别是植被和共同基金市场。根据人们对多元系统的观察与记录，把观察值进行分类，研究观察值之间的关系，被称为 Q 分析。一些事物，比如植被，我们也许只能通过样本来认识它。没有样本就没有植被分析。例如，植被分类是研究取自不同地点的植被样方之间的亲疏关系；植被演替分析是研究取自不同年代的植被样方之间的渐变关系。本书中所谓的共同基金市场分析是对共同基金市场进行纵的和横的比较，选出一个组合，以便取得比市场平均更好的投资效果。研发者 20 多年的研究和实践表明，超球面模型是到目前为止，在英文和华文世界上能够对多元系统进行趋势分析的唯一行之有效的多元分析方法。超球面模型最独特的地方是使用向量（又称矢量），而且特别侧重向量的方向和角度，因此可以测出系统组成的变化。

本书可作为高等学校，科研单位生态专业、数学专业、金融和经济专业的研究生、青年教师，多元数据分析师（本书要催生的一个新的职业）的参考读物，也是望子成龙的家长给孩

子的好礼物。

超球面模型的形成过程和重点文章介绍：

1978年，笔者在内蒙古大学读硕士研究生时，导师李博给笔者布置的任务是，在主修英语和数量生态学的同时，寻找一种新的植被分析的数量方法，促进植被科学从定性的科学进入定量的科学。现在笔者可以告慰导师，经过近30年的研究探索，终于发现了一种听似荒诞无稽，实则科学有效的，不仅可以用于植被，而且可以用于股市分析的数学工具：超球面模型，并且定义了堪与道·琼斯指数相匹敌的高高指数。本书主要包括三个内容：生态监测模型的研究，在股票基金市场分析上的应用、检验，杂文（包括电子邮件，讨论，版权证书，及其他文件等）。

从形成时间上来说，超球面模型的发展，大体可以分为4个阶段：

1982—1985年为国内探索阶段。这个时期的许多论文（除硕士论文外）都没正式发表过，但对后来模型的形成有很大影响。其中有我的硕士论文（1982，首次使用向量夹角余弦值作为相似系数来对植被进行分类），兰州会议的发言（1984，提出状态转移矩阵的通解是对角矩阵），《草原退化趋势预测及对策研究》期中报告和草原监测网的设想。这期间比较典型的应用实例是，在1984年指出了中国草原的退化趋势和大风天数增加的趋势。

1985—1997年为模型形成阶段。笔者到美国科罗拉多州立大学草原系进修，师从 Charles Bonham 和 Donald Jameson 教授读博士学位，继续草原监测方面的研究。导师认为对角矩阵的想法很有新意。一般的样方—变量矩阵不满秩，因此无逆，而对角矩阵肯定满秩且有逆。他要笔者深入探索“非对角元素为0”的意义。这期间是笔者学术活动最活跃的阶段，写了许多

超 势分析 及其在生态股市中的应用



随笔、通讯和报告，与同行进行交流和探索。最后决定突破矩阵的框框，直接使用多元向量，并定义了向量除法，以便于处理指数增长的多元系统。这期间比较典型的应用实例有两个：预报了远在意大利的一个科学实验的数据预测数据和实测数据非常接近。模拟投资三个月，虽没有赚钱，但比市场平均好。两个实验结果以通讯的形式，登在美国生态学会统计生态分会的网站上。这期间代表性的文章是发表在欧洲《生态模型 (Ecological Modelling)》杂志 97 期上的，《Multi-Dimensional Sphere Model and Vegetation Instantaneous Trend Analysis》。

1998—2004 年为检验模型阶段。鉴于公共基金市场的数据更新周期快，公信度高，成本低，1998 年开始使用公共基金市场的数据检验模型和完善模型，并进一步用模型来管理个人退休基金。6 年的投资效果优于道·琼斯、纳斯达克和标准普尔。同时也加深了笔者对模型的认识。2001 年在植物生态学报发表《多维球面模型应用于草原监测数据分析的探讨》，最后完成了模型。

2004 年至今为应用阶段。笔者以模型的知识产权入股，朋友们凑了一笔资金，开始组建投资俱乐部和数据分析公司。我们的关于超球面模型的材料引起了联邦证券会的注意。他们约谈笔者，要笔者出示证据，来支持材料所说的“投资六年回报率高于三个指数”的结论。笔者很乐意地赴约，并回答了他们的问询。

虽然转战到股票市场，笔者仍然关心着内蒙古草原。针对人民日报海外版上的相关报道，笔者先后写了几篇关于骆驼、蒙古包的随笔。笔者的老同学，国家民委的铁木尔委员，为了了却笔者的心愿，在民族干部学院召集了一个民间的草原文化研讨会，介绍笔者与民间的草原保护人士见面、交流；并联系了民族出版社，为笔者出版论文集。在这部文集出版之际，谨

向所有真诚帮助的朋友们致以衷心的感谢。

在本书出版的时候，笔者心里最想感谢的有两个人：第一个是笔者的老母亲，内蒙古大学离休老干部赵栩女士。笔者常年在外奔波，即使在母亲 86 岁高龄，仍不得在母亲身边尽孝，使笔者在惭愧的同时，也深感母爱的伟大。第二个是笔者的太太，白·郭·额尔敦图。太太对模型所知甚少，但对模型的研发贡献巨大。如果没有她在餐馆打工的辛劳和精神上的支持，笔者根本无法完成这项历时 20 多年的研究。

这是一本有价值、但读起来不轻松的书。超球面模型仍处于它的初级阶段，还很不成熟。由于涉及几个不同的专业，一个人很难业业精通。在这里说明一下，一方面是为了提请读者注意并谅解，同时也请各方面的专家多多指教，促进超球面模型的发展。为了方便广大读者更好地理解超球面模型，笔者写了《超球面模型基本知识》一文，并把第一章多维空间和多元向量，第二章向量的四则运算加在这本论文集的后面，供大家参考。

趋

势分析 及其在生态股市中的应用





## 序 言

我和白·图格吉扎布博士是小学同班同学，初中在同一学校。我们那个班似乎特别出人才，三十多个同学中，做学问的且有高级职称的就出了十多人，他是其中的佼佼者。他自小就特别好学，爱思考，而且成绩一直是最好的。曾在同年级六个班的竞赛中夺得数学第一，物理第二。曾是那种五分加绵羊式的尖子生。一九六五年，高二期末考试刚结束，他突然报名到牧区插队做牧民，使同学们都很惊讶。同年初我已转学到北京，暑假回呼市还参加了欢送图格吉的聚会。

他在巴盟牧区插队8年。后来，他先后上了工农兵大学，以及读研究生，但我们一直保持着联系。他在做研究生时，接触到了联合国教科文组织的课题，并自那时起把“生态环境监测”也就是“草原退化趋势分析和对策研究”作为了他毕生的研究方向。

有一次，图格吉给我解释他的模型，他说：“我是用多维空间的点来模拟生态环境动态的。因为只有点，才有动态分析；只有多维空间的点才能够携带足够的信息。因此，生态监测的数学工具只能是多元向量”。他的研究得到中国科学院资深院士阳含熙、应用数学

所前副所长王寿仁教授的指点与赞赏。

如他的研究报告所显示的，1985年，在模型研发的早期，他便预报了中国草原退化和北京地区大风日数增加的趋势。我是做民族工作的，对草原退化的严重程度和恶果深有体会，也很担忧。而他现在的一些观点，例如：草原是不依人们主观意志为转移的自然存在；草原是中华民族的生态屏障，人类文明的摇篮；在草原放牧对草的消费要厉行“十一”律，建立畜草平衡；草原之与国土，相当于绿地之与城市；草原退化引起沙尘暴；要发扬游牧文化中的优良传统；民族关系要“和而不同”等；我认为很在意义，值得提出来供大家讨论，并靠实践来检验。

虽然我在北京工作，他在呼市及美国继续读书和开展他的研究工作，但我们时有来往。1995年我去美国出差，在他家（位于科罗拉多州柯林斯堡市）见了面，他还开车带我去南达科他州参观拉什摩尔山四个总结雕像。途经怀俄明州，是一望无际的大草原。图格吉还讲到那里与内蒙古纬度接近，虽是冬末春初，仍能让人感到草原生态比内蒙古好。我们还看到了在美国电影《与狼共舞》中出现过的野牛。这些都给我留下了深刻印象。

近十年来，我们又有多次见面，当然是在北京的次数为多，另有一次是内蒙古自治区成立五十周年，都回到呼市了，并和同学们一起畅叙友情。

去年十月，图格吉从美国回来，我和朋友们为关



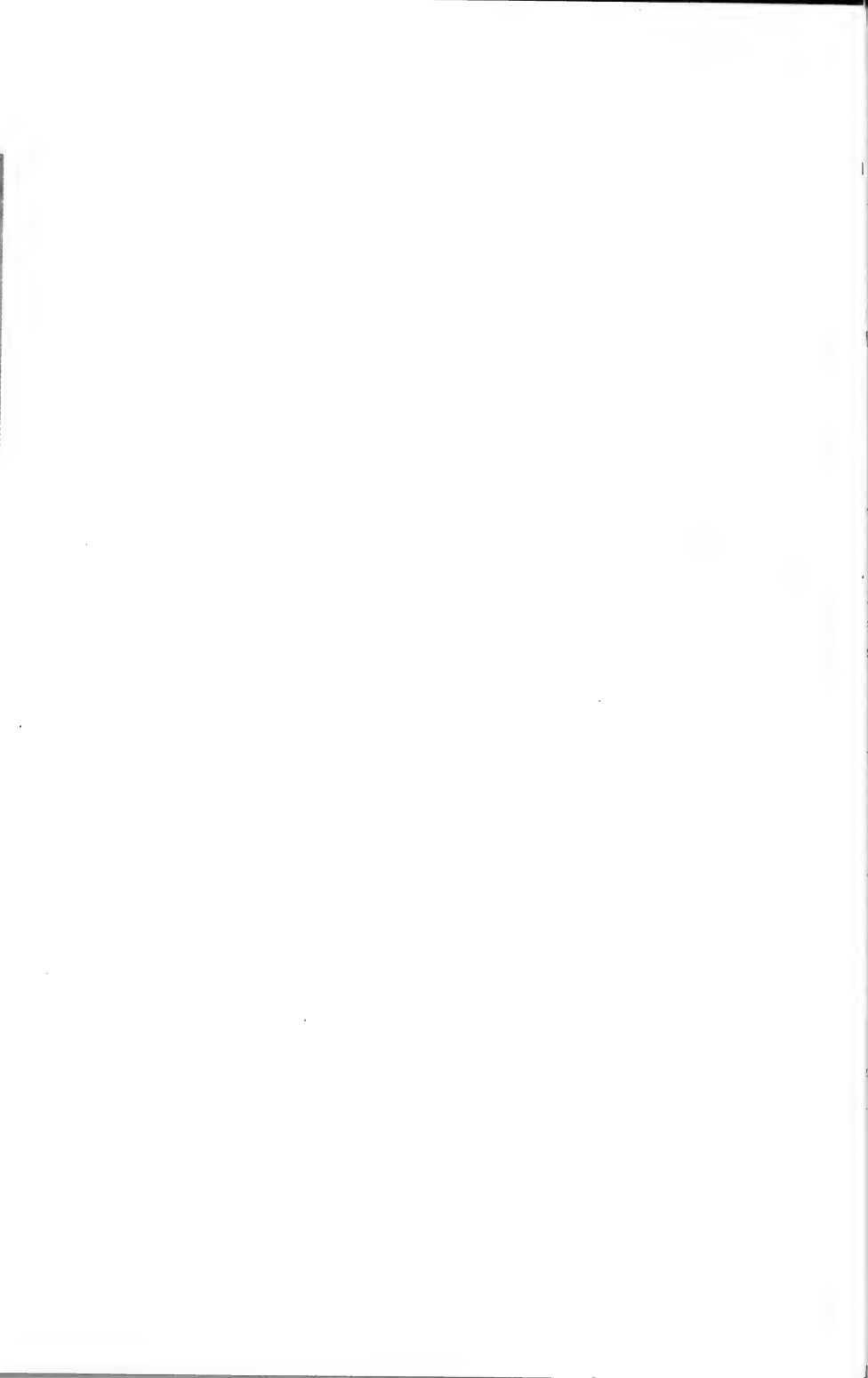
心草原生态，也为让这位老同学一展抱负，开了一个“草原生态与游牧文明研讨会”。大家度过了美好的一天。

图格吉去年就打算出一个集子，把自己写过的论文（包括已发表和未发表的，中文的和英文的）呈现给热衷于草原生态研究的朋友们，也希望更多的学者、官员都更加关心草原的未来，共同保护好中华民族的生态屏障。

我不是什么名人，也不是学者，但作为图格吉的老同学愿意写上这段文字，向大家推介他的这本论文集。

铁木尔

2006年9月





## 目 录

序 言 /1

前 言 /1

- 锡林河流域羊草草原数量分类的探讨 /1
- 草场演替趋势的数字预测 /24
- 关于建立草原监测网的初步设想 /31
- 《草原退化趋势预测及对策研究》期中报告 /38
- 卡森堡植被演替趋势分析——超球面模型应用实例 /46
- 超球面模型与植被演替趋势分析 /56
- 超球面模型应用于股票排序的探讨 /72
- 超球面模型及其在股市分析中的应用 /80
- 多维球面模型应用于草原监测数据分析的探讨 /104
- 关于《超球面模型应用于草原监测数据分析的探讨》  
一文中几个问题的讨论 /117
- 发展骆驼，保护草原，加固中华民族的生态屏障 /123
- 论蒙古包在生态学方面的意义 /127
- 中国沙尘暴起因之我见 /134
- 多维空间与向量分析 /152
- 多元向量基本运算 /171

- What Can We Discover From 1, 2, 3, TO 2, 3, 4? /186
- About "Ocosaic Coefficient" /195
- A Case of Projection /198
- Three Month Summary /217
- Multi-Dimensional Sphere Model and Vegetation Instantaneous  
Trend Analysis /222
- A Critique of Matrix Solutions for Ecology /244
- Testing an Ecological Model using Mutual Fund Data /247
- Vegetation Science and Vector Analysis /248
- Multidimensional Sphere Model and its Application to  
Vegetation Classification /254

词汇、符号 /271

参考文献 /275

## 附 件

1. 谁为他插上飞翔的翅膀 /282
2. 反随机检验  
Anti Random Test /286
3. 科罗拉多州立大学关于知识产权的备忘录  
CSU Memorandum about MDSM Ownership /296
4. 《股票市场监测器》版权证书  
Copyright Certificate of Stock Market Monitor-52 /297
5. 美国证券会来信  
A Litter from Securities and Exchange Commission  
Central Regional Office /299

**趋 势分析** 及其在生态股市中的应用



## 锡林河流域羊草草原 数量分类的探讨

本文是在研究生指导小组诸导师：李博副教授、雍世鹏副教授、刘钟令副教授、孙鸿良副教授、曾泗弟老师指导下完成的。并承中国科学院综考会阳含熙教授、内蒙古大学陈杰教授审阅。

野外工作得到生态——地植物专业八一届毕业生，特别是赵爱芬同学的帮助；数学、计算方面得到邹清莲、谢育先等老师，王厚工程师及郝敦元、阿其拉图同学的帮助。

在此表示感谢。

1981年12月

### 摘 要

以重量为数量指标，依据方差对数据进行简缩，在多维向量空间中，以代表样方的向量的夹角余弦值作为相似系数。通过模糊矩阵的复合，使之满足传递性，成为等价关系，据以完成锡林河流域羊草草原的分类。

在群丛、群丛组和群系一级的分类结果，与传统分类基本相符。亚群系一级的分类则可能要依赖于正确的划分生态种组。

### Abstract

A method is suggested for the rapid classification of ecological surveys by using a calculator.

This method uses cosine angle ( $\alpha$ ) enclosed between position vectors as the coefficient of correlation and operates on composition of fuzzy matrix.

## 前 言

羊草草原 (Form. *Leymus chinensis* (Trin.) Tzvel.) 是欧亚大陆草原区东部特有的一个草原群系, 广泛分布于我国内蒙古高原东部、东北平原、蒙古人民共和国以及(前)苏联的外贝加尔地区。(图 1)

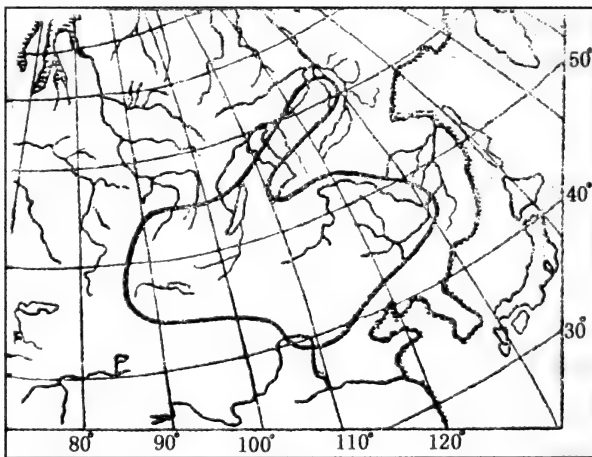


图 1 羊草群系的分布区

(依 Ц. Х. Блюменталь, 1960 年)

在我国, 羊草草原主要分布在草原区的东北部。即东北平





原和内蒙古高原的森林草原地带及其相邻的典型草原地区(图2)。在分布区内,常呈大面积连片分布,为我国草原的优势群系之一。

羊草草原的生态幅度较广,见于多种生境条件,从开阔的草原到低山丘陵;从地带性的黑钙土、栗钙土到隐域性的碱化草甸上,甚至柱状碱土上都有本群系的分布。

建群种羊草(*Leymus chinensis*)为广旱生根茎禾草,其地上枝均匀分布于地表,营养苗高30—60厘米,生殖苗高70—100厘米。它具有强烈的根茎繁殖能力。约在地下10厘米处匍匐生长的根茎,能穿过板结的土壤而迅速占领空间,形成群落。由于密集的羊草常常排斥其他植物的进入,羊草群落与同一地区相近生境的其他群落比较,种类组成常比较简单;但由于羊草草原的生态幅度广,它可与生态特点十分不同的其他植物生长在一起,因此羊草群系的种类组成总计起来是相当丰富的。羊草草原的伴生植物的生态性质常具指示意义,因此常作为羊草草原分类的依据。

羊草草原生产力较高,每公顷产干草2000—3000公斤。质量也好,群落组成中优良牧草比例占60%—80%。羊草为上繁草,适于刈割,是我国草原植被中主要的割草场类型。同时,羊草草原也适于放牧各种家畜,尤适于牧牛。

因为羊草草原面积辽阔,经济价值高,引起不少地植物学工作者的重视。早在50年代末,N.X.布留民塔尔已对本群系的类型做了概括。他把羊草草原划分为5个亚群系,47个群丛。

我国对羊草草原的研究工作也很重视。解放以来,政府有关部门组织的黑龙江流域综合考察(1956—1959),黑龙江省土地资源综合考察(1973—1977),内蒙古宁夏综合考察(1961—1965),内蒙古现代草原试验研究中心地区综合考察队

(1964—1965)，以及有关高等院校、科研所，都对羊草草原做了大量工作。1980年出版的《中国植被》一书，对我国建国30年来的植被研究工作进行了总结，其中对羊草群系也进行了很好的概括。

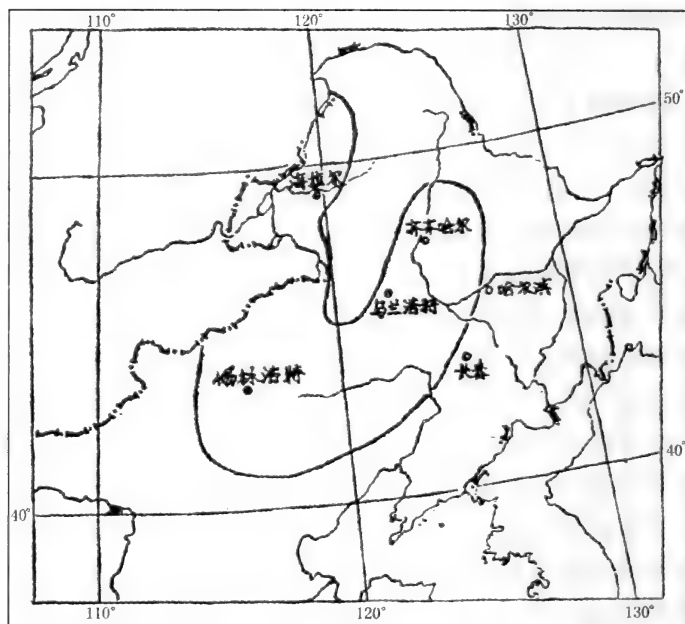


图2 羊草草原在我国境内的主要分布区

(依吴征镒等 1980年)

在此期间，阳含熙教授曾利用主分量分析等方法对呼伦贝尔地区的部分羊草群落进行了数量分类的研究。这大概是我国草原群落数量分类的开端。主分量分析由于既有严格的数学基础，又有较好的应用效果，已受到人们的重视。但因该方法需要矩阵的特征值和特征向量，计算量很大，离开电子计算机就实际上不能工作，使应用受到了一定限制。



阳含熙教授在讨论数量分类的原则、基础和方法时指出：“迄今为止，植物群落分类都采用不重叠的原则。即一个实体只能属于一类，……近年来，很多学者认为，在自然界内，中间类型是很多的，人为地加以割裂会损失许多信息。有些实体只能说在多大程度上属于一类，而又在多大程度上属于另一类，亦此亦彼，界线是模糊的。……数学方面出现模糊数学（Fuzzy mathematics）这一新的领域，就与重叠分类的原则有关。”

在本文中，我们试图引入模糊聚类法对草原群落进行分类。希望在满足传递性的前提下，使数量分类方法有所简化，既适于电子计算机计算，也适于野外条件下利用小型计算器进行工作。本文的目的：

1. 探讨草原群落分类中引入模糊聚类法的可能性。
2. 摸索模糊聚类法应用于草原群落分类的条件。
3. 把模糊聚类的结果和传统分类的结果相比较，说明模糊聚类的分类效果。

考虑到羊草草原研究资料较多，群落类型也比较复杂，我们选定锡林河流域的羊草草原作为研究对象，以便于和传统分类结果进行比较。

## 锡林河流域羊草草原概况

锡林河流域位于内蒙古高原的东南部，主要分布在锡林郭勒盟阿巴哈纳尔旗境内。基本上属典型草原栗钙土地带。往东进入大兴安岭南端山地森林草原及森林地区。地理坐标为北纬 $43^{\circ}26'$ 至 $44^{\circ}08'$ ，东径 $116^{\circ}04'$ 至 $117^{\circ}05'$ 。地势从东南往西北逐渐降低。海拔从1400米至1000米。

羊草草原是锡林河流域的主要草原类型之一，总面积达

1700多平方公里，占本地区面积的三分之一以上。在锡林河流域内，羊草草原可见于不同的地貌单元与不同的生境条件。据统计，锡林河流域羊草草原的种子植物共有184种，分属35科116属。其中以菊科、豆科、禾本科、蔷薇科、百合科的植物种类最为丰富。就上述植物的生态学组成而言，早生草原种占绝对优势，但有一定数量的草甸中生植物（约占总种数的10%）和盐中生植物（占4.4%）。生活型组成及层片结构多种多样，主要层片有：

1. 根茎禾草层片——为本群系的优势层片，主要由羊草组成。

2. 丛生禾草层片——由几种针茅与隐子草、溲草等丛生小禾草组成，为羊草群系的优势层片之一。

3. 多年生杂类草层片——种类组成复杂。按水分生态类型又可进一步分为早生、中旱生、旱中生、中生、耐盐中生等不同类型，形成相应的杂类草层片，组成不同性质的羊草草原。

4. 丛生苔草层片——由日阴菅（*Carex pediformis*）组成，在偏中生生境中起优势作用。

5. 根茎苔草层片——由寸草苔（*Carex duriuscula*）或黄囊苔草（*Carex korshinsky*）组成，为群落下层的优势层片。

6. 小半灌木层片——由冷蒿（*Artemisia frigida*）等组成。分布很广，尤其在过牧地段，常成为优势层片。

7. 草原灌木层片——主要由小叶锦鸡儿（*Caragana microphylla*）组成。一般郁闭度5%—15%，多呈斑点状出现在草原群落中。在过牧丘陵坡地相当密集，而在偏中性生境中很少出现。

本区羊草草原的类型，根据1964—1965年草原研究中心的资料，以及1979—1981年草原生态定位站资料有如下亚群



系、群丛组和群丛：（每一群落类型后面的号码是本文所选用的样方序号）

I. 羊草、中生杂类草亚群系

- |                    |       |
|--------------------|-------|
| 1. 羊草、日阴菅、杂类草群丛组   | 21    |
| 2. 羊草、贝加尔针茅、杂类草群丛组 | 23    |
| 3. 羊草、线叶菊、杂类草群丛组   | 22、36 |
| 4. 羊草、无芒雀麦、杂类草群丛组  |       |
| 5. 羊草、羽茅、杂类草群丛组    | 3     |
| 6. 羊草、早中生杂类草群丛组    | 1、2、4 |

II. 羊草、丛生禾草亚群系

- |                   |           |
|-------------------|-----------|
| 1. 羊草、大针茅群丛组      |           |
| 1) 锦鸡儿—羊草、大针茅、羽茅  | 8、9、11、13 |
| 2) 锦鸡儿—羊草、大针茅、杂类草 | 10、15     |
| 3) 锦鸡儿—羊草、大针茅、小禾草 | 24        |
| 4) 羊草、大针茅         | 12、17、40  |
| 5) 羊草、大针茅、黄囊苔草    | 14        |
| 6) 羊草、大针茅、冰草、苔草   | 25、33     |
| 7) 羊草、大针茅、隐子草、冷蒿  | 31        |
| 2. 羊草、克氏针茅群丛组     | 16、30     |
| 3. 羊草、羊茅、杂类草群丛组   | 5         |
| 4. 羊草、丛生小禾草       |           |
| 1) 羊草、丛生小禾草、冷蒿    | 32        |
| 2) 羊草、小禾草、山葱      | 39        |
| 5. 羊草、黄囊苔草群丛组     |           |
| 1) 羊草、苔草          | 20        |
| 2) 羊草、苔草、杂类草      | 6         |
| 6. 羊草、早生杂类草群丛组    |           |
| 7. 锦鸡儿—羊草，丛生禾草    |           |

8. 羊草、早熟禾杂类草群丛组	18
9. 羊草、隐子草、早熟禾	19
10. 羊草、葱、杂类草群丛组	7
11. 羊草、知母	34、35
Ⅲ. 羊草、盐中生杂类草亚群系	
1. 羊草、芨芨草群丛组	27
2. 羊草、马蔺群丛组	28、29
此外，有二个演替系列类型：	
Ⅳ. 羊草、香茅	26
羊草、冷蒿	

除此之外，尚有二个样方即 37、38 分别属于线叶菊、羊草和线叶菊、羊茅群落。

## 分类方法与分类过程

### (一) 取样与数据处理

1979—1981 年生长季，结合中国科学院内蒙古草原生态系统定位站的工作，先后在锡林河流域进行了羊草草原的调查研究，共取样方 83 个，样线 26 个，合计 109 个。（其中 20% 是老师、同学和其他同志提供的）均采用主观取样。根据植被情况，选择有代表性的典型地段设置样方、样线。其中样方是取 1 平方米面积，逐一登记各种植物名称、盖度、株数，分种剪下称鲜重，烘箱干燥后称干重。样线是用 50 米长 0.4 厘米直径的测绳沿地面拉直，登记基部和测绳相接触的植物，距原点距离，高度、截线长和测绳垂直方向的最大宽度。回室内整理，计算盖度、频率和密度，填统一的样方表格。

在数量分类中，我们采用了优势种原则，并选择干重作为



种的数量指标来确定种的优势度。

分类使用的是 Alpha micro 型电子计算机。由于计算机内存的限制，我们从所做样方中选出 40 个样方，进行数量分类的尝试。

40 个样方中，共有植物 109 种，分属 27 科 78 属（种名见附录 5）。第一步，我们对 109 个种进行了数据简缩。删除、归并了一些重要值（这里是重量）很小的种，仅保留了其中的 40 个种（种组），以与计算机的内存相匹配。种的取舍的原则是看它的方差大小。若某个种的方差较大，则意味着在相应的座标轴方向上，代表样方的实体比较分散，保留这个种，分类效率就较高；若某个种的方差小，无论是因为频度小而引起，或者是数据小，或者是因为趋近于一个平均数，都表明在相应的座标轴方向上，样方分散性小，因而删除、或把它们归并于生态意义相近似的种，对分类效果也影响不大。

对 40 个样方 109 个种的重量值的数据，进行简缩的结果，我们得到了 40 个样方 40 个种的数据，列成表，得到表 1：“TABLEQ $\times$ S”（见附录 1，略去以节省篇幅）。横行是种的数据，纵列是样方的数据。

## （二）分类方法

我们以多维属性空间中边界模糊的实体（向量、多维空间的点）来代表分类对象，以向量的夹角的余弦值作为相似系数，引入模糊聚类的方法进行分类。

考虑到植物群落的分类过程中，存在着大量的模糊现象，模糊概念，为精确数学所不易处理。因为“对自然植被来讲，内部绝对的一致性是不存在的，因为群落内部常常沿某一环境因子的梯度而发生连续的渐变，从而只能划分出相对一致的若干等级。而不同群落之间通常也是沿着许多关系复杂的环境因

子的梯度彼此发生关系，群落间的界线也不是截然分开的，最多只能说是相对间断的，……”作为我们分类的对象来说，一个样方，我们只能说它相对地归于某个群落较好，对于一个命名了的群落，我们也仍不能规定它的数量一定是多少，而允许在某一范围内波动。也即是说，一个样方对于一个群落的隶属程度并不仅是  $\{0, 1\}$  二个值，一个确定的群落，它在数量特征上的界限也不是分明的。况且植物群落之间并没有类似于生物种之间的亲缘关系，而“只是群落在不同级别上的相似而已”。所谓分类，多是“根据植物群落的相似或相异程度进行比较，划分为不同的群落和不同的类别，……”。而据以进行比较的群落之间的相似（或相异）关系也不是一种分明关系，它同样也不能简单地取  $[0, 1]$  二个值，而是取  $[0, 1]$  之间的值，相似关系也是模糊关系。所以，我们想，用模糊集的理论来解释群落学的一些现象和概念，采取模糊数学的方法进行植物群落的分类，也许能取得事半功倍的效果。

方法论确定之后，采取下列具体步骤进行分类。

### 1. 定义相似系数

以多维空间中具有模糊边界的实体代表分类对象  $Q$ 。实体  $Q$  的坐标就是分类对象  $Q$  中各组成成分的重量。 $Q \equiv$  样方  $Q \equiv$  实体  $Q \equiv$  向量  $Q \equiv \{\text{样方 } Q \text{ 的植物种}\}$ ；坐标  $\equiv$  分量  $\equiv$  数量指标。

把样方分类，就是按实体在多维空间分布的疏密，通过分布相对稀疏的带，将空间划分为若干个区域。在这些区域内，代表样方的实体分布得相对密集。这样就可以做到使同一区域的实体彼此近——样方尽可能相似；不同区域的实体彼此远——样方尽可能相异。

设  $Q_a$ 、 $Q_b$  是多维空间中代表分类对象的两单位化向量，一般的有：





$$Q_a \times Q_b = 1 \Leftrightarrow Q_a // Q_b;$$

$$Q_a \times Q_b = 0 \Leftrightarrow Q_a \perp Q_b.$$

再考虑到中间状态，我们定义分分类对象  $Q_a$ 、 $Q_b$  间的相似系数  $R_{(a,b)}$ ：

$$R_{(a,b)} = \frac{\sum Q_a \times Q_b}{|Q_a| \times |Q_b|} = \frac{\sum A_{(i)} \times B_{(i)}}{\sqrt{\sum A_{(i)}^2} \times \sqrt{\sum B_{(i)}^2}}$$

这里， $R_{(a,b)}$  可以理解为两向量夹角的余弦：

$R_{(a,b)} = \cos \langle Q_a, Q_b \rangle$ 。它的取值在第一象限里从 1 到 0，两样方愈相似，欧氏距离愈短，夹角则愈小，余弦值也愈大。这是 ORLOCI (1967) 提出的弦距离（标准化距离）

$$D'_{(j,h)} = \sqrt{\sum \left[ \frac{X_{(i,j)}}{V_{(j)}} - \frac{X_{(i,h)}}{V_{(h)}} \right]^2}$$

的修定，也可看作是不平移的相似系数。

## 2. 建立模糊矩阵

相似关系是种模糊关系。相似矩阵  $R$  = 模糊矩阵  $\tilde{M}$ ：

$\tilde{M} = R$ 。模糊矩阵的元素  $\tilde{M}_{(i,j)} = R_{(i,j)}$  计算如下：

### ① 向量单位化

把样方的各组份平方求和取算术根，得到代表样方的向量的长度，或曰模。它表示实体到原点的距离：

$$ABS(Q_a) = \sqrt{\sum A_{(i)}^2}$$

例如：对于第 16 号样方来说，它的向量长度是：

羊草、克氏针茅、隐子草、冰草、冷蒿、黄囊苔草、星毛委陵菜、二裂委陵菜、菊叶委陵菜、麻花头、猪毛菜平方和的算术根是：

$$\sqrt{54^2 + 42.6^2 + 18.7^2 + 15.6^2 + 3.72^2 + 5.3^2 + 0.4^2 + 12 + 0.2^2 + 0.8^2 + 0.3^2} = 73.3$$

把各分量除以向量长度，使绝对数转化为相对数，在保持

每个组份在样方中的组成比例的同时又使各组份的平方和为 1。这样，在二个完全相同的样方相乘时，乘积等于 1。

$$Q_a Q_b = \sum \frac{A_{(i)}}{\sqrt{\sum A_{(i)}^2}} \times \frac{B_{(i)}}{\sqrt{\sum B_{(i)}^2}}$$

当  $A_{(i)} = B_{(i)}$  时，上式变为：

$$\frac{\sum A_{(i)}^2}{\sqrt{\sum A_{(i)}^2} \times \sqrt{\sum A_{(i)}^2}} = \frac{\sum A_{(i)}^2}{\sum A_{(i)}^2} = 1$$

也即下文所说的，满足反身性。

以上，求向量长度和各分量除以向量长度二个步骤合起来，被称做向量单位化，或标准化。它使代表分类对象的实体投影在超球体上，它们到原点的距离都为一个单位，抵消了样方大、小，欠年、丰年的影响，仅保留了组成比例，以便用来进行比较。

### ②求向量内积

我们进一步比较二个样方的共同点。共同点越多，意义越重要，样方就越相似。我们是这样来实现比较的：把二个样方同有种种的值相乘后累加起来，也即向量的内积作为二样方相似程度的度量。例如，第 16 号样方和第 17 号样方的同有种种为：羊草、隐子草、冷蒿、黄囊苔草、二裂委陵菜、菊叶委陵菜、麻花头、猪毛菜，两者的相似系数为：

$$R_{(16,17)} = (54 \times 300 + 18.7 \times 22 + 3.7 \times 2 + 5.3 \times 34 + 1 \times 8 + 0.2 \times 7 + 0.8 \times 15 + 0.3 \times 1) / 73.3 / 315 = 0.7$$

作为相似矩阵第 16 行（列）、第 17 列（行）的元素。……

用同样方法算出每一个样方和其他 39 个样方的相似系数，共  $40 \times 39$  个数据，写入  $40 \times 40$  表的相应位置上，而对角线元素全写 1。这样，得到了模糊矩阵  $\underline{M} = R$ 。 $\underline{M}$  是对称矩阵，例：

$$\underline{M}_{(i,j)} = \underline{M}_{(j,i)} \quad (\text{见附录表 2 "MATR", 略去})$$



### 3. 模糊关系等价化

由相似系数的定义可知，模糊矩阵  $M$  所代表的分类对象之间的模糊关系满足反身性： $\tilde{M}_{(i,i)} = 1$ ；对称性： $\tilde{M}_{(i,j)} = \tilde{M}_{(j,i)}$ ；但不一定能满足传递性： $\tilde{M} \circ \tilde{M} \leq \tilde{M}$ 。一般地说传递性是指：若  $B \sim A$  且  $B \sim C$ ，能否保证  $A \sim C$ ，即能否由  $B$  把相似关系“传递”过去。

一般来说，相似系数不能保证满足传递性。所以，在以相似系数为基础的数量分类中，每产生一个新的并组，新组和其他实体的相似系数都要重新计算，先后共要进行  $(N-1)^2$  次相似系数的计算。且存在新组的数量指标该是多少的问题。实际上，如前所述，在群落分类中，我们所接触的多是模糊现象。分类对象都是互无亲缘关系的模糊集，是多维空间中的边界模糊的实体。当我们在某个水平下归并两个实体，我们就认为这二个实体同属于某个更高水平的模糊并集，这二个实体对并集有同等的代表性，归并不影响二个实体和其他实体的关系。设有二个样方，其中  $A$  有羊草 200 株， $B$  有羊草 201 株，其他成分相同。当在它们的并组中考虑再并入其他样方时，我们不必拿 200.5 和去和别的样方考虑而是拿  $A$ 、 $B$  中的任一个去比较。况且，属于一个群丛的样方可有无穷多个，在实际工作中，我们不可能，也没必要去和所有可能的样方进行比较，一般和一个典型样方比较后，即可决定归属。

在群落分类中，模糊数学允许我们在实践中按相似系数的大小来归并、划分实体。在野外现象，或样方较少时，我们就这样做\*。例如在模糊矩阵  $\tilde{M}$  中， $\tilde{M}_{(19,20)} = 0.973219$  是第 19

\* 可以证明，在我们的问题中，直接按相似系数的大小，归并实体，与通过模糊矩阵复合划分实体，结果是一样的。

行(列),同时也是第20列(行)中最大的(除对角线元素)元素。我们就可以把19号和20号样方并为一类。……同样,我们还可以把样方34、35并为一类,……。这时,我们是把实体视为模糊集(分明集的特例)把实体之间的相似系数视为模糊集之间的隶属函数。这里需要指出,一般来说,相似系数并不等于隶属函数,但在群落分类这个具体问题上,我们是可以把相似系数作为隶属函数的。从而,有隶属原则保证我们可以直接归并它们。这就给我们的分类工作带来很大方便,依据样方上的同有种,经过几次加、减、乘、除、开方,即可确定归属。

数学上,模糊关系的等价化是通过模糊矩阵的复合来完成的:

$\tilde{M} \circ \tilde{M} = \tilde{M}^2$ , 其中,“ $\circ$ ”是模糊矩阵的复合运算,定义为:

$$A \circ B = C \Leftrightarrow C_{(k,l)} = \text{Max} (\text{Min} (A_{(k,i)}, B_{(i,l)}));$$

二向量对应分量小者集合中的最大者。也即是把普通矩阵的乘法中的乘运算“ $\times$ ”改为取较小值,加运算“ $+$ ”改为取最大值。

例如,40维模糊矩阵  $\tilde{M}_{(i,j)} = R_{(i,j)}$  (见表2)的第26行、第27行分别是:

$$\begin{aligned} \tilde{M}_{(26)} = & (0.6, 0.8, 0.9, 0.8, 0.5, 0.9, 0.9, 0.7, 0.7, \\ & 0.7, 0.8, 0.8, 0.9, 0.8, 0.9, 0.7, 0.9, 0.8, \\ & 0.8, 0.9, 0.7, 0.7, 0.9, 0.9, 0.9, 1.0, 0.8, \\ & 0.9, 0.9, 0.7, 0.8, 0.8, 0.7, 0.9, 0.9, 0.8, \\ & 0.3, 0.1, 0.7, 0.9) \end{aligned}$$

$$\begin{aligned} \tilde{M}_{(27)} = & (0.6, 0.7, 0.9, 0.7, 0.5, 0.8, 0.8, 0.7, 0.7, \\ & 0.6, 0.7, 0.7, 0.8, 0.7, 0.8, 0.7, 0.8, 0.7, \\ & 0.7, 0.8, 0.7, 0.6, 0.8, 0.8, 0.8, 0.8, 1.0, \\ & 0.8, 0.8, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 0.7, \end{aligned}$$



0.3, 0.1, 0.7, 0.8)

则  $\text{Min}(\underline{M}_{26}, \underline{M}_{27}) = (0.6, 0.7, 0.9, 0.7, 0.5, 0.8,$   
 $0.8, 0.7, 0.7, 0.6, 0.7, 0.7,$   
 $0.8, 0.7, 0.8, 0.7, 0.8, 0.7,$   
 $0.7, 0.8, 0.7, 0.6, 0.8, 0.8,$   
 $0.8, 0.8, 0.8, 0.8, 0.8, 0.6,$   
 $0.7, 0.7, 0.7, 0.8, 0.8, 0.7,$   
 $0.3, 0.1, 0.7, 0.8)$

其中最大值  $\text{Max}(\text{Min}(\underline{M}_{(26,i)}, \underline{M}_{(27,i)})) = 0.9$

所以复合后的矩阵  $\underline{M}$  的第 26 行 (列), 第 27 列 (行) 的元素是 0.9,

$$\underline{M}_{(26,27)}^2 = \underline{M}_{(27,26)}^2 = 0.9$$

重复同样的过程, 我们可以得到全部的  $\underline{M}^2$  的元素  $\underline{M}_{(i,j)}^2$  进而可以得到  $\underline{M}^4, \underline{M}^8 \dots$ 。

有一定理保证, 模糊矩阵复合有限次后可满足传递性:  $(\underline{M}^k)^2 = \underline{M}^k$ 。其中  $2K < N$ ,  $N$  是矩阵的维数。

在本题中, 模糊矩阵复合 4 次后,  $\underline{M}^{16} = \underline{M}^{16} = \underline{M}^*$ ,  $\underline{M}^*$  是模糊等价关系。(见附录 3 表 “MATM\*”, 略去)

#### 4. 利用模糊等价关系完成分类

模糊矩阵  $\underline{M}^*$  对应 40 个样方之间的模糊等价关系。

$\underline{M}^* = \underline{M}_{(i,j)}^*$  中的元素都是  $[0, 1]$  上的数, 表明分类对象之间的不同程度的相似; 而且, 这种相似关系已满足反身性、对称性和传递性。

当我们任意确定一个置信水平, 选一个实数  $\lambda$  ( $0 < \lambda \leq 1$ ), 把大于等于  $\lambda$  的  $\underline{M}_{(i,j)}^*$  作为 1, 小于  $\lambda$  的  $\underline{M}_{(i,j)}^*$  作为 0:

$$\tilde{M}_{(i,j)}^* = \begin{cases} 0 & \tilde{M}_{(i,j)}^* < \lambda \\ 1 & \tilde{M}_{(i,j)}^* \geq \lambda \end{cases}$$

则得到  $\tilde{M}^*$  的  $\lambda$  截集, 代表 40 个实体在  $\lambda$  置信水平下的一个分明等价关系。据此分明等价关系, 可将对象按 0、1 分类:  $\tilde{M}_{(i,j)}^* = 1$ , 则  $Q_i Q_j$  在  $\lambda$  水平下可归为一类;  $\tilde{M}_{(i,j)}^* = 0$ , 则  $Q_i Q_j$  在  $\lambda$  水平下不能归为一类。

例如, 当我们令  $\lambda_1 = 1$  时,  $\tilde{M}^*$  的对角线元素全等于 1, 其余元素全小于 1,

$$\tilde{M}_{(i,j)}^* \begin{cases} < 1 & (i < j) \\ = 1 & (i = j) \end{cases}$$

所以  $\tilde{M}^*$  的 1 截集  $\tilde{M}_1$  是单位矩阵 E, 对角线元素全为 1, 其余元素全为 0, 表明在  $\lambda = 1$  的置信水平下, 所有样方各为一类。

当我们取  $\lambda_2 = 0.975943$ , 除对角线元素外, 还有第 28 行 (列) 第 29 列 (行) 元素  $\tilde{M}_{(28,29)}^* = \tilde{M}_{(29,28)}^* \geq 0.975943$ 。所以  $\tilde{M}^*$  的 0.975943 截集  $\tilde{M}_{0.975943}$  除对角线元素外, 尚有  $\tilde{M}_{0.975943(28,29)} = \tilde{M}_{0.975943(29,28)} = 1$ 。表明在  $\lambda = 0.975943$  置信水平下, 第 28 号和第 29 号样方可以归并为一类, 而其他样方各自为一类。

当我们取  $\lambda_3 = 0.975622$  时, ……。

这样, 当  $\lambda$  的取值从  $\text{Max}(\tilde{M}_{(i,j)}^*) = 1$  逐级下降,  $\lambda_1 = 1$ ,  $\lambda_2 = 0.975943$ ,  $\lambda_3 = \dots\dots$  直到  $\lambda_{40} = 0.587209$ , 全部样方归为一类为止, 得到了 40 个不同水平下的分类。

**趋势分析** 及其在生态股市中的应用



## 结果与讨论

按 40 个不同水平下的分类，将其以树枝状聚类图表示出来归并的动态如图 3。

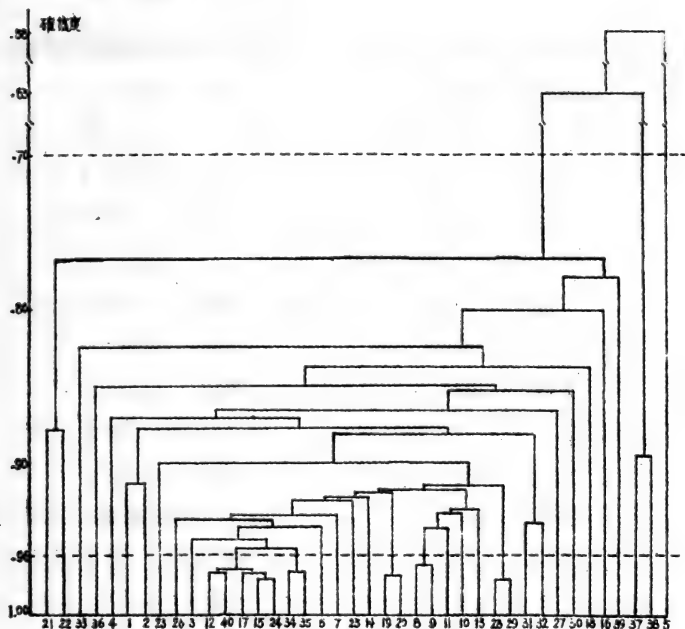


图 3 锡林河流域羊草草原聚类图 (1981)

图中的纵坐标是  $\lambda$  值 (确信度)，横坐标是样方序号。

根据聚类图和我们分类的需要，我们选二个确信度： $\lambda_1 = 0.70$ ， $\lambda_2 = 0.96$ ，得到了二个不同水平下的分类。

1. 在  $\lambda_1 = 0.70$ ， $\arccos 0.7 = 45^\circ 34'$  的水平下，40 个样方分为三组： $\{5\}$ ， $\{37, 38\}$ ， $\{\text{其他}\}$ 。我们可以把它们视为三个群系。 $\{37, 38\}$  是线叶菊草原， $\{\text{其他}\}$  是羊草草原 (样方 5

留待下文讨论)。而  $\arccos 0.7 = 45^{\circ}34'$  很接近于  $45^{\circ}$ ，这仅是一种巧合，还是二者之间有某种必然的联系；今后，我们是否可以以 0.7 做为群落异质的一个界限尚有待于进一步证明。

2. 在  $\lambda = 0.96$ ,  $\arccos 0.96 = 16^{\circ}15'$  的水平下，羊草草原的 37 个样方，聚为 29 个组。看来它们相当于群丛一级的分类单位。

把 29 个组进行整理、命名，列出如下：羊草草原群系

### I. 羊草、杂类草亚群系

#### 一、羊草、日阴菅群丛组

1. 羊草、日阴菅 21

2. 羊草、日阴菅、线叶菊 22

#### 二、羊草、无芒雀麦群丛组

1. 羊草、无芒雀麦、杂类草 33

#### 三、羊草、线叶菊群丛组

1. 羊草、线叶菊、黄芩 36

#### 四、羊草、杂类草群丛组

1. 羊草、荆芥、柳穿鱼 4

2. 羊草、麻花头 1、2

3. 羊草、地榆 23

演替系列：羊草、香茅 26

### II. 羊草、丛生禾草亚群系

#### 一、羊草、羽茅群丛组

1. 羊草、羽茅、荆芥 3

#### 二、羊草、大针茅群丛组

1. 羊草、大针茅 12、40、17、15、24

2. 羊草、大针茅、杂类草 14

#### 三、羊草、知母群丛组

1. 羊草、知母 34、35

趋势

势分析 及其在生态股市中的应用





四、羊草、苔草群丛组	
1. 羊草、苔草、杂类草	6
2. 羊草、苔草	19、20
五、羊草、杂类草群丛组	
1. 羊草、扁蓄豆	7
2. 羊草、星毛委陵菜	18
六、羊草、冰草群丛组	
1. 羊草、冰草	25
七、锦鸡儿——羊草、针茅群丛组	
1. 锦鸡儿——羊草、针茅、(羽茅)	8、9、10、11、13
八、羊草、克氏针茅群丛组	
1. 羊草、克氏针茅、隐子草	16
九、羊草、冷蒿群丛组	
1. 羊草、冷蒿	31、32
2. 羊草、冷蒿、克氏针茅	30
十、羊草、葱群丛组	
1. 羊草、葱、苔草	39
Ⅲ. 羊草、盐湿杂类草亚群系	
一、羊草、马蔺群丛组	
1. 羊草、马蔺	28、29
二、羊草、芨芨草群丛组	
1. 羊草、芨芨草	27

上述数量分类的结果和传统分类的结果相比较，我们可以看到，两者互相吻合得是比较好的，尤其是在群系和群丛一级。而且，模糊聚类比较准确、严格。这里，“准确”是指能始终坚持一定的分类原则，不受主观经验的影响。例如，传统分类，把 22 号和 36 号样方归并，使之属于羊草，线叶菊群落。而按照优势种原则，计算机把 22 号样方和 21 号样方归

并,使之属于羊草、日阴营群落。从样方数据看出(见附表1),样方22号含有线叶菊,但它的日阴营的分量更大于线叶菊。虽然,在我们看来也许线叶菊具有指示意义,但计算机并没有特别看重它。在本题中,计算机只承认重量。若欲贯彻特征种原则,需特别加强某个特征种的生态指示意义,我们要预先给以加权,给它乘以一个大于1的系数。而哪个种需要加权,加多大的权,则有赖于我们对种类成分的生态性质的深入了解。在不清楚时,我们就把它们的权都作为1。

由于模糊聚类准确、客观,又比较简便,灵活、因此可以用来在野外工作,检验现场所取样本。虽说,在理论上,最好的办法,甚或是唯一被承认的办法是随机取样,但在实践中,由于劳力、时间的限制,尤其在路线考察中,往往不得不采取主观取样。而在主观取样中,由于受经验、素养、精力等影响,我们并不能保证会经常地洞察植被的细微变化,所以在取样后,通过计算进行监测是十分必要的。

一些貌似均匀、同质的群落,往往隐含着错综复杂的分布格局,会严重地影响取样结果。比如样方5,从生境看,从与之平行做的样线看,确是取自羊草、羊茅、杂类草群落,传统分类也将它归入羊草,羊茅群丛组。但模糊聚类的结果显示它和所有的羊草草原群落的相似性都很差,从夹角说已超过了 $45^\circ$ 。检查它的数据,发现它的山萝卜的分量已超过了羊草。看来,样方5是不巧取在了山萝卜的斑块上。因此用模糊聚类法,在现场检验样本,会使取样工作更可靠,效率更高。

模糊聚类对于归并的实体,指出归并的确信度,对于分立的实体,指出它们的相似程度。这种性质对于我们处理多个样方时是很有用的。因为分类方案随样方的增加而急速增加。当样方多时,比较它们的细微差别、归并后的动态,即使很有经验的人,也很难保证不发生判断错误。而有相似系数做指标,

趋

## 趋势分析 及其在生态股市中的应用



分类就直观、明晰得多了。

同时，我们也可以看出，群丛组一级的分类与传统分类吻合得不够好。并且，单依赖目前的聚类图想进一步归并出亚群系是困难的。例如，根据经验，我们可以将羊草、芨芨草群丛组和羊草、马蔺群丛组进一步归并为盐湿草原（亚群系）。但从图上，我们却看不出它们有任何归并的趋向。如前所述，我们是按实体在多维空间中的分布来划分它们。数学上要求各坐标轴之间要互相垂直，也即是要求作为坐标轴的种类之间，互相独立，没有相关。既然作为不同的坐标轴，它就不承认芨芨草和马蔺之间有任何相关性，而认它们为截然不同的种。但是，我们知道群落中的种之间往往存在着某种相关关系。因此，有的学者认为，坐标轴之间应该是不垂直的。但如果不垂直，那么夹角应该是多少？这将是一个十分复杂的问题。我们认为这种相关关系也是一种模糊关系，在一定的阈值下可认它为分明关系。我们曾对 1981 年的 40 个样方按水分生态类型划分了生态种组，同一种组中的植物被认为是完全等同的。如糙苏、地榆属于一个生态种组，芨芨草和马蔺属于一个种组，群落中出现糙苏或出现地榆被认为是一样的；出现芨芨草，或出现马蔺也被认为是一样的。一个样方里的马蔺可以和另一个样方里的芨芨草相乘，增加了这些样方的相似性，使分类在亚群系一级也取得了比较满意的结果。

所以建议，在大规模的分类时，是否可分两步走，先按种，划分出群丛和群系。需要进一步归并亚群系和群丛组时，按层片的概念归并出相应的生态种组。同一生态种组，表现出种类成分对某一生态因子反应的一致性，反应了种类成分之间的相关。然后，以生态种组为坐标轴，划分出群丛组和亚群系。

从聚类图上，还可以得到有关植被的一些信息。比如，似

乎有一种倾向，群落的羊草比重<sup>①</sup>越大，则确信度越高。按模糊集理论，我们把羊草草原视为一个模糊集。在多维空间中，它是一个具有模糊边界的实体，我们可简单地把它视为超球体，则它的核是纯的、单一的羊草群落。而呈同心球向外，羊草比重逐渐降低。羊草草原的分类和羊草的比重很有关系。这一点似和其他的方法，例如关联分析所依据的理论有所不同。（在关联分析中，建群种是有指示意义的。）而和生态学中的分布中心说、生态位学说有某种相似。如果这种假设成立，那么我们要寻找稳定的、羊草比例高的割草场类型，人工草地较好的生态组合，单从生态的角度，似乎应该从羊草、如母，羊草、扁蓄豆，羊草、羽茅，羊草、丛生禾草这几个组合中去寻找。

就这 40 个样方而言，羊草、大针茅群丛组处于距核最近的位置。这可能说明：①40 个样方取自羊草草原和针茅草原的过渡带上。②在样区范围内，羊草和针茅相关性好，两者的生态性质相近，或互补。这有待于进一步的证明。在模糊聚类的基础上，证明②并不复杂，把输入的“TABLE Q \* S”矩阵转置，改对样方聚类为对种（种组）聚类即可求出种对之间的相关来。

关于数量指标，本文中我们采用重量作数量指标，因为重量比较客观且单一，而且从应用效果看，用重量来描述种的优势度还是比较好的。它的缺点是取样比较费事，而且对植被有破坏性。

我们也曾采用过盖度级，对群落组成成分五个等级，按种评分，作为数量指标，进行分类，也取得了比较好的结果。其优点是迅速、简便，其缺点是主观性较大，不易和其他数量指标

① 这里的比重理解为组份在向量长度中的比例。



进行比较和转换。所以，选取怎样的数量指标才能更好地体现优势种原则，充分地发挥模糊聚类的功能，是个尚待进一步探讨的问题。

模糊聚类与其他的以多维空间为模型的分法，在分类效果、理论上的异同还有待于在实践中予以检验和证明。

## 结 论

(一) 群落中存在着大量的模糊现象，如群落的相似，种的相关等。模糊数学在群落学中的应用中有着广泛的前景。它将成为我们认识、揭露植被中模糊现象的有力武器。

(二) 本文引入模糊数学的一个方法应用于草原群落的分类。它比较简单、灵活、准确。

(三) 在运用模糊聚类法进行草原群落分类时，需注意：正确选择数量指标，给每个坐标轴以一定的生态意义且互不相关；选择一个好的相似系数——隶属函数。

(四) 以向量的夹角余弦值作为两样方相似程度的度量是可行的。它具有反身性和对称性，在作聚类图时，为了在很高的相似程度下进一步细分样方，可以进行反余弦变换。

(本文为硕士学位论文，发表于1982年《内蒙古大学硕士学位论文研究生论文集》)

## 草场演替趋势的数字预测

本文在写作过程中得到数学硕士郝敦元和阿其拉图的帮助。初稿经李博先生审阅，并提出宝贵意见。在此一并致谢。

### 前 言

草场演替趋势的预测研究，不仅具有一定的理论意义，更具有重大的实践意义。草场是畜牧业的物质基础，草原植被生产力的高低、升降，关系到草原畜牧业的兴衰。同时草原植被又是人类生存环境的重要生态屏障。现在，关于草场沙化、盐碱化、退化为裸地的报告越来越多。关于水库淤满、风沙增大的报告也每见于各种文字。但是，草场退化不可能是突然发生的。它必然有个过程。从顶极群落到裸地之间有成分变化、生物量变化的渐变过程。我们希望能及时地发现草场的细微变化，同时从细微的变化中察知草场的变化趋势，以便据以制定合适的利用方式和利用强度。这可使我们的畜牧业生产管理，草原利用建立在更加科学稳妥的基础上，减少盲目性。同时，我们也希望这种方法简便易行，同时又有可靠的数学基础，能够定性定量地描述，能够互相比较。

草场演替一般是一个漫长的过程。造成演替的原因也很复杂多样，就整个演替过程来说，它不可能，也不应该是线性过



程，但就某个局部而言，就某一小段时间来说，它可以是，可以被看作是线性的，可以用线性代数的方法来处理、研究。

## 方 法

如果我们把草群产量  $S$  作为纵坐标，以时间  $T$  为横坐标，在时间  $T_0$ ，初始产量为  $S_0$ 。过一段时间后，在  $T_1$  时间，产量为  $S_1$ ，则可以得到点  $A_{(t_0, s_0)}$  和点  $B_{(t_1, s_1)}$ 。因为  $T_1 - T_0$  总是正值，一般地有，若  $S_1 - S_0 > 0$ ，我们认为草场产量在上升，草场处于上行演替。反之，若  $S_1 - S_0 < 0$ ，认为草原产草量在下降，草场处于下行演替。一般来说，在植被演替的多维空间里，在无限的演替过程中，截取其中的一段  $A_{(t_0, s_0)} B_{(t_1, s_1)}$ ，若  $AB$  的斜率大于零，则上行演替，若斜率小于零，则下行演替。

随着生产的发展，人们认识到草原的退化与恢复不仅是个产量的问题，而更主要的是质量的问题，即草原群落的组成问题。在一块草场上，有几十种植物，除了总产量有升降外，每种植物的产量，分种产量，也都有升降，有的升高，有的降低。草原退化恢复的概念不仅表现在草群总产量上，而且还表现在质量上，表现在分种产量上。比如一块羊草地实施浅耕翻后，第一年长满了猪毛菜和蒿子，产量非常之高。人们不会认为草地是在上行演替。第二年，虽然羊草的比重有提高，但总产量却有所下降。对此，人们都会认为草原植被正在恢复，而不是在退化，即使总产量不如头一年。因此草原产草量  $S$  在这里不是一个数，而应该是数组。其中各分量表示草群的各个组份。如第一个分量，羊草；第二个分量，猪毛菜……对于数组  $S$ ，我们不再是简单地求和， $\sum (S_{1,i} - S_{0,i})$ ，来说明演替趋势，而是对分量一个一个地分别研究。比如，先研究第一个分

量，看羊草是怎样变化的；再研究第二个分量，看猪毛菜是怎样变化的；然后再研究第三个、第四个。研究每个分量的变化趋势是怎样的，是变大还是变小，变化的速率有多大？待把各个分量分析透了，再综合起来，就得到整个植被变化的概貌了。

设在  $T_0$  时，草原生产力是向量  $A$ ，通过时间  $\Delta t$  后，在未知综合因素  $X$  的作用下，草原生产力在  $T_1$  时成为向量  $B$ 。

用代数表示： $AX = B$ 。这里  $X$  是未知综合因素，以  $n \times n$  矩阵表示。我们若能解出  $X$ ，就可以进一步求  $C = AX^k$ 。向量  $C$  为在  $X$  因素作用下  $k$  个时间段后的草场状况。此时，即使  $A, B$  的变化很微小，当  $k$  足够大时，也应该能够看出明显的变化趋势。

现在，问题的关键是求出转移矩阵  $X$ 。在  $A, B$  都为向量的情况下，问题转为解方程组：

$$AX = B,$$

其中  $A, B$  是已知向量； $X$  是  $n \times n$  未知矩阵。由于未知数多于线性无关的方程个数，所以解不确定。有多组解，表明在群落演替的多维空间中，从  $A$  运动到  $B$  有多条途径。

为使解确定，我们把向量与矩阵的乘法引向矩阵与矩阵的乘法。对于非奇异方阵  $A$  方，由  $A$  方  $X = B$  方，可推出  $X = A$  逆  $\times B$  方。因此，我们用方阵  $A$  来代替向量  $A$  表示草场状况  $S$ 。向量  $A, B$  乘以单位矩阵  $E$ ，得到方阵  $A$  方， $B$  方。方阵  $A$  的对角线元素  $A_{(i,i)}$  等于向量  $A$  的对应分量  $A_{(i)}$ ； $A_{(i,i)} = A_{(i)}$ 。方阵  $B$  的对角元素  $B_{(i,i)}$  等于向量  $B$  的对应分量  $B_{(i)}$ ； $B_{(i,i)} = B_{(i)}$ 。同时，非对角元素全为零：

$$A_{(i,j)} = B_{(i,j)} = 0, (i \neq j)。$$

观察得到， $A$  逆 =  $(0, 1/A_{(i,i)}, 0)$ 。

趋势

势分析 及其在生态股市中的应用





因此, 转移矩阵  $X = A^{-1} \times B = (0, B_{(i,i)}/A_{(i,i)}, 0)$   
 $= (0, B_{(i)}/A_{(i)}, 0)$ 。

进一步,  $C = B \times X^k = B \times (B_{(i)}/A_{(i)})^k$ 。

比较 A, B, C, 和 k, 即可对草场演替的趋势进行预测。

一个例子

将上述方法应用于羊草草原浅耕翻的研究结果报告如下:

样方取自草原研究所改良课题组的资料。

测产, 聂素梅等取样整理。原始资料存草原研究所科研处。数据处理可参照聂文, 1984 年。

优势度 = (高度 + 密度 + 盖度 + 频度 + 重量) / 5。

分种计算优势度得到植物优势度如下:

年度 \ 植物种	羊草	苔草	猪毛菜	隐子草	冷蒿	黎
1982	81.6	17.3	66.8	7.6	39.4	61.8
1983	96.1	28.7	28.7	15.8	30.5	6.0

即  $A_{82} = (81.6, 17.3, 66.8, 7.6, 39.4, 61.8)$

$B_{83} = (96.1, 28.7, 28.7, 15.8, 30.5, 6.0)$

分别乘以单位矩阵 E, 得:

$$A \text{ 方} = \begin{bmatrix} 81.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 17.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 66.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 39.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 61.8 \end{bmatrix}$$

$$B \text{ 方} = \begin{bmatrix} 96.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 28.7 & 0 & 0 & 0 & 0 \\ 0 & 0 & 28.7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 30.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6.0 \end{bmatrix}$$

草场 1981 年浅耕翻后即封闭, 没有放牧, 由于当地的水热土壤条件和植物群落内部的矛盾斗争, 群落由 1981 年的裸地演变为 1982 年的猪毛菜 - 黎 - 羊草为主的组合, 继而演变为 1983 年的羊草 - 冷蒿群落。假设各种条件不变, 按照目前的这种变化势头, 多少年后, 会变成什么群落呢?

根据前面的讨论, 以 1982 年为起点 A, 1983 年为 B。由  $A \times X = B$ , 得  $X = A^{-1} \times B$ 。

$$A^{-1} = \begin{bmatrix} 1/81.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/17.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/66.8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/7.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/39.4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/61.8 \end{bmatrix}$$

$$X = A^{-1} \times B = \begin{bmatrix} 1.18 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.66 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.43 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.08 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.77 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.097 \end{bmatrix}$$

$$\text{当 } k=3 \text{ 时, } C = B \times X^3 = B \times (B/A)^3 = B^4/A^3$$

**趋** 势分析 及其在生态股市中的应用



向量  $C = C_{(i,i)} = (134, 47.5, 68, 18, 0.06)$ 。

也即，若保持目前的势头，到 1985 年，藜将完全消退，猪毛菜将仍有一小部分。群落将以羊草，134.4，隐子草，68，苔草，47 为主，同时冷蒿也占一定比例。根据过去对实验站植被所作的本底调查，这个结果是可信的。说明浅耕翻 + 封闭所造成的羊草草原的植被演替是上行演替，是使植被恢复的演替。转移矩阵作用二次，也即四年后植被即可恢复。相比于植被演替的漫长地质过程，速度也是相当快的。

## 结果与讨论

从我们所作的例子看来，以去年、今年两年的草场状况通过除法求出转移矩阵，从而监测草场的演替趋势，并预测若干年后的草场状况的方法是可行的。以草场连续两年的对应分量的比值作为转移矩阵的对应分量的元素的解，比较简单、直观，其实质是在多维空间里草场演替的轨迹中取相邻二点 A、B，连接 AB，延长 k 倍到 C。当作用于植被的诸因子保持不变，植被变化较平稳的情况下，把 A、B、C 和 k 加以比较，可以预测草场未来的变化。

对于转移矩阵 X 的各分量： $X_{(ii)} = B_{(i)}/A_{(i)}$ ，若  $X_{(i,i)} > 1$ ，则此分量随时间而增长，若  $X_{(i,i)} < 1$ ，则随时间而衰减。 $X_{(i,i)}$  的值偏离 1 越多，变化越剧烈。

若对应分量不相等： $A_{(i)} \neq B_{(i)}$ ，当 k 足够大时， $C_{(i)}$  必定收敛于 0 或无穷大： $C_{(i)} \rightarrow 0$ ，或  $\sum C_{(i)} > \text{无穷大}$ 。

在计算 C 时，有一个或几个分量趋近于零，或显著大于其他分量的和时，即可停止计算。因为它表示，演替到了一个种消亡或趋近于一个单纯群落。作为演变的渐变过程，量变过

程已经终结。这时发生了质的改变。比较发生质变时的  $k$  值，可用来说明演替的速度。在没有其他资料的情况下，我们暂时主观地把  $k < 5$  的称为快速演替， $k < 30 - 50$  的称为正常演替。

因为转移矩阵的元素  $X_{(i,i)} = B_{(i)}/A_{(i)}$ ，所以  $A_{(i)} \times B_{(i)}$  不可等于零。若  $A_{(i)} \times B_{(i)} = 0$ ，说明在二次取样的期间，草场成分发生了突变：或者一种成分无中生有，或一种成分突然消失。如果这种成分是有生态指示意义的种或种组，则演替的数字预测就已经没有必要了。因为此时，生态学家应当可以指出草场成分变化的意义。反之，如果单凭这种突变尚不能说明草场演替的趋势，则在草场演替的趋势预测问题上，这种成分所提供的信息量太少，这个坐标轴是多余的，可以去掉。

在运算过程中也是这样。如果某个分量已经趋近于零，但仍不能明确草场演替的趋势，则略去这个分量，继续往下运算，直至显示出明确的生态意义时为止。

若  $A_{(i)} = B_{(i)}$ ，则此处的转移矩阵  $X$  为单位矩阵  $E$ 。或有时虽  $X$  不严格等于  $E$ ，但即使  $k$  较大时（如在草原群落中，我们暂时主观地规定，当  $k > 50$  时）任一个  $C_{(i)}$  都不趋近于零，且又没有哪个次要成分转化而占明显优势，我们就认为群落是稳定的。此时的利用方式，强度是最合适的。

此外，不言而喻，我们以上的讨论都是建立在所取样方能代表植被的基础上。这里完全没有讨论样方的代表性问题。实际上，我们预测的是“样方的演替”。至于“样方的演替”能否代表植被的演替是另外一个问题，不属本文的讨论范围。

（1984年在兰州草原生态会议上的报告，中国农业科学院草原研究所）

趋势分析 及其在生态股市中的应用



## 关于建立草原监测网的初步设想

为了开展生产力的动态研究，及时地指导牧业生产，掌握牧业生产的主动权及大范围、系统地积累资料，以便更深刻地揭示草原生产力与诸生态因子之间的关系，把草原科研推向新的高度，根据草原处李玉堂同志提议，8月份《草原退化趋势预测及对策研究》课题组碰头会上提出建立草原监测网的初步设想。

建立草原监测网的必要性与可能性。

我们这次承担全国的草原退化趋势预测，痛感草原科研上信息的缺乏，质不高、量不足，费劲找到的一点东西也是规格不统一，数据不完整，资料不系统，时间性不强，更不要说形成完整的信息网了。这与国家对草原牧业的要求，草原在国土资源与生态环境中的重要地位，与当今的信息化社会是十分不相称的。

在6月份西北四省区调研中，接触到许多草原科技工作者及有关专家学者，大家不同程度地提到，我们的草原科研体制基本上没有社会分工，仍然处于手工业作坊阶段。取样，资料整理，信息加工分析，撰写报告往往都要由同一个人来完成，许多信息在传递加工过程中散失，或因过期失效而失掉了价值。各研究者之间缺乏应有的流通与交换。许多有造诣的学者把大量精力耗费在多次重复的野外常规取样上，而有时又往往面对残缺不全的资料，不得不把希望寄托于下一个生长季。即

使组织大规模的资源普查也往往互相独立，前后缺乏对比，左右缺乏呼应；或者由于研究方法上持静态研究的观点，把资料、信息看作是僵死的不变的。生产周期很长，待成果出来时，往往时过境迁，失掉了用来指导生产的意义。

为了解决这一问题，过去曾经建了一些定位站，希望通过坚持长期观测，把草原的变化和时间联系起来。在加入时间轴后，将一个个孤立的点连成线，来开展动态研究。同时定位站上有固定的工作人员，科研有所分工。但由于当时的条件限制，信息的传递、转化、反馈较慢，未能用来及时指导生产，因此各定位站的生命力不够强，十年动乱中几乎都被冲掉了。

现在由于四个现代化对草原科研的要求，政府的重视，加之30年来的资源普查，使我们摸清了家底，掌握了一批基础资料。特别由于数学向生物学的渗透及电子计算机（器）的推广使用，使信息加工的速度、规模、质量较之过去都有了很大的提高；而且信息加工日益程序化、简单化，更易为普通科技人员所掌握，使我们有可能提出建立草原监测网的初步设想。

设想中的全国监测网将分为三个层次：基本单位——点（哨）、站，二级单位——台，最高为监测中心。

## 一、监测点（哨）、站

监测点是由上级监测台指定的，由监测站定期进行观测的样地。它应有一定的代表性，或设在具有重要经济意义的场、社内。其代表面积、植物名录、观测项目由上级监测台确定。

哨是指由非专业人员（经过短期培训的职工、知青、专业户）就近进行观测的监测点。

监测站为监测网的基本单位，负责对划定的固定监测点进

趋

势分析 及其在生态股市中的应用



行观测登记，实施监测工作、野外取样、积累原始数据，上报监测台。工作人员主要由初级科技人员组成。具备中专文化程度即可胜任。一般由5至10人组成。

### 基本装备

PC-1500 程序计算器、照相机、复印机、电话或电台，工作室、档案室、暗室，野外工作车（具有探照灯、无线报话机、简单实验室装备如天平、土壤箱等，车顶有观察台〈架〉）。

监测站一般下辖8—16个监测点及哨。监测站可由现有的旗县草原工作站扩充后担任。建站前首先要明确其监测范围，辖区内的草场类型和面积，辖区内的植物名录及频度顺序，取得重要植物的标本，并分别绘图、编号、存档。然后按草原类型面积，经济重要性逐步设立监测点。把有关信息存入监测中心的数据库。监测点一经设立后，即不得擅自中断观测。

### 观测项目

1. 日观测：限于重要的点、哨的气象观测。
2. 旬观测：各种植物高度，物候。以便编制物候谱。
3. 月观测：

a. 频度调查：横贯样地，按编好的植物名录以10平方米的样园，随机作20—100个记名样方，统计频度，登记。一般用于横向比较，历次结果可叠加。

b. 密度调查：取一平方米样地，分种计数登记。一般用于纵向调查，不叠加。单位：株（丛）数/平方米。

c. 盖度调查：以整个地段或测产样方为对象，主观分种评定盖度。分种盖度之和应当等于或大于总盖度。盖度（覆盖率）是衡量植被作用的重要指标，可用于水土保持等。为使各地的盖度估计趋于一致，需不断地培训交流，修正。在初期，盖度估计需伴有照片。

d. 产量调查：主观选定具有代表性的匀质地段，取一平方米样方，先按名录分别登记高度、估计盖度、计算密度，然后剪下称鲜重，登记。测产样方一般要做三个以上，互相拉开距离。产量样方主要用于计算生产力。做测产样方的同时，要兼作高度、盖度、密度调查。

e. 干鲜比：按植物名录，剪取定量的各种新鲜植物称重后登记。把样品置烘箱内摄氏 100 度干燥 48 小时后，再称重，登记，把样品粉碎后装样品袋。

f. 土壤含水量：测产后，随即用土钻取土样，装入标有准确重量的铝盒内称重。烘箱烘干后再一次称重，登记。土样装样品袋。

把频度表、产量表、干鲜比、土壤含水量及有关气象资料等表格复制两份。一份报监测中心，一份和样品袋上报上级监测台。

#### 4. 季观测

土壤风蚀的观测。把特制的有刻度的钢钎打入预定地点并设标志。逐季观察刻度有无变化。表土被风刮走，土层变薄则钢钎相对上升；风积沙土使表土层加厚，则钢钎相对下降。可作为水土保持的重要资料。

#### 5. 不定期项目：包括鼠害、虫害的调查。

点的布局、点上的观测项目、技术规程由监测中心每年召集的台站长会议统一制定及修改。

## 二、监测台

监测网的二级单位，领导监督站的监测工作，并对下级站报来的资料，样品进行核实分析，提出分析报告。每年 10、

趋势

势分析 及其在生态股市中的应用





11月份提出当年草场的演替趋势及有关年底存栏数的建议的报告。每监测台下辖8—16个站，工作人员主体应当是中级科研人员，具有相当于大专以上文化水平。一般由10—20人组成。

装备：微型计算机，化验室，野外工作车。

工作内容：

a. 对样方进行取舍：用标准程序

$$R_{(a,b)} = \frac{\sum A_{(i)} \times B_{(i)}}{\sqrt{\sum A_{(i)}^2} \times \sqrt{\sum B_{(i)}^2}}$$

计算一组样方两两之间的相似系数，画出星系图。舍掉相似系数很差的样方。当一组样方明显取于不同的异质群落时，可要求下级站重新取样。

b. 对各站上报的样方表等登记造册。依样方的代表面积为权，加权平均，求出更高一级单位的高度、盖度、产量等存档。

c. 计算优势度。植物种的优势度 = 相对盖度、相对产量、相对高度、相对密度和频度的平均值，取值范围0—1。以种的优势度为元素，列出样地的草场状态（对角）矩阵存档。

d. 将当月的草场状态矩阵与去年同期相比，与上月相比。用  $AX = B$ ,  $\rightarrow X = B/A$ ,  $X_{(i,i)} = B_{(i,i)} / A_{(i,i)}$  公式求出草场状态矩阵的各元素。  $X_{(i,i)} > 1$  时物种是上升的，  $X_{(i,i)} < 1$  则是下降的，说明草场的演替动态。预报下月，来年同期的草场状况存档。

e. 把当月的草场状态与预报的上月、去年同期草场状态相比较，相似系数大于0.7者为三级精度，0.8以上者为二级精度，0.9以上者为一级精度。对于精度达不到要求者，分析原因，并修正方程，或进一步加入二次项系数矫正。写出当月的草场状态分析报告存档。

f. 每年的8、9月份，根据当年的草场演替趋势，当年的

产草量, 实际牲畜密度, 在 10、11 月份分站提出当年的草场状态分析报告, 预报当年的牲畜存栏数存档。

g. 每年年底, 根据当年的降水总量、产草量、对降水—产量回归方程进行校正, 分站提出新的降水—产量回归方程存档。

h. 根据钢钎的刻度, 求当年的侵蚀模数 (吨/平方公里) 存档。

i. 对样品进行化验分析, 提出营养成分报告存档。

j. 根据历月的营养成分报告, 年底提出营养动态报告。

k. 根据资料, 列出不同地段的产量—株高, 产量—密度回归方程, 积累提供基础资料。

### 三、监测中心

监测中心为监测网的一级单位。向上对国家草原、环保、科研部门负责, 向下领导协调整个监测网的工作。每年负责召集台、站长会议, 对监测网的布局, 工作内容、规程、技术指标进行讨论修订。研制、提供草原监测网的运作方法及计算机通用程序。

监测中心建立数据库, 以印刷物或磁带的形式向国家及有关科研、教学单位提供不同年代不同地区的草场资料。监测中心以草原、生态、计算机方面的科研人员为主, 大约由十人组成。

装备: 配有数据库的计算机, 卫星照片判读设备, 卫星信号接收设备, 刻录像机。

建立全国的草原监测网是一件大事情, 新事情, 涉及面很广, 许多领域都是我们所不熟悉的。因为监测网是开展以草定



畜，掌握牧业生产主动权，保护草场资源和草原生态环境的重要手段，对发展草原科学也有重要意义。我们在这里提出来，希望草原界的各位前辈、各位同行都来关心、充实、完善它，促使它早日实现。

**2006年出版补充说明：**这是20年前写的一个关于建立监测网的设想。当时的侧重点是盖度，“草原是牧业生产的物质基础”，所以特别强调地上生物量和草原生产力的监测、记录。20年后的今天看来，至少还要加上根量的调查，因为草本植物的根系是维系固定土壤的主要机制、是防止土壤风蚀的重要指标。

(中国农业科学院草原研究所，1984年)

## 《草原退化趋势预测 及对策研究》 期中报告

根据(83)城环字373号文件精神,农牧渔业部环境办公室,畜牧局科技处于1984年3月份召集上海畜牧所、中国农科院草原研究所有关人员下达了牧业环境预测的任务。经过4月份北京上园饭店会议、郑州中州宾馆会议,进一步明确了:课题定名为《草原退化趋势预测及对策研究》;从属于《2000年中国环境》生态专题;由中国科学院环境委员会、农牧渔部(环办、畜牧局)共同负责,中国农科院草原研究所主持。

草原研究所5月16日在呼和浩特市召开了课题论证会,邀请所内外有关专家就课题意义、技术路线进行了讨论,充实完善了原设计方案。

按设计方案,6月份到甘肃、宁夏、青海、新疆进行了调查。在当地政府有关部门的大力支持下,搜集了畜牧、气象、草原、水文等方面的大量数据、资料。目前正在对资料进行初步整理,以便输入内蒙古大学的电子计算机系统。

据现有资料的初步分析,我国牧区畜牧业的基本生产条件,或说生态环境条件正在发生严重地、缓慢而持续地变化;草场有明显的退化趋势,水源、居民点附近退化现象十分严重。

以新疆八一农学院许鹏老师提供的紫泥泉种羊场(1966—1977)定位研究的数据看,该场各类型草场的平均产草量在按 $y = 141 - 5.2x$ 的方程式递减。若今后仍按此种势头递减,则不

趋

势分析 及其在生态股市中的应用



等到 2000 年将成为裸地。

新疆乌鲁木齐县。据 1965 年草场调查，平均亩产为 170.3 斤，据 1982 年调查亩产为 105.2 斤。17 年下降 65.1 斤，平均每年下降 3.8 斤，列出方程是  $y = B - 3.8x$ 。虽然近一两年雨水好，以至危机似有所缓和，但仍不可掉以轻心。

再以青海为例，据重点牧区草场资源调查的资料，几乎每个县都有草场退化的报告。其中海晏县 1980 年的报告指出，1980 年（平年）与 1973 年（欠年）调查资料比较，可食亩产青草由 1973 年的 146 斤下降到 1980 年的 114 斤，下降 22%。回归方程为  $y = B - 4.6x$ ，按当地的划分标准，退化草场占可利用草场面积的 32.4%，草原载畜量下降 28%。因草场退化而损失的可食草量折合绵羊后约占存栏数的十分之一。

草原植被的退化不仅严重地影响了畜牧业生产的稳步发展，而且引起了水土流失、气象条件恶化等严重的生态环境问题，其影响面之广还不仅限于约占全国二分之一面积的草原，必须引起全国上下的高度重视。

若以每年 8 级以上大风日数作为指标，据内蒙古镶黄旗 1959—1980 年的气象资料，每年大风日数由 60 年代的 30 天增加到 80 年代的 80 天，总的增长趋势是  $y = 20.8 + 2.5x$ ，预计到 1990 年大风日数将达到 95 天，到 2000 年将达到 120 天，即全年约有三分之一的天数要处于 8 级以上大风的威胁之中。

北京的大风日数也有增加的趋势，增加的速度是每年 1.1 天，10 年增加 11 天。方程是  $y = -38 + 1.1x$ 。就是说 40 年代似乎是没有大风的，50 年代开始有大风。预计到 1990 年大风日数将达 55 天，2000 年大风日数将达 66 天，差不多每周一天。

在 6 月份西北四省区的调查中，我们和当地的科技人员，有关专家学者、领导进行了接触，就课题的设置、实施交换了

意见，对退化草场的有关技术问题进行了商讨，加深了对课题的认识。有关部门对课题表示了很大的热情和支持，为我们提供了许多工作上的便利，包括允许我们查阅、摘录大量尚未公开的第一手资料。青海、新疆、甘肃对把课题继续搞下去，进一步深入开展草原的动态的定量定位研究表示了很大的兴趣。他们希望尽快地在畜牧局立上题。由业务主管部门正式下达此项任务。

后期工作安排：

8月下旬，在锡盟达里诺尔地区2号机井附近退化草场上布置一个空间系列，以补现有资料时间系列的不足。此项工作由草原研究所石永怀同志负责。

同时，8月下旬即开始数据输入工作，编制调试程序。预计11月份计算出结果。此项工作由内蒙古大学郝敦元负责。

9、10月份，内蒙古畜质、草场、气象调查重点蒙东地区。由草原研究所硕士昭那顺负责。争取1985年1月份在呼市召开《草原退化科学讨论会》。参加课题各省区11月底前交出草场退化单行报告。联系人：新疆程向蛟，甘肃车文信、内蒙于中流……。

原设计方案中的云贵川藏调查等因为经费尚未落实，暂不确定。

年底交初步报告，3月份成果鉴定。

对策部分。我们准备本着：少花钱、多办事；牧业产值翻两番的原则，重点放在探讨牧业生产经营方针，从维护生态平衡着眼，分析牧业生产内部矛盾入手，对来自外部的因素，如盲目流入人口、滥垦滥牧（购）野生动植物资源，不过多纠缠的精神，提出以下初步设想。

1. 扩大放牧半径

A. 努力稳步提高大畜比例。

趋势

分析 及其在生态股市中的应用



B. 研究把现在的定居点、饮水点同一，改为定居点与饮水点分开一定距离，或改现在放射状的放牧方式（每日从定居点出发，晚间归牧）为围绕定居点，沿简易休憩地轮牧，夜间不回定居/饮水点的放牧方式。如此要考虑引入牧羊犬及解决输水工具，增设临时流动饮水点。

2. 以草定畜、合理控制牲畜密度。提高商品率、发展季节畜牧业。坚决处理、淘汰超载的牲畜，以减轻草场压力，节约人力资金，提高经济效益，改善牧民生活，提高畜质。这是维持畜草平衡，使牧业生产及生产环境向良性循环转变的关键措施。

我们现在实行的是商品生产。因此提高商品率不仅是生产措施，而且是我们的生产目的。

由纯增定义式：纯增 = 总增 - 死亡 - 自食 - 出售，移项可得：商品率（出售）= 总增率 - 死亡率 - 自食率 - 纯增率。式中，纯增率可正可负，而其他各项一般都大于零。为使商品率最大，在自食不变情况下，就要使：

A) 死亡最少，这除应加强疫病防治外，还要求及时处理乏弱畜，使之转化为商品率。

B) 要使总增是最大。总增率 = 母畜率 × 配种率 × 受胎率 × 保胎率 × 成活率。所以在提高其他几个指标的同时，调整畜群结构、提高母畜比例很有必要。

C) 减掉纯增。积三十年经验，丰年的纯增率往往最后转化为欠年的死亡率，所以根据年景，尽量使纯增小，以至必要时取负值是可行的、经济的、科学的。

3. 为了实行以草定畜，正确确定当年的商品率、纯增率、年底存栏数，掌握牧业生产的主动权，同时开展草原的动态研究，把我国的草原科研推向新的高度，建议逐步建立全国的草原监测网。每年 8、9 月份测产，10、11 月份处理超载牲畜，

关于建立草原监测网的初步设想见附件。

4. 为了在有限的牲畜基础上无限地发展生产，满足社会需要，提高人民生活水平，要提高牲畜质量，提高畜均产值。改良畜种，提高大畜比例，因地制宜保护发展牛、驼、马。良种畜与土种畜、大畜与小畜的畜均产值可差几倍以至几十倍，牧业生产的无止境应体现在生物体的无穷的进化和变异上，而不是依赖于有限的头数上，为了贯彻推广“种植业与畜牧业并重、畜草平衡”、“数质兼顾以提高质量为主”的原则，建立开展生态牧场的研究。

贯彻“优质优饲劣质淘汰”的科学原则，保留肥壮、生产性能高的家畜，坚决淘汰老弱病残，以解放草场、资金、人力，保证畜质、提高经济效益，并形成一种推动畜质提高的生态压力。

5. 尽快颁布草原法，固定草牧场使用权。保护草原是关系到国计民生、子孙后代的大计，一切局部暂时利益都应服从这个大局，励行禁止开荒、保护牧场。滥垦草场不仅侵犯了牧民的切身利益，也损害削弱了中华民族的生态屏障，危害了全国人民的长远利益，应予明令禁止，对经说服教育无效者应予以制裁。

解决牧民燃料、牧区能源问题。开展风能、太阳能和沼气的研究、推广。国家把资金用于解决牧民燃料问题，较之投资搞基本草牧场建设而任其砍伐灌木，效益可能更高。

关于编写年终报告的几项技术要求：

8月份课题组碰头会要求，青、新、蒙、甘四省区分别拿出本省区草场退化趋势预测及对策的单行报告，以便12月份汇总。同时报地方政府主管部门，以引起地方政府的重视。在编写报告时，要尽量争取区域预测的合作。报告中请采用下述单位、格式：





1. 所有单位采用公制，草原面积采用平方公里，以利于图上作业，测产、计算载畜量及牲畜密度。

2. 把理论载畜量与实际载畜量分别定义，一个称载畜量，由草原产草量克/平方米 = 吨/平方公里，家畜采食量吨/年。只计算得到。另一个称牲畜密度，由草场上实际牲畜数（折成羊单位）/草场面积得来。前者是草场容量，后者是实际负荷。两者差值为正，则表现为生产潜力，差值负则表现为超载。

3. 各地草场在最后报告中均按匀质对待。测产时，搜集资料时应按草原分类系统，分类统计汇总，然后以面积为权，加权平均，求出各地区草场的平均高度、平均盖度、平均产量。不可利用草场面积产量为零。

4. 草原畜牧业综合生产指标暂订为畜均产值，单位：元/畜，畜牧业生产的整个流程是：水光→草→畜→畜产品→货币。以单位草场提供的商品来衡量是比较理想的。但鉴于目前我们生产者及生产指挥者多有追求头数的自发倾向（因为在小生产者看来：头数，而且只有头数才是财富的象征、生产水平的标准）而危及基本生产条件，为矫正，我们建议并在报告中取产值/总头数作为衡量指标，以利于加强经营管理、加强改良、提高大畜比重，改变畜群构成，提高商品率。

#### 5. 草场退化的划分标准

在本文中“由于人类不当的经济活动所引起的草原第一性生产力明显持续下降的现象”称为草原退化。草原退化分为三个等级：轻、中、重。

轻度退化：草原建群种（组）优势度已有下降，年平均产草量下降但不超过四分之一者。

中度退化：建群种（组）优势度下降到与其他种（组）共建，产草量下降二分之一者。

重度退化：原建群种（组）优势度已下降到次要地位以至

消失，产草量下降到原来的四分之一以下者。

优势度：取相对盖度、相对密度、频度、相对产量、相对高度（某几个指标）的平均值。

6. 草场退化趋势预测采取一次函数模型。当植物群落分种计算时，由  $AX = B$  推导出  $X = A^{-1} \cdot B$ ，推导出  $C = BX^k$ ，其中  $A$ 、 $B$ 、 $C$  都是对角矩阵， $X$  是转移矩阵， $A$ 、 $B$  是连续二次的草场状态， $C$  是  $k$  个时间间隔后的草场状态。

在不分种时，以年度为自变量，产草量为因变量，根据历史数据求  $y = B + AX$ ，回归方程。解出  $B$ 、 $A$  后，预测 1990、2000 年草原产草量（或盖度、高度、优势度）。

在有条件时，重点地区可做降雨量，产草量的回归分析；降雨量与产草量比值（暂定名为降水有效系数）的变化趋势预测。

7. 要算经济账，与总产值翻两番联系起来。程式：退化草场面积、强度、速度 → 牧草损失量 → 家畜、畜均产值 → 经济损失。

同时尽可能算生态环境账，把草原产草量和气候因子（大风日数）降雨量和河流含沙量、输沙量、侵蚀模数进行回归分析。

（中国农业科学院草原研究所，1984 年）



# 多维超球面模型通讯

Vol. 2 No. 4

一个新的，用于趋势分析，系统监测的多元分析方法——超球面模型。

1995年5月28日

中国科学院应用数学所负责人：

老师，您好。

我叫白丽格吉扎布，中国公民，蒙古族，植物生态学博士，专业方向数量生态学，美国科罗拉多州立大学自然资源学院森林系副研究员。近年从事生态环境和自然资源的趋势分析和系统监测保护方面的研究。在研究和实践中，我认识到，我们可以用变量空间（而非样本空间）中的向量（而非矩阵）来代表系统的状态，而状态向量在多维空间的角度（而非距离），携带着系统各组成成分的信息，我们可以通过监测代表系统状态的向量在多维变量空间中方位角的变化，来监测系统的变化。

因此，我逐渐形成“多维超球面模型”的概念，其基本想法是：用向量代表系统的状态，用后两个时间段的状态向量的商作为趋势向量，趋势向量和第二时间段的乘积作为第三时间段的状态向量预报值。在预报的同时，进行取样，预报值和取样值的加权平均作为系统状态的期望值。由于期望值有两个信息来源：来自历史资料的预报值和实际现场的取样值，可以互相校正，降低误差，减少取样规模，降低监测成本。通过后两个期望值计算趋势值，及下一时间段的预报值；届时再取样。...如此循环往复达到监测系统动态的目的。

具体做法包括数据中心化，标准化，计算趋势向量，预报值，期望值，和误差。中心化就是归并样方，用样方组在多维空间的形心代表系统的状态，标准化是状态向量除向量长度，把多维空间中的点投影到单位超球面上，标准化向量表明系统的状态；而标准化向量的元素表明相应变量在系统中的重要性，某变量在后两个时间段的重要性的比率表明该变量在系统中的演变趋势。了解了所有元素的演变趋势，我们就了解了整个系统的演变趋势。反过来，当我们了解了系统的演变趋势后，对个别的变量就有了更深刻的了解。（比如，当我们了解了整个股票市场的趋势，市场上每个股票的走向后，将更有助于我们对某一特定股票采取行动，或买，或卖，或观望。）预报值是趋势值和后一时间段的取样值的乘积，期望值是预报值和取样值的加权平均，误差是预报值和期望值的夹角正弦值。

超球面模型的应用对象是：资源有限（平方和为一），变量互相独立（变量构成空间的坐标轴），变化趋势的期望值为一的系统。因此将在资源管理，生态系统，环境科学，财务金融方面有一定的用处。我自觉方法有一定的独到之处，尽管仍有待于进一步改进，完善。但由于我原来是工农兵学员，又不是数学专业毕业，所有的一点一知半解的数学知识多是在工作实践中积累的，很不足以在短期内完成，推广这一数学方法，希望能得到你们的指点，帮助，和扶持。

我将于六月5日到13日到北京参加第18次太平洋科技大会，希望届时能面见贵所的有关人员当面请教。

祝好

白·丽格吉扎布，于美国科罗拉多州立大学1995年5月28日

白丽格吉扎布博士的来信  
读后模型颇有新见，并有言回作佳话得悉  
经济出版社  
1995.6.13

用多系统超球面模型  
进行趋势分析有一定的应用价值，希望在实践中积累经验使模型更加精确，这样才能真正付诸实践。

模型更加精确，才能真正付诸实践。

1995.6.2

《草原退化趋势预测及对策研究》期中报告

## 卡森堡植被演替趋势分析

### ——超球面模型应用实例

**Abstract** The Multi-Dimensional Sphere Model (MDSM), a new multivariate analysis method, was used to analyze vegetation successional trend on Fort Carson using Land Condition Trend Analysis (LCTA) data collected in 1989 and 1993. MDSM analyzes vegetation successional trend by species, and is designed to include as many species as required. MDSM expresses the result by vector, instead of a scalar. Data on 188 species were collected at 203 transects (quadrats) on Fort Carson in both 1989 and 1993. This concept of trend analysis was explained using a student's report card. In this analogy the trend was determined by comparing a student's (installation) mid-term and final grades (1989 and 1993 data) by subject (species). The parent (land manager) analyzes the trend to determine what action to take. The trend analysis was conducted by first centralizing and standardizing the 1989 and 1993 data. The trend value for each species was then obtained by dividing the standardized 1993 value by the standardized 1989 value. There was no change in trend if the 1989 standardized value was identical to that of 1993, for the trend value would be one. The trend was up if the 1993 standardized value was greater, and the trend was down if the 1993 value was less than that of 1989. The data was used to evaluate range successional trends of Fort Carson at both the in-



stallation level and the management area level. The data indicated that installation range succession was in an upward trend and that range succession trends varied by management area.

MDSM was also used to estimate the future states of the range condition. The results can assist the land manager in determining what action, if any, would be required. All analyses were accomplished using Quattro Pro 5.0 for Windows.

## 一、前言

自然资源监测,趋势分析曾经是植被科学家间很热门的话题。尤其是自1971年联合国教科文组织发起“人与生物圈”计划后,各国政府纷纷制订长期生态定位研究计划。如美国的“长期生态研究(Long Term Ecology Research, LTER)”,中国的“生态研究网(CERN)”,投入巨大的人力,物力、年复一年地定位取样,积累数据。希望通过定位研究,加入时间因素,发现植被的演替规律。然而,提出定位研究的生态学家当初不知道有没有意识到,在提出生态监测的同时,他们提出了一个新的研究方向:研究样本之间的关系;历史的样本、现在的样本和未来的样本之间的关系。1985年,美国陆军发起“土地条件趋势分析(Land Condition Trend Analysis, LCTA)”计划。其协调中心(Center for Ecological Management of Military Lands, CEMML)办公室,1993年转移到科罗拉多州立大学森林系。科罗拉多州的卡森堡是LCTA试点地区,自1989年开始执行LCTA。其数据库包括203个6米×100米的永久固定样条。变量包括:土壤、水文、植物、野生动物等。到1994年,虽然全美已经有30多个基地执行LCTA,但是,至今仍然没有能利

用数据写出趋势分析的报告。原因是 LCTA 没有解决数据（趋势）分析及预测预报的方法问题，没有提出相应的分析软件。按 USLE (Universal Soil Lose Equation) 所揭示的土壤流失规律 (Wischmeier & Smith, 1978)，土地条件趋势分析的核心问题应该是植被演替趋势分析。而由于植被是由多种植物组成的，其分析工具应当属于多元分析的范畴。目前的多元分析技术，多用 R 分析，其基本作法是在样本空间中分析变量之间的关系。通过随机取样，假定样本之间是独立的，而假定变量之间是相关的。通过分析，在许多变量中提炼出主要变量或变量的线性组合，描述变量之间的关系。近年，数量生态学家也经常采用变量空间，进行 Q 分析。以植物种为轴构成变量（物种）空间，在种空间中分析样本。其基本指导思想被小高奇 (Gauch, 1982) 阐述如下：“种空间完全代表了原来种 - 样方数据矩阵的多度值（就像样方空间那样）。问题仅仅在于，种空间的维数是高的，要将原来的高维种空间投影到一较低维的种空间，比如说二维或三维空间上，这样在选定的二维或三维上，其结构是清楚了，但其他维上的信息却失掉了。”显而易见，小高奇所说的种空间的方法，更接近于我们分析样本之间关系的目，问题在于如何保持所有维上的信息。因为在植被演替趋势分析中，我们更关心的是整个植被组成的变化，而不仅仅是两三个建群种优势种的变化。在趋势分析中，任何微小的植物种的细微变化，都可能传递着植被演替的重要信息。问题是，我们能否找到一种新的投影方法，既可以保留种空间所有维上的信息，又便于人们理解掌握呢？答案是肯定的。作者在硕士论文 (1982) 中提出的“样方除以样方长度，向单位超球面投影”，便既保留了所有维上的信息，又为人们进一步的分析研究提供了可能。投影到超球面，植被向量保留了方向角度，而被滤掉了距离。因为我们认为是向量的方向承载着植被组成的

趋

势分析 及其在生态股市中的应用



全部信息。比如，三维种空间的向量  $A = (1, 2, 3)$  所代表的植被，不同于  $B = (1, 2, 1)$ ，不同于  $C = (1, 1, 3)$ ，而相同于  $A_2 = (2, 4, 6)$ ，或  $A_{10} = (10, 20, 30)$  所代表的植被。向量  $A = (1, 2, 3)$ ， $A_2 = (2, 4, 6)$ ，和  $A_{10} = (10, 20, 30)$  在多维种空间中的方向是一致的，位于同一条发自原点的射线上，尽管与原点的距离各不相同。从植被组成角度来说，植被在物种空间的射线和射线在单位超球面上的投影点是一一对应的。植被的每一个组成变化，射线的每一偏转，都将反映在超球面上代表植被的点的位移。这就为我们进行植被监测提供了可能。本文使用超球面模型 (Multi-Dimension Sphere Model, MDSM) 对卡森堡 1989 年和 1993 年的数据进行汇总、标准化的基础上，进行植被演替趋势分析。分析结果被草场管理员认为符合实际情况，可行，并推荐到美国草原学会年会和全美土地利用—恢复会议上发表。

## 二、方 法

为介绍抽象的数学概念，我们使用比喻。假定期末有个学生回家带着成绩通知单。成绩单分为：语文、算术、自然常识、舞蹈及总评五项，每项有个分数。对学生来说，也许总评分数是最重要的。因为总评成绩决定其能否升级。对家长来说，他们也许更希望全面了解孩子的每个科目的成绩。更细心的家长可能还会想到把期末和期中的通知单进行比较，以便知道孩子的发展趋势。而且不仅要比较总评，还要进行分科比较。其趋势分析结果可能是：语文持平，算术显著下降，自然常识、舞蹈有进步。所以总评成绩看起来可能有上升，但那是由于舞蹈、自然常识把总分拉高的结果。由于算术成绩显著下

降，实际问题可能比较严重，需要采取适当措施：谈话，家庭辅导……

现在，设想我们是家长，卡森堡植被是学生，我们收到了前后两张成绩通知单。一张是 1989 年的植被调查数据，来源于 203 个永久样条；另一张是 1993 年的植被调查数据，来源于同样 203 个永久样条。每个植物种的重要值相当于科目成绩。30 个植物种（根据方差和相关分析从 188 个植物种中选出 30 个）在两个状态矩阵中按相同顺序排列，如演替趋势分析表所示。表的第 1 列是植物种名码，用属 - 种拉丁学名的头两个字母表示。如数据的第 23 个元素 BOGR 表示蓝格拉玛草 (*Bouteloua gracilis*)，北美草原的一个主要建群种。表的第 23 行是 BOGR 的有关数据。30 个植物种的数据到第 30 行为止，第 31 行是向量长度 (Length)。第 32 行是趋势指数 (Index)，只有一个元素。

数据第 2 列 AVG89，是卡森堡 1989 年的 203 个样方的平均值，30 维向量。第 31 个元素是向量长度。如表所示，1989 年植被向量长度等于 39.088，是 30 维向量 AVG89 的 30 个元素的平方和的算术根。第 3 列 STD89，是 1989 年的植被状态向量，表明植被的相对组成，也是 30 维向量。STD89 的值来自于  $AVG89/Length89$ ，各分量除以向量长度。第 5 列，第 6 列是相应的 1993 年的植被数据。

以上是趋势分析前的数据简缩工作，包括中心化和标准化。为了简明清晰，方便易懂，在应用上，我们先将数据中心化：在种空间中，归并所有的样条，用它们各分量的平均值（几何上是形心）所组成的向量代表整个植被；然后标准化，向量除以向量长度。超球面模型定义标准化，中心化后的向量为植被状态向量，其分量为相应植物种的重要值。

趋势分析表第 4 列是趋势向量，前后两时间段植被状态向





量的比值。MDSM 用前后两状态向量的商表示演替趋势。其基本思想是：植被组成于植物，了解了所有的植物，我们就了解了植被。同时，一植物种在后前植被中的重要值表明了植物种在植被演替中的动态。了解了所有植物种在植物群落中的动态，就了解了植被的动态。MDSM 把一个植物种 1989 年的重要值和 1993 年的重要值的商定义为趋势值，来表示相应植物种的演替趋势。趋势值大于 1，该植物种在上升。如第 13 个趋势值等于 1.16，表明相应的植物种冰草（AGROP）在植被中的重要值在上升，上升率是 1.16。相反，如果趋势值小于 1，表明相应的植物种在下降。如第 23 个趋势值等于 0.97，表明相应的植物种蓝格拉玛草（BOGR）在植被中的重要值在下降，下降率是 0.97。分析优势种，建群种，各已知的生态指示种，以至全体植物种的趋势值，就可以全面地了解植被的演替趋势变化。MDSM 定义各植物种的趋势值的平均值为伪趋势指数，第 32 个元素，表明整个植被的一般演替趋势。卡森堡的伪趋势指数是 1.23，意味着，根据 203 个随机永久样条 30 个植物种的数据，卡森堡的植被总的演替趋势呈上升趋势，其平均上升率为 1.23。

为了估计植被的未来状况，我们需要假定植被将保持目前的演替趋势。我们不能指望所有的植物在所有的情况下都会保持目前的趋势，但我们至少可以指望大多数的植物在相当一段时间里将保持目前的演替趋势。因此我们把植物种按趋势值的大小排列。在两端各删去百分之十的取极端值的 6 个植物种：上端的 SIHY, ORHY, BRIN2, 和下端的 HIJA, GUSA2, SPORO。根据剩下的 24 个植物种对未来 1997 年和 2001 年卡森堡的植被状态进行了预报。预报值和标准化后的预报值列在第 7, 第 8 列。未来 1997 年的植被状况的预报值是趋势向量继续作用于 1993 年植被向量的结果： $97 = 93 \times T$ 。向量 97 再投影

到超球面上，得到 STD97。以上过程可以重复多次，当然预报误差也就会增加。预报越远的状态，我们的把握就越小。

### 三、讨 论

超球面模型是很有潜力的植被分析工具。它所涉及的多是植物数量生态学中很基本的问题，诸如：我们感兴趣的是样本之间的关系，还是变量之间的关系？以我的理解，生态定位研究的目的是要了解历史的、现在的和未来的样本之间的关系。那么，我们的出发点是样本空间还是变量空间？植被分析的基本数学工具是数组的，还是矩阵的？如果我们用向量（数组）代表植被，则是向量的方向还是距离承载着主要的植被信息？物种空间里作为坐标轴的植物种要互相独立。我们的植物种是相对独立，还是高度相关？MDSM 用标准化后的向量表示植被状况。如果把标准化解释为百分比向多维的扩展，则植物的重要值是无名数。这是否意味着 MDSM 允许不同规格的样方、样线混合使用？这是否意味着扩大样方的来源，以至植被数据在世界范围的交流？限于篇幅和我们现在对 MDSM 的粗浅理解，我们还不能在这里讨论这些问题，下面仅仅就投影到超球面的几何意义进行一些浅显的讨论。

通过除以向量长度，MDSM 把代表植被的多维种空间中不可琢磨的点投影到单位超球面上。从远距离向圆心看，我们得到超平面，其上分布着代表植被的点。从另一角度，如果用两个向量确定的平面切割超球面，我们得到单位圆，圆上分布着代表植被的点的投影。这样，我们可以使用现成的几何工具，分析处理平面上或圆上的点。在初始的多维种空间中，向量的和是以两向量为边的平行四边形的对角线；若干个样本的平均

超

势分析 及其在生态股市中的应用



值是它们在高维空间点的几何形心，平均距离是方差。投影到超球面，保持角度，滤掉距离可能会给数据带来某种程度的失真。但当向量长度远远大于1时，这种失真可以小到忽略不计。我们仍然可以在超平面上组合、分解它们，进行加减运算。超球面上两点的和是其圆心角平分线在超球面上的投影。两个向量夹角的余弦值（一向量在另一向量上的投影），可以显示两个样本的相似性，正弦值显示距离，误差。

1. MDSM一般是先标准化数据，然后在超球面上聚类，中心化数据。在植被演替趋势分析时，中心化和标准化的先后次序，对分析结果影响很小。

2. 向量乘除法的应用，已经比较广泛见于算表软件。

3. 此定义为伪趋势指数。真正的趋势指数在指数模型中的定义应当是： $M$ 个趋势值的连乘积的 $M$ 次根。

4. 超球面的半径等于1的生物学意义可以解释为，资源共享，此消彼长。可以解释群落中一些貌似负相关的现象。

## 后 记

本文向国内同行简介一种新的多元分析技术——超球面模型及其应用。原稿是在全美土地使用、恢复第三次会议（Aberdeen, MD., USA. 1994年）上的发言，对象是草场管理员，重点在于如何应用个人计算机进行趋势分析。1994年9月应母校内蒙古大学邀请，回校参加现代生态学学术讨论会。会后根据幻灯片和英文稿翻译整理，并做重大修改。

演替趋势分析表

	AVG89	STD89	T	AVG93	STD93	1997	STD97
SIHY	0.7931	0.02	3.57	2.8227	0.07	<u>0.26</u>	<u>0.24</u>
ORHY	1.2611	0.03	2.91	3.6601	0.09	<u>0.27</u>	<u>0.26</u>
BRIN2	0.6257	0.02	2.04	1.2759	0.03	<u>0.07</u>	<u>0.06</u>
JUNCU	1.1724	0.03	1.82	2.1330	0.05	0.10	0.09
SALIX	0.5665	0.01	1.50	0.8473	0.02	0.03	0.03
ATCO	0.3202	0.01	1.45	0.4631	0.01	0.02	0.02
SAVE4	0.1232	0.00	1.44	0.1773	0.00	0.01	0.01
ATCA2	0.3793	0.01	1.38	0.5222	0.01	0.02	0.02
FRJA	0.4187	0.01	1.37	0.5714	0.01	0.02	0.02
CHNA2	0.4384	0.01	1.34	0.5862	0.02	0.02	0.02
CEMO2	2.2217	0.06	1.19	2.6305	0.07	0.08	0.08
QUGA	2.6256	0.07	1.17	3.0591	0.08	0.09	0.09
AGROP	7.8867	0.20	1.16	9.1379	0.23	0.27	0.26
ORMI2	0.6207	0.02	1.15	0.7094	0.02	0.02	0.02
YUGLG	1.5862	0.04	1.11	1.7537	0.04	0.05	0.05
PIED	11.6946	0.30	1.08	12.576	0.32	0.35	0.33
EULA5	0.5468	0.01	1.07	0.5813	0.01	0.02	0.01
ARIST	4.5616	0.12	1.05	4.8177	0.12	0.13	0.12
JUMO	3.4433	0.34	1.05	14.024	0.36	0.38	0.35
RHTR	1.0788	0.03	1.03	1.1133	0.03	0.03	0.03
PIPO	1.8670	0.05	1.02	1.9064	0.05	0.05	0.05
STIPA	3.7685	0.10	0.98	3.6847	0.09	0.09	0.09
BOGR	32.2020	0.82	0.97	31.004	0.80	0.77	0.72

趋势

趋势分析 及其在生态股市中的应用



ANGE	1.0788	0.03	0.96	1.0345	0.03	0.03	0.02
BOCU	2.6946	0.07	0.94	2.5369	0.07	0.06	0.06
OPIM	0.2611	0.01	0.89	0.2315	0.01	0.01	0.00
OPUNT	0.9606	0.02	0.85	0.8128	0.02	0.02	0.02
HIJA	1.9901	0.05	0.66	1.3153	0.03	<u>0.02</u>	<u>0.02</u>
GUSA	1.3202	0.03	0.58	0.7586	0.02	<u>0.01</u>	<u>0.01</u>
SPORO	5.2266	0.13	0.53	2.7537	0.07	<u>0.04</u>	<u>0.03</u>
Length	39.088	1.00		38.991	1.00	1.07	1.00
Trend			1.2341				

(1995年发表于李博主编《现代生态学讲座》)

## 超球面模型与 植被演替趋势分析

**摘要:** 多维超球面模型 (MDSM), 被设计用来处理多变量时间系列——三下标变量  $D_{(i,j,k)}$ 。当用于植被动态分析时, 三下标分别为植物种、样方和时间。MDSM 用植物种建立变量空间并用变量空间中的向量来代表植被的状态, 是为状态向量。状态向量在多维空间的角度携带着植被组成的信息。因而, MDSM 以度量状态向量在多维变量空间中的方位角的变化来监测植被的变化。在实际应用中, 通过除以向量长度, MDSM 把变量空间中代表样方的组成的点投影到单位超球面上, 再根据向量之间夹角余弦值的大小聚合它们, 是为中心化, 以取得植被的状态向量。状态向量的分量, 是各植物种在植被中的重要值。当进行趋势分析时, MDSM 定义后前后两时间段的状态向量的比值为趋势向量, 用以描述整个植被的演替趋势。而趋势向量的分量, 植物种的趋势值表明该物种在植被中的重要值随时间的变化。由于 MDSM 可以容纳尽可能多的植物种, 因而是对植被更好的模拟。由于 MDSM 在容纳多变量多样方的同时兼顾到了时间因素, 所以可用于植被动态的研究。用美国卡森堡军事基地土地条件趋势分析的数据 (1989—1993) 试行趋势分析表明, 多维超球面模型可以对每个植物进行单独分析, 同时又兼顾到植被整体的变化, 从而取得了比较可信、满意的趋势分析结果。MDSM 还被用来估计未来 1997 年的植被状态。

**趋势分析** 及其在生态股市中的应用



**关键词:** 趋势分析 状态向量 多变量时间系列 三下标变量  
超球面 相关分析

**Abstract** The Multi-Dimensional Sphere Model (MDSM), a new method of multivariate analysis, is introduced here for analyzing trends in vegetation succession. The MDSM uses individual plant species, or species groups, as dimensions of a multi-dimensional space, and quadrats as points (vectors) in the space. All quadrats are projected onto the unit supersphere. The model analyzes all species simultaneously and independently as each species is treated as an orthogonal axis. The MDSM defines the quotient of elements of previous and current standardized state vectors as a successional trend, expressed as a trend vector, analyzing vegetation composition change over time. This successional trend can then be extended to predict future states of the vegetation. The model was tested with data from the Land Condition Trend Analysis at Fort Carson in southeastern Colorado, showed good results for vegetation succession trend analysis.

**Key words:** Land condition trend analysis, Supersphere, Euclidean distance, Trend analysis, Trend vector, State vector

自 1971 年联合国科教文组织发起“人与生物圈”计划以来,自然资源的保护、生态环境的监测日益引起人们、各社会组织、政府的重视。继美国科学基金会的 LTER (Long Term Ecological Research 长期生态研究计划, 1970) 后, 美国军事部门又提出 LCTA (Land Condition Trend Analysis 土地条件趋势分析, 1984) 课题。美国军事基地卡森堡自 1989 年正式执行 LCTA 课题至今, 今年开始使用一个新的数据分析技术——多维超球面模型 (MDSM), 用以分析植被演替趋势、速度, 并估计未来

的植被状态。长期以来，人们一直非常关心、担忧我们赖以生存的环境，特别是植被的状况，希望能认识、了解其变化规律，监测它的动态变化。过去，植被演替趋势分析多定性的描述，而且仅限于几个优势种。由于植被是由多种植物组成的，其定量分析工具属于多元分析的范畴，而植被数据一般取双向数据矩阵  $D_{(i,j)}$  的形式。其中  $i$  和  $j$  分别是植物种和样方。由于趋势分析还要考虑时间因素，所以植被演替趋势的分析所面临的是三向数据矩阵或三下标变量  $D_{(i,j,k)}$ 。其中  $i$ 、 $j$ 、 $k$  分别表示植物种、样方、时间。然而，群落生态学中常用的基本方法（如排序和分类）的计算只能接收双向数据矩阵作为输入，对于如何处理三下标变量或者说多元时间系列，目前还不见有确定、成熟的数学方法。中国兰州大学赵松岭等提出使用线性系统和转移矩阵来解决植被演替的趋势分析的可能性，并成功地预报了青柑林的演替。但赵松岭在文中没有给出状态转移矩阵的通解。中国农科院草原研究所白·图格吉扎布等继续赵松岭的研究，引入对角矩阵，给出了状态转移矩阵的一个特殊解，即通解。差不多同时，美国科罗拉多州立大学的 Jameson 在引入矩阵伪逆（Pseudoinverse）的基础上导出了系统监测的基本框架，提出预报值和采样值加权平均，用现场采样来矫正预报，以提高监测精度，降低监测成本。综合上述研究，我们在美国第 47 届草原学会年会上首次提出超球面模型，并用于卡森堡的植被演替趋势分析。

MDSM 最突出的特点使用向量来代替对角矩阵，因此可以同时分析所有可能的变量，以反映多元的植被的成分变化及状态变化。在进行植被演替趋势分析时，对于三向数据矩阵或三下标变量  $D_{(s,p,t)}$ ，MDSM 认为变量（Species）之间是独立的，并且以物种为多维空间的坐标轴建立多维变量空间。对于样方（Plot），MDSM 认为样方不是演替的基本单位，不是





MDSM 的基本分析对象。由于每个样方的每个数据的实现都带有很大的偶然性，样方在传递植被演替信息的同时夹杂有很强的“噪音”。因此 MDSM 在样方的方向上聚类，希望通过归并来自同一母体的样方抵消噪音，加强信号。对于一块理想的匀质植被，尽管取样面积可以不同，数据可以不同，比如 200 米的样线和 2000 米的样线，虽然后者的数据是前者的 10 倍，但它们反映着同一植被。因此，在聚类时，MDSM 使用向量的夹角的余弦值，而非距离，作为相似系数。通过聚类，MDSM 使每个植被类型最后由一个状态向量来代表，三向数据  $D_{(s,p,t)}$  被简缩为二向数据  $D_{(s,t)}$ 。最后，MDSM 在时间方向上进行趋势分析。MDSM 定义后前时间段的状态向量的对应分量的比为趋势向量的分量；趋势向量和状态向量的对应分量的积为预报向量的分量。若  $A$ 、 $B$ 、 $C$  是植被演替系列中的三个连续状态，且趋势向量为  $T$ ，则有  $AT = B$ ， $T = B/A$ ， $C = B^2/A$ 。

MDSM 的另一优点是计算简便，只牵涉加、减、乘、除、乘方、开方。方法简便到可能会使一些数学家不屑一顾：“不就是百分比和除法向多维的扩展吗，太简单了。”事实上，化繁为简是科学研究的基本原则之一。介绍一种简便可行的数学工具正是本文的目的。

## 一、方 法

MDSM 用于植被的演替趋势分析时，首先建立物种空间，也即确定参加分析的植物种，建立植物种一样方表格，而暂不考虑时间的变化。如表 1 所示，横行是种的数据，两个植物种：物种甲和乙；纵列是样方数据，三个样方： $A$ 、 $B$  和  $D$ 。样方  $A$  含有物种甲 40，物种乙 80；样方  $B$  含有物种甲 64，物

种乙 77；样方 *D* 含有物种甲 117，物种乙 53。MDSM 以植物种作为多维空间的坐标轴；而以样方作为多维空间中的点，或者一个由原点指向代表样方的点的向量。表 1 所示数据反映到 MDSM 模型中可以用图 1 显示；横坐标代表植物种甲，纵坐标代表植物种乙，点  $A = (40, 80)$ 、 $B = (64, 77)$ 、 $D = (117, 53)$  代表三个样方 *A*、*B* 和 *D* 在二维空间的相对位置。

### 1. 样方聚类（中心化）

如上所述，在趋势分析中多维超球面模型分析的对象是植物类型而不是样方。在建立空间后，MDSM 把样方分组，用样方组来代表不同的植被类型。因为不同的植被类型对环境压力的反应不尽相同，演替趋势也不尽相同。因此，在演替趋势分析中保持植被的同质性（Homogeneous）十分重要。组成相近的植被，对环境压力的反应相近，在多维空间中的运动轨迹也相近；依照相似系数归并样方，可提高样方组内的同质性及组间的异质性。而且样方组较之单个样方更具稳定性，趋势分析结果也更可靠。尤其分组（合并）能降低 0 的出现几率，保证趋势分析的正常进行。因此，样方聚类是趋势分析的有机组成部分。样方聚类可分为三个步骤：数据标准化，计算相似系数；根据相似系数聚类。

#### (1) 数据标准化

MDSM 首先标准化数据，把多维空间中的点投影到多维超球面上。标准化投影的过程是通过各分量除以向量长度来完成的。向量长度等于各分量的平方和的算术根。定义如下：

$$P' = \frac{P}{L_{(p)}}$$

$$L_{(p)} = \sqrt{\sum P_{(i)}^2}$$

合并二式得：



$$P' = \frac{P}{L_{(p)}} = \frac{P_{(i)}}{\sqrt{\sum P_{(i)}^2}}$$

式中  $P$  (Plot) 代表样方,  $P_{(i)}$  是样方  $P$  的第  $i$  个分量,  $L_{(p)}$  是样方  $P$  的向量长度,  $P'$  是标准化后的向量,  $P'_{(i)}$  是标准化后的第  $i$  个分量, 在本文中暂时称之为物种  $i$  的重要值 (Importance Value);  $\sqrt{\quad}$  表示平方根,  $\Sigma$  表示求和, 上标 2 表示乘方。标准化后的向量长度都是一个单位, 同时保持各分量的相对比例不变。因为长度都是 1, 所以所有的标准化向量的端点都落在超球面上。

向量标准化的过程与结果可以用图来更清楚地显示, 如图 1。图 1 坐标有两套刻度; 0—100 用于原始样方数据, 而 0—1 用于标准化后的向量。两套刻度用于一图, 以便比较投影前与投影后的异同。三个点  $A$ 、 $B$ 、 $D$  的坐标是三个样方  $A$ 、 $B$ 、 $D$  的成分:  $A = (40, 80)$ ;  $B = (64, 77)$ ;  $D = (117, 53)$ 。 $B$  的向量长度为 1 可以免去讨论,  $A$  和  $D$  的向量长度分别为:

$$L_{(a)} = \sqrt{40^2 + 80^2} = \sqrt{8000} = 89$$

$$L_{(d)} = \sqrt{117^2 + 53^2} = \sqrt{16498} = 128$$

标准化的向量  $A'$  和  $D'$  分别为:

$$A' = A/L_{(a)} = (40/89, 80/89) = (0.45, 0.89)$$

$$D' = D/L_{(d)} = (117/128, 53/128) = (0.91, 0.41)$$

标准化的向量保持原来的组成成分

$$\begin{aligned} A' (\text{甲}: \text{乙}) &= 0.45:0.89 = 40/89:80/89 = 40:80 \\ &= A (\text{甲}: \text{乙}) \end{aligned}$$

$$\begin{aligned} D' (\text{甲}: \text{乙}) &= 0.91:0.41 = 117/128:53/128 = 117:53 \\ &= D (\text{甲}: \text{乙}) \end{aligned}$$

标准化后,  $A'$  和  $D'$  向量的长度为一个单位, 各分量的平方和为 1:

$$L_{(a')} = (40/89)^2 + (80/89)^2 = (1600 + 6400) / 8000 = 1$$

$$L_{(d')} = (117/128)^2 + (53/128)^2 = (13689 + 2809) / 16498 = 1$$

(2) 计算相似系数

图 2 显示标准化的向量  $A' = (0.45, 0.89)$ 、 $B' = (0.64, 0.77)$  和  $D' = (0.91, 0.41)$  在单位超球面上的相对位置。 $B'$  位于  $A'$  和  $D'$  中间，且距  $A$  较近。显然，当样方数量太大或坐标轴数目超过 3 时，人们无法单凭肉眼，而只有借助相似系数才能洞悉样方在多维空间中的相互位置及相互关系。MDSM 定义两向量夹角的余弦值作为两向量（一样方）的相似系数。相似系数的计算公式如下：

$$\begin{aligned} S.C. (a, b) &= \cos AOB \frac{A_{(i)} \times B_{(i)}}{\sqrt{\sum A_{(i)}^2} \times \sqrt{\sum B_{(i)}^2}} \\ &= \sum A'_{(i)} \times B'_{(i)} \end{aligned}$$

式中  $S.C.$  是相似系数， $AOB$  是向量  $A$  和向量  $B$  的夹角， $A_{(i)}$  和  $B_{(i)}$  是向量  $A$ 、 $B$  的分量， $A'$  和  $B'$  是标准化向量。MDSM 定义的相似系数是两个标准化向量的对应分量的积的和，向量夹角余弦值。因为标准化向量的长度是 1，以下计算单位中分母中的 1 被省略。

$$\begin{aligned} S.C. (a, b) &= \cos AOB = OF = (0.45 \cdot 0.64 + 0.89 \cdot 0.77) \\ &= 0.97 \end{aligned}$$

0.97 的反余弦  $\arccos 0.97 = 13 = AOB$ ， $A$  和  $B$  的夹角是 13 度。

$$\begin{aligned} S.C. (a, d) &= \cos AOD = OE = (0.45 \cdot 0.91 + 0.89 \cdot 0.41) \\ &= 0.78 \end{aligned}$$

0.78 的反余弦  $\arccos 0.78 = 39 = AOD$ ， $a$  和  $d$  的夹角是 39 度。

至此，余弦值的计算证明了肉眼观察的结果，即  $B$  距  $A$  较近，而距  $D$  较远。表明样方  $B$  和  $A$  更相似（较之样方  $D$ ）。

趋势

势分析 及其在生态股市中的应用



### (3) 归并样方

成分相似的样方投影相近，成分相异的样方投影相距较远(图2)。样方标准化或投影到超球面为样方聚类奠定了基础。不同的样方组代表不同的植被类型，其内部个体之间的同质性可用它们的相似系数来描述。如图3所示样方A、B和D投影到 $A' = (0.45, 0.89)$ 、 $B' = (0.64, 0.77)$ 、 $D' = (0.91, 0.41)$ 。 $B'$ 与 $A'$ 近，而距 $D'$ 远。因 $A'$ 和 $B'$ 可归并，从而形成一新的并组G。并组G的坐标是 $A'$ 和 $B'$ 的平均值：

$$A' = (0.45, 0.89),$$

$$B' = (0.64, 0.77)$$

$$G = (A' + B')/2 = [(0.45 + 0.64)/2, (0.89 + 0.77)/2] \\ = (0.54, 0.83)$$

标准化G：

$$G' = G/L(g)$$

$$= [0.54 / (0.54^2 + 0.83^2), 0.83 / (0.54^2 + 0.83^2)]$$

$$= (0.55, 0.84)$$

从图中可以看到，G是弦 $A'B'$ 的中心。一般来说，MDSM用众样方的几何形心(重心)作为样方并组的坐标，其数量值等于各样方坐标的平均值。

### 2. 趋势分析

至此，MDSM解决了三向数据 $D_{(s,p,t)}$ 在变量方向和样方方向上的问题；用植物种为坐标轴建立物种空间；通过投影、分组，用样方的平均值形成状态向量代表植被状态。以下开始讨论时间系列。对于时间系列1989—1993，定义1989年状态向量(State Vector) $A' = 89 = (0.45, 0.89)$ ，1993年状态向量 $B' = 93 = (0.64, 0.77)$ 。它们分别代表1989年和1993年的植被状态。如图4所示。一般来说，时间系列中的连续两点 $A'$ 和 $B'$ 在多维空间中的不重合，表明在这段时间中植被成分发

生了变化。从自然监测和趋势分析的角度，MDSM 要回答如下问题：(1) 植被成分发生了怎样的变化；(2) 每个植物种的变化速度多大；(3) 如果这个变化趋势保持不变，目前的植被能维持多长时间；或在下一个取样时间，MDSM 预期什么样的植被状况；以及其后的植被状况……即使植被变化的原因并没有被充分了解，我们仍然可以假定所有现存的状态条件足以用来预报未来的发展，一般，可以用方程式表述如下：

$$A' \cdot T = B'$$

其中  $A'$  和  $B'$  是 1989 年时间段和 1993 年时间段的状况向量。 $T$  是未知的多维变量，表示导致植被从  $A'$  状态转移到  $B'$  状态的未知因素的总和，然而，在一般情况下，当  $T$  取矩阵形式且是对角矩阵时，状态转移矩阵才有确定的解。

回顾 MDSM 的前提假定：所有的物种都是独立、互不相关的，则非对角元素为零，状态转移矩阵可以用向量、趋势向量来表示：

$$A' \cdot T = B'$$

$$t = B' / A'$$

定义：趋势向量的分量是后前两状态向量的对应分量的比。趋势向量  $T_{(i)}$  的元素是相应植物种在时间系列里不同的重要值的比，其期望值为 1，表明相应的植物种在植被中的重要值随时间的变化。分析  $T$  值（在以下的讨论中我们用  $T$  表示趋势向量），可以了解每个植物种及整个植被在监测期间的变化。进一步，趋势向量  $T$  可以作用于状态向量  $B' = 93$ ，以估计未来的植被状态。MDSM 定义预报向量的分量是趋势向量和状态向量相应分量的积：

$$B' \cdot T + R = C = 1997$$

$$B' \cdot T^2 + R_2 = D = 2001, \dots$$

$$B' \cdot T^{k-1} + R_{k-1}$$

趋势

## 趋势分析 及其在生态股市中的应用



此处  $R$  是预报误差，时间间隔是 4 年， $k$  是预报的重复次数。

用图显示可能会更清楚。在图 4 中， $A' = 89$  (0.45, 0.89) 是 1989 年的状态向量； $B' = 93$  (0.64, 0.77) 是 1993 年的状态向量。连线  $A'—B'$  的斜率用来表示趋势向量  $T$ ，表明植被在监测期间的变化率。将  $A' = 89$ 、 $B' = 93$  的值代入  $T_{(i)}$  的定义式：

$$\begin{aligned} T &= B' / A', \\ T_{(i)} &= 93_{(i)} / 89_{(i)} \\ &= (0.64/0.45, 0.77/0.89) \\ &= (1.42, 0.865) \end{aligned}$$

趋势向量的值表明物种甲从 0.45 增加到 0.64，增加比率为 1.4；而物种乙从 0.89 下降到 0.77，下降比率为 0.86，趋势向量  $T$  作用于状态向量 93，得到第一个预报值  $C = 97$ ：

$$\begin{aligned} C = 97 = 1997 &= B' \cdot T = A' \cdot T^2, \\ A' \cdot T^2 &= (0.45 \cdot 1.42^2, 0.89 \cdot 0.865^2) = (0.91, 0.67)。 \end{aligned}$$

向量  $C = 1997$  标准化得到  $97'$ ：

$$\begin{aligned} 97' = 1997' &= 97 / L_{(97)} 97_{(i)} / L_{(97)} \\ &= [0.91 / (0.91^2 + 0.67^2), 0.67 / (0.91^2 + 0.67^2)] \\ &= (0.81, 0.59) \end{aligned}$$

趋势向量再作用一次，得到  $D = 2001$ ：

$$\begin{aligned} D = 2001 &= B' \cdot T = A' \cdot T^3 = (0.45 \cdot 1.42^3, 0.89 \cdot 0.865^3) \\ &= (1.29, 0.58) \end{aligned}$$

$D = 2001$  标准化得到  $2001'$ ：

$$\begin{aligned} 2001' &= [1.29 / (1.29^2 + 0.58^2), 0.58 / (1.29^2 + 0.58^2)] \\ &= (0.91, 0.41) = D' \end{aligned}$$

1997' 和 2001' 分别代表 1997 年、2001 年的植被状况的标准化的估计值。原则上线段  $89'—93'$  可沿指数函数曲线延长到

和某一坐标轴逼近。从现在到与坐标轴逼近这段时间  $k$  可以用于表明现阶段植被利用是否合理。图中估计线用发散的虚线表示。发散表示非直线，虚线表示有误差。

## 二、一个试运行

表 2 显示 MDSM 用实际数据对美国卡森堡植被条件试行趋势分析的初步结果。MDSM 使用相关分析结合野外工作经验从 188 个植物种中筛选了 35 个植物种，建立 35 维变量空间。199 个样方是空间中的 199 个点。数据样方取自 1989 和 1993 年“土地条件趋势分析 (LCTA)”野外样方数据库。为了突出趋势分析，所有的 199 个样方合并为一组，表示整个卡森堡的植被状况。表 2 的第一列是 35 个植物的拉丁名缩写。为了强调模型的效果，试运行包括了几种一年生植物。第二列代表 1989 年植被的状态向量，是 199 个样方的平均值的标准化后的值，也即 1989 年群体在多维空间的几何形心在单位超球面上的投影。第四列是 1993 年的状态向量。两列中间夹着的第三列是趋势向量。如以上讨论， $T$  值大于 1，表明相应植物种的成分在增加；反之，小于 1，表明下降。例如第 9 个  $T$  值， $T_{(9)} = 6.1$ ，表明相应的植物种 BRJA 在监测期时增加了 6 倍。而  $T_{(6)} = 0.94$  表明植物种 6，BOCU 基本没有变化。第五列是预报的 1997 年的状况，第七列是预报的 2001 年的状况。第六列、第八列分别是标准化的预报的 1997、2001 年的植被状况。表 2 横行显示植物种变化。以第一行物种 AGROP 为例， $T_{(1)}$  值是 1.61，表明 AGPRO 的分量在上个监测期间有增加。根据预报值，到 2001 年将增加到 0.32。但是根据重要值，则仅达 0.06。同一物种有两个不同的预报值的原因在于 MDSM 不仅能





计算到每个单独植物种的变化,分量的变化,而且要计算到整个植被的变化,反映向量长度的变化。根据表 2,考察整个植被的变化和其他植物种的变化,当物种 20, MEOF 从 1989 年的 0.004 上升到 2001 年的 0.86 和 BRJA 从 0.01 上升到 0.40 时,物种 AGROP 的分量则相对减少,重要值相对下降。由于标准化向量的长度是一个单位,所有分量的平方总和为 1。一个植物种增加,其余植物种必将相对减少。资源有限,此消彼长。当 MEOF 取值 0.86、BRJA 取值 0.4 时,为其他物种所余空间便很有限了。试运行包括几个一年生植物,就是为了显示 MDSM 的这个性质。因为一年生植物对当年的降水量敏感,变化显著,更易于说明各分量之间的变化关系。但是,一般来说,一年生植物不宜用于长期的植被演替趋势分析。

回到卡森堡的例子,由于绝大多数的多年生植物和主要优势种的  $T$  值接近 1, MDSM 试运行的结果表明,卡森堡的植被状况在 1989—1993 年监测期间没有发生显著的变化。

### 三、结果与讨论

MDSM 以互相垂直的坐标轴表示植物种,以空间的向量表示样方。通过向量标准化,把空间的实体投影到超球面上,使标准化的向量长度为一个单位,不但使多维空间变得可以捉摸,便于操作,而且比较好地刻画了植物种和植被的关系:物种独立,资源共享。然后,在多维空间中,用两向量夹角的余弦值为相似系数表示空间实体的相对位置:在超球面上,用多边形的几何形心代表多边形;用植被在多维空间运动的轨迹表示植被演替趋势;分析趋势以估计未来的植被状态。MDSM 在数学上是可行的,用 LCTA 数据进行试运行,效果也比较好。

使用 MDSM 的前提条件是所有植物种相互无关。虽然我们不知道,生态系统中所有的植物种都具有某种程度的相关。问题是多大程度的相关,是否大到影响植物种的独立性。在三向数据  $D_{(s,p,t)}$  中,与时间的变化、样方的变化相比,物种的变化是较为绝对的。不同的植物种就是不同的数据,而不同时间的数据可以是相同的。1989 年到 1993 年的植被变化可以是零。样方也是一样,取自同质群落的不同样方具有差别等于零的倾向。因此, MDSM 以变量为坐标轴建立变量空间是应该被接受的。在其他以多维空间为基础的模型中,一般以实体间的距离、欧氏距离或弦距离来表示实体之间的差别,而 MDSM 使用的余弦值夹角来度量实体之间的差距。实际应用中, MDSM 使用夹角余弦值度量相似程度,数学上可行,实际效果较好。余弦的取值范围在 0—1 之间。当两实体完全相同时,  $\sum A_{(i)} = \sum B_{(i)}$ , 相似系数值取最大值,  $S.C = 1$ 。而当两实体完全不同,没有共同成分,  $\sum A_{(i)} \cdot \sum B_{(i)} = 0$  时,相似系数取最小值,  $S.C = 0$ 。

MDSM 合并相近的空间实体,用它们的几何形心代表它们。物理学上是重心,代数统计上是平均值,向量加法形成植被单位,然后以植被单位进行趋势分析。设想:如果不用植被类型,而是引入某个控制因子或处理,比如以坦克的碾压程度——重、中、轻来组分样方,然后按组分别进行趋势分析。比较其结果,则可能提示不同控制因子、处理,在此是坦克碾压对植被演替的影响。

MDSM 被用于多变量分析,要求纳入尽可能多的变量。增加一个具有小刻度的坐标轴,不会很大地影响分类结果,但却增加了一个趋势分析的坐标轴,即增加了一个信息来源。对于植被演替趋势分析,使用尽可能多的植物种是重要的。数学模型是对真实植被的模拟,越多的植物种被选用,模型越逼近真

**趋** 势分析 及其在生态股市中的应用



实。而且选用越多的种，种对间的相互作用将相对减弱，种的独立性加强，使分析的结果更可靠。

对于互相垂直的不同的坐标轴，MDSM 允许使用不同的度量，可具有不同的单位、刻度。如用密度表示羊草，而用盖度表示针茅……在不同的坐标轴上使用不同的刻度，不会影响趋势分析；但是由于将影响实体间的相互距离和角度，所以将影响分类。反过来，对于有些研究，在使用 MDSM 时，我们可以给某个植物种加权，如树木，夸大它的作用，伸长它的坐标轴刻度（以弥补二维取样用于三维植被存在的不足），以便取得较好的分类效果，而同时又可以保持趋势分析的结果不变。

$T$  值，转移向量的各分量，并不是相应各变量的简单的比，而是标准化后的值（重要值）的比：

$$\begin{aligned} T &= B'/A' = [B/L_{(B)}] / [A/L_{(A)}] \\ &= [B \cdot L_{(A)}] / [A \cdot L_{(B)}] \end{aligned}$$

$$T_{(i)} = B_{(i)} / A_{(i)} \cdot L_{(A)} / L_{(B)}$$

如定义式所示，趋势向量的值是当前的变量值和前一个时间段的变量值的比，与它相应的向量长度的反比是乘积。换一个说法。 $T$  值是物种的重要值的变化率和植被的变化率的比。如果物种重要值上升的幅度大于整个植被上升的幅度，则  $T_{(i)}$  大于 1。如果物种的重要值虽然上升，但上升的幅度小于植被的上升幅度，则  $T$  值小于 1。反之亦然。 $T$  值中不仅包含变量本身的变化信息——后前不同时间段的重要值，而且包含整个植被状况的信息——后前不同时间段的向量长度，也即整个植被的状况。通过分析状态转移向量，人们可以得到比较准确、全面的植被变化的信息。

在我们所使用的例子里，取样时间间隔为 4 年，估计间隔也为 4 年：1997，2001……事实上，当取样时间间隔为 4 年，预报时，时间间隔  $k$  可为分数，从而使模型更灵活。比如对

于同样的 LCTA 数据,  $T = 93'/89'$ 。若令  $k$  的增量 =  $1/4 = 0.25$ , 我们可以估计 1989 年以后任何一年的植被状态。

$$90' = 89' \cdot T^{0.25}$$

$$91' = 89' \cdot T^{0.5}$$

$$92' = 89' \cdot T^{0.75}$$

$$93' = 89' \cdot T^1$$

$$94' = 89' \cdot T^{1.25} \dots$$

或, 反之, 使  $T = (93'/89')^{(0.25)}$ , 1989 年到 1993 年变化率的 4 次方根, 则  $k$  可取正整数。

当趋势分析时, 随  $k$  值增加, 预报值愈来愈远离初始值。当  $k$  值大到一定程度, 使某年的预报值  $P_{(k,i)}$  趋近于 0, 或  $P_{(k,i)}$  趋近于 0.7 时, 趋势分析可以停止。生物学解释: 因为  $P_{(k,i)}$  趋近于零, 意味着一个植物种从植被中消失; 而  $P_{(k,i)}$  趋近于  $0.7^2 = 0.49$  植被演替达到了一个质变的新阶段。而比较此时的  $m$  值, 可作为植被演替的速率参考。

MDSM 的局限: 趋势分析的核心是趋势向量的定义。趋势向量的元素, 植物种的  $T$  值, 是后前两个时间段状态向量  $89'$ 、 $93'$  的比值。因此状态向量的元素不可为 0, 否则, 比值无意义。状态向量元素为 0 可以有 3 种情况: (1)  $0/0$ , 可以省略讨论; (2)  $0/89'$ , 表明一个物种在监测期间消失了; (3)  $93'/0$ , 表明一种植物种无中生有。如上段的讨论, 两种情况都表明植被在两次取样期间已经发生了质的变化, 进行趋势分析已属多余。如果这种质变, 物种的消失, 或出现仍不足以说明问题, 则这个物种可以被忽略。

另一个局限, 图中发散的虚线表示估计线。表示估计有误, 而且估计差随  $k$  值的增加而增加。状态向量来自取样, 都有取样误差。状态向量的误差被传递到趋势向量。当估计未来状态趋势向量被乘幂时, 其中的误差也被乘幂。也许乘/除



法模型仅适用于长期预报。有文献指出，提高预报精度的方法是配合预报进行取样，把预报的数值和取样的数值加权平均，这即所谓的卡门滤波，我们在这里不讨论。

MDSM 用状态向量乘趋势向量的乘幂估计未来的状态。估计值所构成的线不是直线，只有在转移向量的元素的值  $T$  接近 1 时，估计值曲线才接近于直线。 $T$  值接近 1，表明 MDSM 仅适用于被监测的植被处于稳定状态，没有剧烈的成分变化。实际上，这也正是趋势分析所要解决的。如欲将 MDSM 用于剧烈变化的系统，要求更准确的预报结果，是否可以考虑使用加法模型。在加法模型中，预报值不是转移向量乘状态向量的乘幂，而是状态向量加增量向量。而增量向量来自后前两个时间段状态向量差。若  $A$ 、 $B$ 、 $C$  为三个连续状态向量，则

$$C = B + (B - A) = 2B - A$$

其中， $B - A$  是增量向量。关于乘法模型  $C = B \times B/A$  和加法模型  $C = 2B - A$  的比较，我们将在以后的文章里进行讨论。

(《中国草地》1966 年第 2 期)

## 超球面模型 应用于股票排序的探讨

**摘要:** 本文探讨把超球面模型 (MDSM), 一种基于定义了向量除法的多维空间的数据分析技术, 应用于股票排序的可能性。MDSM 把股票市场作为多维指数增长系统, 用多元向量表示股市状态, 并定义余弦向量的商为多维即时趋势, 以表示股票市场的动态, 用即时趋势排列股票的结果与使用股票回报率的排序一致。

**关键词:** MDSM, 多维空间, 多元向量, 即时趋势商高指数, 系统监测, 多维指数方程

**Abstract:** The Multi-Dimensional Sphere Model (MDSM), a new data analysis method based on multivariate space ( $m$ -space), is applied to stock market data for temporal dynamic analysis. MDSM uses multi-component vectors ( $m$ -vectors) to express the state of the market. The rotation of the vector in the  $m$ -space is used to express the dynamics of the market. The quotient of the cosine values of the vectors is defined as Multivariate Instantaneous Trend (MIT), and the stocks are ranked by their trend values.

**Key words:** MDSM,  $m$ -space,  $m$ -vectors, multivariate instantaneous trend, ShangGao Index, System monitoring, multivariate exponential equation



## 引言

超球面模型 (MDSM) 是为了资源监测设计基于定义了向量除法的多维空间 ( $m$  空间) 的数学方法, 可以用于系统的趋势分析和动态监测, MDSM 用  $m$ -空间的点 ( $m$  向量) 表示系统的状态, 向量有长度和方向, MDSM 用向量的长度表示系统的综合状况 (量, quantity), 而用向量的方向表示系统的组成 (质, quality), 由于定义了向量的乘法和除法, MDSM 能够处理系统的指数增长, 比较准确地描述多维指数增长系统, 股票市场被认为是典型的多维指数增长系统, 本文提出利用超球面模型, 使用即时趋势进行股票排序的方法。

### 1. 传统的方法

股票按股票名称排列, 便于客户查找, 显示收盘价, 列出增幅, 用差来表示变化便于股票持有者计算增益亏损。指数 (如: 道·琼斯指数) 产生于若干预选的被认为有代表性的股票, 随着时代的变迁, 地域的不同, 人们需要设计不同的指数, 或更换成份股。而各指数之间不能/不易转换。每支股票有多项纵的统计量, 但缺乏横向比较。

### 2. MDSM 建议的增补改进的方法

(1) 指数的计算使用全体股票的信息 (商高指数 =  $m$  元向量长度)。

(2) 股票的变化用经过调整的增率 (即时趋势) 来表示。

(3) 以增率 (即时趋势) 来进行横向比较, 并排列股票。

优点: 客户可以一目了然地了解整个股市的变化以及给定的股票相对于其他所有股票的表现。即时趋势具有可比性, 且可累积为长期趋势。由于 MDSM 用序位显示数据分析结果, 可

能比较容易为大众传媒所接受。

### 3. 具体做法

#### (1) 定义 $M$ 维空间

选择  $M$  个股票（假定深圳的上市股票数为 275，则  $M = 275$ ），记录它们的收盘价，得到 275 个变量值（也即 275 元向量），用 275 支股票的价格来表示股票市场的状况。

表达式： $Y_{i,k} = (Y_{1,k}, Y_{2,k}, \dots, Y_{275,k})$  或  $Y_{i,k}, i = 1, 2, \dots, 275$ ，下标  $i$  表示股票，下标  $k$  表示时间。

#### (2) 计算商高指数

计算 275 个股票价格的平方和的算术根（向量长度），命名为商高指数（Shang Gao Index）。商高指数是一个数（标量），表示股票市场当天的综合状况。（以商高来命名是为了纪念发现勾股定理的中国古代数学家商高）

公式： $SGI_k = (Y_{i,k})^2$ ，其中  $i = 1, 2, \dots, 275$ 。

#### ① 计算商高率

商高指数的变化比率为商高率（SGI%）。商高率表明股市的变化（275 支股票的综合变化）。商高率大于 1，整个股市有增长；小于 1，整个股市在下降。

公式： $SGI\%_k = SGI_k / SGI_{k-1}$ ，其中  $k$  表示当天， $k-1$  表示前一天。

#### ② 计算股票率

每支股票的变化比率股票率，得到 275 个股票率。股票率等于 1，表示该股票没有变化；大于 1，有增长；小于 1，在下降。

公式： $Y\%_{i,k} = Y_{i,k} / Y_{i,k-1}$ ， $i = 1, 2, \dots, 275$ 。其中， $k$  表示当天， $k-1$  表示前一天。

#### ③ 计算趋势值

用商高率来标定（除）每个股票率，标定后的股票率称为



## 趋势分析 及其在生态股市中的应用





多维即时趋势 (Multivariate Instantaneous Trend), 从而得到 275 个即时趋势值。即时趋势值表明该股票相对于股市的变化。它的期望值是 1。如果股票 A 的增率是 106%, 而同时整个股市的增率是 107%, 则股票 A 的即时趋势值是 0.9963, 表明该股票是在相对下降, 而不是在上升。

公式:  $T_{i,k} = Y\%_{i,k} / SGI\%_k, i = 1, 2, \dots, 275$

#### (5) 用趋势值将股票排序

用 275 个即时趋势值对股票进行横向比较, 并排列 275 支股票。即时趋势值大, 增率高的排在前面, 即时趋势值小, 增率低的排在后面, 使客户对整个股市所有股票的动态一目了然。

#### 4. 一个实例 (数据抄录自 2 月 21 日、22 日洛基山报 (Rocky Mountain News, <2/21&2/22, 1997>))

令  $m = 5$ 。从报纸录取 5 支在纽约股票交易所上市的股票。选择的股票是 2 月 22 日分别在 A, B, C, D 和 E 部中增幅最高的股票。表头中, 股价是当天的股票价格, 2/21 和 2/22 分别表示 2 月 21 日和 2 月 22 日。增量是连续两天股价的差, 增量 = 股价 (2/22) - 股价 (2/21)。增率 (%) 是连续两天股价的商, 增率 (%) = 股价 (2/22) / 股价 (2/21)。最后一行是有关商高指数的数据。股价和增量的单位是美元, 取两位小数。增率用三位数表示, 趋势值取四位小数。

表 1 用实例计算趋势值, 并以趋势值排列股票。算法、定义如 3。具体做法如下:

变量名	股价(2/21)	股价(2/22)	增量	增率%	趋势值 T
CentenTs	2.00	2.50	+ 0.50	125	1.1704
Echlin	31.00	33.38	+ 2.38	108	1.0081
AberFit	15.63	16.63	+ 1.00	106	0.9963

变量名	股价(2/21)	股价(2/22)	增量	增率%	趋势值 T
Dimon	24.50	26.00	+1.50	106	0.9937
Banctec	21.25	22.50	+1.25	106	0.9914
商高指数	47.549	50.782		107	

表2(见下页)是模拟投资。比较表1、表2,可以看出趋势值与投资回报率(投资效率)的排序是一致的,都是C, E, A, D, B。趋势值高者投资效率也高。根据趋势值定义、以上结论不仅在五维空间成立,在任意多维空间也成立。

表2模拟投资。并以投资效率排列股票。模拟投资每股100美元,所以投资效率可简单地以产出来表示。

股数 = 投入/股价<sub>k-1</sub>, 产出 = 股数 × 股价<sub>k</sub>

	投入 \$	股价(2/21)	股数	股价(2/22)	增量	产出 \$
CentenTs	100.00	2.00	50.00	2.50	+0.50	125.00
Echlin	100.00	31.00	3.23	33.38	+2.38	107.66
AberFit	100.00	15.63	6.40	16.63	+1.00	106.40
Dimon	100.00	24.50	4.08	26.00	+1.50	106.12
Banctec	100.00	21.25	4.71	22.50	+1.25	105.88

以上分析表明,股票市场的投资效率(产出投入比)完全取决于价格比(即时趋势),而与价格差(增量)没有直接联系。

## 5. 背景介绍

MDSM 是介于向量分析和投影变换之间的新的多元分析方法。按照多媒体 1995 版 GROLIER 百科全书的说法,“向量分析”是处理既有量值又有方向的变量的数学分析工具。与传统

**趋势分析** 及其在生态股市中的应用



的向量分析 (Davis, 1979), (Crowe, 1967) 不同的是 MDSM 把向量分析从三维扩展到了任意多维, 并引入向量除法。而与向主分量投影的方法 (Gauch, 1982) 比较, 其不同之处在于, MDSM 实行两次投影来保持所有分量。首先, 通过标准化, MDSM 把多维空间中代表股市状况的点投影到单位超球面上。然后, 从超球面上分别向所有的  $m$  个坐标轴投影。这样, 整个分析过程始于  $M$  元向量, 终于  $m$  元向量, 实现了全息投影。经过投影, MDSM 用  $m$  元余弦向量来表示股市的状态 (状态向量, State Vector)。对于线性系统, 有了余弦, 就有可能计算正弦和余切, 可以分别用来表示差别和变化率。但由于 MDSM 假定股票市场是多维指数增长, 因此它不用切线, 而直接定义状态商来表示变化趋势。具体到股市分析, MDSM 用超球面上的点, 或余弦向量, 代表股市状态。并定义状态商 (状态向量对应分量的商) 为即时趋势; 用  $m$  元趋势向量来表示股市动态。由于  $m$  元向量是  $m \times 1$  矩阵, 可被视为一维里的  $m$  个数值, 因此可以在一维里排序。MDSM 根据即时趋势值, 将 275 支股票排序, 实行横向比较。

根据在上文 3. 具体做法中交代的计算公式, 即时趋势值 ( $T$ ) 是用商高率 (SGI%) 调节后的股票率 ( $Y\%$ ):

$$T_{i,k} = (Y_{i,k} / Y_{i,k-1}) / (SGI_k = SGI_{(k-1)}), i = 1, 2, \dots$$

275,  $k$  表示当天,  $k-1$  表示前一天。

同时, 即时趋势的定义式是后前状态的比, 即时趋势值 ( $T$ ) = 状态<sub>(k)</sub>/状态<sub>(k-1)</sub>, 而状态用余弦向量表示:

$$T_{i,k} = (Y_{i,k} = SGI_k) / (Y_{i,k-1} = SGI_{k-1}), i = 1, 2, \dots,$$

275,  $k$  表示当天,  $k-1$  表示前一天。

也即, 多维即时趋势是状态向量 (多维超球面上的点) 在各坐标轴上后前投影 (余弦) 的比。超球面模型的名称由此而来。由于即时趋势值包含了股票自身的变化和所有其他股票的

变化，它反映着相应股票在股市中份额的变化。同时，它还显示，一元钱的相应股票经过一个交易日后，扣除市场因素后的价值。而中长期趋势可以是即时趋势的简单的积累。如5日连乘积是周趋势，20日连乘积是月趋势，等等。另一方面，由于即时趋势的时间段是任意的，可以是天、周、月、年，所以趋势分析相当灵活，可以应用于各种不同的场合。

表3从2/20到2/22的双日趋势是21日和22日两个即时趋势的连乘积： $T_{20-22} = T_{21} \times T_{22}$ 。表中，Y，T和%的定义同上，分别表示价格、趋势和增率。下标是日期。趋势值取4位，价格取2位小数。

	Y <sub>20</sub>	Y <sub>21</sub>	%21/20	T <sub>21</sub>	Y <sub>22</sub>	%22/21	T <sub>22</sub>	T <sub>21</sub> × T <sub>22</sub>	%22/20	T <sub>20-22</sub>
AberFit	15.25	15.63	102	1.0117	16.63	106	0.9963	1.0079	109	1.0079
Banctec	21.13	21.25	101	0.9933	22.50	106	0.9914	0.9847	107	0.9847
CentenT	2.00	2.00	100	0.9874	2.50	125	1.1704	1.1557	125	1.1557
Dimon	24.00	24.50	102	1.0080	26.00	106	0.9937	1.0016	108	1.0016
Echlin	30.75	31.00	101	0.9954	33.38	108	1.0081	1.0035	109	1.0035
SGL	46.95	47.55	101		50.78	107			108	

上文4.部分已用实例显示，用即时趋势值所排的序列和投资回报率（投资效率）的序列是一致的。同时，根据趋势值的定义可以导出趋势值的序列是唯一确定的。也即，两个股票A和B的排序，在m支股票的股市和在m+1支股票的股市中是一致的。更进一步，MDSM所定义的多维即时趋势不仅引出了向量除法的定义，而且推导出多维指数增长方程。后者可被用来描述和监测股票市场：

$$Y_{i,k+x} = \frac{\alpha \cdot Y_{i,k} \times T_{i,kx} + \beta \cdot D_{i,k+x}}{\alpha + \beta} \quad i = 1, 2, \dots, m, 10$$



式中,  $x$  是时间自变量,  $Y$  是股市状况,  $T$  是即时趋势,  $D$  是取样值,  $\beta$  和  $\alpha$  分别是预报值和取样值的权。虽然当差别小到接近于 0, 且状态商接近于 1 时, 指数增长和线性增长没有太大的分别, 但多维指数增长方程更易于描述股票的大起大落, 也与利滚利, 息生息, 人跟人的股票市场规律相吻合。“向量除法”和“多维指数增长方程”似乎在传统数学中仍属于处女地, 这既是推广 MDSM 的阻力, 同时也是 MDSM 的潜力和希望之所在。

本文是向母校 40 周年校庆的献礼。作者在研究过程中得到中国科学院应用数学研究所陈培德研究员的指点和科罗拉多州立大学数学系博士生马建敏的帮助。初稿曾送资深院士阳含熙审阅, 特此致谢。

(1999 年 5 月发表于《内蒙古大学学报》30 卷 3 期)

# 超球面模型及其 在股市分析中的应用

——金融数学的新思考

**摘要：**超球面模型是从资源监测发展起来的基于向量的多元分析方法。向量是既有量值又有方向的变量。超球面模型把传统的向量分析从三维扩展到任意多维，可以同时处理任意多个变量。超球面模型定义了向量除法，可以处理指数增长，因此特别适用于生物、金融、信息、经济等领域的变量。应用于基金市场时，模型以基金为坐标轴建立多维基金空间；用多元向量表示基金市场；以向量长度做指数表示市场的量值；以向量的方向表示市场状态；并用后前余弦的商表示市场的动态，称多维即时趋势；进而根据趋势值将基金排序，根据基金的序位决定售购。在7个月（1998.6.19—1999.1.29）的投资实验中，超球面模型指导的个人退休基金的增率是25.82%。这个增率不仅超过了20个基金市场的平均增率（2.64%），而且超过了市场中最好的基金的增率（18.98%）。目前，尚不能用传统的风险或投机的概念来解释MDSM的高增率。超球面模型假定变量无关，所以它的应用可以被扩展到任意多支基金。

**关键词：**多维空间，多元向量，指数增长，向量分析，超球面模型，多维即时趋势，多元分析，多变量时间系列

**Abstract:** The Multi-Dimensional Sphere Model is a new multivariate

趋势分析 及其在生态股市中的应用



data analysis method based on extended vector analysis derived from natural resource monitoring. A vector is a quantity stating both a magnitude and direction. MDSM extended the classic vector analysis from 3 dimension to multidimension, so it can handle more than three variables simultaneously. MDSM defined vector division, so it can manipulate the exponential growth. Thus, it is suitable for variables in sciences of biology, financing, information, and economics. When applied to mutual fund data, MDSM builds a multidimensional space with funds, expresses the market with a multicomponent vector. It expresses the quantity of the market with vector length, and expresses the state with direction. It uses the quotient of cosine values, present over previous, to express the changing trend of the funds, and trade them based on their trend values. In a seven-month test, an account directed by MDSM returned 25.82%, not only higher than the average of the 20 involved funds (2.64%), but also better than the best individual fund performance (18.98%).

**Key words:** MDSM,  $m$ -space,  $m$ -vectors, exponential growth, multivariate exponential equation, vector analysis, Trend, multivariate analysis, competitive resource management.

## 一、前言

超球面模型 (Multi-Dimensional Sphere Model, MDSM), 是从资源监测发展起来的新的多元分析方法。超球面模型以多维空间 (Multi-dimensional Space,  $m$ -space) 为基础, 使用多元向量 (Multi-component Vector,  $m$ -space 的夹角来表示分析对象之间的关系。本文所指的多维空间,  $m$ -空间是多变量空间,

与样本空间不同，它是线性空间的扩展，是定义了向量乘法和向量除法的空间。我们用  $m$ -空间表示，以便与样本空间线性和空间区别。超球面模型是从对角矩阵衍生出来的数学工具。从本质上说，超球面模型是扩展了的向量分析 (Vector Analysis)。根据 Grolier 百科全书 1999 年光盘版，“向量分析”是数学的一个分支，从属于分析 (Analysis)。而分析与统计、几何、代数、微积分等并列，从属于数学 (见图 1)。据美利坚百科全书，Encyclopedia Americana，1999 网络版，向量分析最初是由 Stevin (1548—1620) 开创，并最后由 Gibbs (1839—1903) 归纳综合，奠定基础。经典的向量分析一般被限于三维。Gibbs 生前曾经设想，他的向量分析可以被扩展到任意多维，但由于当时生产力水平的局限，他的这一想法没有得到推广和发展。近年来，由于资源管理、环境科学的发展，人们对植被分类问题、系统内部资源分配问题以及多元系统动态监测问题的研究，促进并扩展了传统的向量分析和多维空间的研究，产生了超球面模型。超球面模型把传统的向量分析从三维空间扩展到任意多维空间，可以同时处理任意多个变量，大大地扩大了向量分析的应用范围。经典的向量分析中，向量的方向基本上被忽略，或者说没有独立的特定的意义，而仅是用来进行量值运算，如计算平行四边形的面积，对角线的长度等。甚至向量的定义，也往往不是很确切的，向量的一个重要特性——方向往往被抹杀了。例如，有的数学教科书定义向量为“ $N$ 元数组  $n$ -tuples”，而另外一些书，特别是计算机方面的书，却定义向量为“一维数组，one-dimensional array”。在本文中，我们取 Grolier 百科全书的定义：“向量是不但有‘量值’ (Magnitude)，而且有‘方向’ (Direction) 的变量。A vector, in mathematics, is a quantity stating both a magnitude and a direction.” 或美利坚百科全书的说法：Vector Analysis, a branch of mathematics

趋势分析 及其在生态股市中的应用





that deals with quantities having both magnitude and direction. Such quantities are called vectors.”与经典的向量分析侧重于量值比较，MDSM更侧重于角度。在做系统的动态分析时，MDSM认为系统的状态向量（State Vector）的方向携带着系统组成的重要信息，因而人们可以通过状态向量在多维空间的指向的变化来感知和度量系统组成的变化。经典的向量分析没有定义除法，仅适用于线性空间（Linear Space）。而MDSM定义了向量除法，可以处理指数增长（又称几何增长），因此特别适用于生物、金融、信息、经济等领域的有自我复制（Replicate）功能的变量。

超球面模型利用向量除法推导出了多元指数方程（Multivariate Exponential Equation）。方程描述了初始值向量  $Y_{(i,0)}$ 、即时趋势向量  $T_{(i,0)}$ 、时间  $k$  和终值向量  $Y_{(i,k)}$  之间的关系，是我们认识多维指数增长规律，进而监测、指导国民经济可持续发展的有潜力的工具。

$$Y_{(i,k)} = [\beta \times Y_{(i,0)} \times T_{(i,0)}^k + \alpha \times D_{(i,k)}] / (\alpha + \beta)$$

其中， $\alpha$  和  $\beta$  分别是观察值， $D_{(i,k)}$ ，预报值  $Y_{(i,0)} \times T_{(i,0)}^k$  的权； $Y_{(i,0)}$ 、 $T_{(i,0)}$ 、 $Y_{(i,k)}$ ，和  $D_{(i,k)}$  是  $m$ -向量，用黑体表示， $i=1,2,3,\dots,m$ 。时间下标 0 和  $k$  分别表示初始和终止时刻。

多维指数方程的推导过程大体上可以分为三个步骤，如图 2 所示：

1. 在指数增长的通式上增加一个下标  $i$ ，表示变量，从而使公式的适用范围从一维扩展到多维；
2. 用经验增率  $T_{(i,k)}$  代替内禀增率（Intrinsic Rate）和常数  $e$  的组合， $e^r$ ，使这个核心参数可以从已知的数据中产生，并命名为多维即时趋势（Multivariate Instantaneous Trend, MIT） $T$ ；
3. 用卡门滤波（Kalman Filter）连接历史与现实。

本文不讨论多维指数方程的展开及其应用，而主要讨论利用多维指数方程的核心参数“多维即时趋势” $T_{(i,k)}$  进行投资

指导的原理，并报告一个投资实验。MDSM 是基于  $m$ -空间和  $m$ -向量分析的新的多元分析工具，在分析——综合多元数据方面有很大的潜力和广泛的应用。它曾经被应用于植被分类、土地条件趋势分析、植被演替趋势分析、反随机检验和股票排序，等等。目前，正在美国农业部的农牧场模型（GPFARM）中被用来作“反应检验”（Response Analysis）。当然，最令人瞩目，最引人兴趣的还属下面将要报告的基金投资实验。如上所述，MDSM 原本是为资源监测研制的，处理多变量时间系列的数学工具，而我们的实验表明，它也可以被应用于证券市场。由于资本有自我复制功能，利息可以产生利息（相当于细胞分裂），证券市场可以作为多维指数增长时间系列来处理。在广义上，证券市场宏观管理的问题是个资源分配的问题。它可以被归纳为：如何善用市场机制，吸引并‘优化配置’民间闲散资金给上市企业，使经营好的企业多得，经营不好的企业少得或不得，Competitive Resource Management。由于证券市场的数据较植被数据更精确、变量多、历史长、数据更新周期短、出结果快且容易验证，成本低，实验结果统一用货币单位‘元’来表示，一目了然，更容易为普通读者所接受。因此，自 1995 年 MDSM 诞生以来，我们一直在使用证券市场的数据来检验  $m$ -空间理论，完善超球面模型。自 1998 年，进一步用个人退休基金（Individual Retirement Account, IRA）实际投资。因个人退休基金已经有账户和资金，而且没有手续费，简化了程序。现将投资实验的原理和结果报告如下。

趋

势分析 及其在生态股市中的应用



## 二、超球面模型应用于金融 市场监测的基本原理

超球面模型原本是为多元分析设计的，它可以同时处理多个证券，但是为了更好地说明问题，首先让我们从最简单的例子，两个证券开始讨论。假定，我们有两个证券，分别用  $X$  和  $Y$  来表示，并用直角坐标系  $XOY$  来表示这两个证券所组成的市场。证券  $X$  和证券  $Y$  分别是两个坐标轴，原点是  $O$ ，单位是元。我们并假设，实验开始时的价格为： $X$  证券每股 3 元， $Y$  证券每股 4 元。MDSM 用二维空间的点  $A_0 = (3, 4)$  来表示这个证券市场的初始状态，向量  $OA_0$  称市场在时间 0 时的状态向量。

这里需要说明的是，在金融科学中，股票与基金是两个不同的概念，分属于不同的投资工具，有不同的性质和操作规律。但在本文的讨论中，我们把基金股票统统作为证券变量，或系统的分量，而把股市、基金市场作为多变量系统，来研究讨论系统和分量之间的关系。在这个意义上，本文把股票和基金等同看待，统称证券，而把股市和基金市场等同看待，统称市场。

下面，我们分三种情况展开讨论。

1. 第一种情况，我们假设每支证券各涨 1 元，增量  $X =$  增量  $Y = 1$ 。市场状态向量的端点从  $A_0 = (3, 4)$  运动到  $A_1 = (4, 5)$ 。

我们的问题是：两支证券中，哪支证券好，哪支证券涨势高？

答案是：如果当初我们向两支证券分别投入 12 元，买  $X$

证券 4 股，买 Y 证券 3 股，由于证券涨价，证券 X 的价值从 12 元增值到 16 元，Y 从 12 元增到 15 元，所以 X 的涨势高，记为： $T_{(x)} > T_{(y)}$ 。T 表示趋势 (Trend)，是多维即时趋势 MIT 的简缩。关于趋势值和增率之间的关系，趋势值的定义式和计算式之间的关系，请参阅下文。

2. 第二种情况，我们假设每支证券各跌 1 元，增量  $X =$  增量  $Y = (-1)$ ，代表市场的点由  $A_0 = (3, 4)$  运动到  $A_2 = (2, 3)$ 。同样的问题，哪支证券好，哪支证券涨势高？

答案是：根据同样的计算，X 证券从 12 元跌到 8 元，Y 证券从 12 元跌到 9 元。

所以，Y 证券跌势小，或 Y 证券涨势高，用  $T_{(x)} < T_{(y)}$  表示。

3. 第三种情况，我们假设每支证券价格各翻一番，市场从  $A_0 = (3, 4)$  运动到  $A_3 = (6, 8)$ 。问，哪支证券涨势高？

答案是：12 元投资双双增值到 24 元，两支证券涨势相等，用  $T_{(x)} = T_{(y)}$  表示。

这个数学小游戏告诉我们什么呢？

当我们用直角坐标系 XOY 中的一个点 A 表示二元系统的状态时，连接并延长 OA，射线 OA 把 XOY 平面分成三个区域：扇形 YOA、射线 OA 和扇形 AOX。代表市场状态的点在空间的运动可以被分解为“径向运动”和“横向运动”。如果代表市场状态的点 A 沿射线 OA 运动，则向量 OA 的长度延伸或压缩，但向量所代表的系统状态、系统组成却没有变化。好比，我们用“元”表示证券价格，或分别用“角”、“分”表示证券价格，虽然相应的量被扩大了 10 倍，100 倍，但不影响市场的状态。这是 MDSM 的基本性质、特点。在 MDSM 分析中， $nA = A$  (详见讨论)。

如果表示市场状态的点进入 AOX 区域，向量 OA 朝 X 轴





方向偏转，则向量所代表的证券市场有变化，而且  $X$  的趋势值大于  $Y$  的趋势值。反之，如果  $OA$  向  $Y$  轴偏转，则  $Y$  的趋势值大于  $X$  的趋势值。如图 3 所示。总结以上的讨论，我们可以看到证券的趋势值和证券的增量 ( $\Delta$ ) 无直接关系，而与增长率 (%) 和初始值有关系。也就是说：

1. 证券是指数增长， $Y_k = Y_0 \times \lambda^k$ ；而非线性增长， $Y_k \neq Y_0 + \Delta \times k$ ；
2. 市场的结构变化可用市场状态向量在空间的偏转来表示。

以上结论可以推广到  $m$ -空间。在  $m$ -空间中，多元向量的方向可以用向量的余弦来表示。而向量的余弦，可以通过向量除以向量长度（标准化）来取得，也即，从多维空间的点向单位超球面投影来取得。这便是“超球面模型”名称的由来。证券市场的变化，表现为代表市场的点在高维证券空间的运动，这种运动可以被分解为互相垂直的两种运动：径向运动和偏转运动。市场状态向量在高维空间的偏转，表现为它的投影点在超球面上的位移。超球面上的点的运动，反映为超球面上的点在  $m$  个坐标轴上的投影的变化，也就是代表股市的状态向量和  $m$  个坐标轴夹角余弦值的变化。我们可以把上面的描述概括为，系统的组成比例的变化，表现为系统状态向量在高维空间的偏转，而向量在高维空间的偏转可以用它夹角余弦值的变化来监测，来度量。MDSM 定义状态向量后前余弦值的商（注意，是商，不是差）为多维即时趋势向量，表示证券市场在给定时间的运动趋势；而用它的分量，趋势值，来描述相应证券的变化趋势。

这样我们就清楚了，用 MDSM 进行证券市场分析的基本步骤如下：

- (1) 建立多维证券空间， $m$ -空间；

- (2) 用多维证券空间中的点,  $m$ -向量表示市场状态;
- (3) 用向量长度作指数, 定名为商高指数, 简称 SGI, 表示市场的量值;
- (4) 用方向(余弦)表示市场的结构状态, 证券价格在市场结构中的份额是相对稳定的;
- (5) 定义后前余弦的商为多维即时趋势 MIT 表示市场状态的变化趋势;
- (6) 以趋势值排序证券;
- (7) 根据序位决定售购。

根据定义式, 趋势值是时刻  $k$  的余弦值除以时刻  $k-1$  的余弦值:

$$T_{(i,k)} = \frac{Y_{(i,k)}}{|Y_{(k)}|} / \frac{Y_{(i,k-1)}}{|Y_{(k-1)}|}$$

其中,  $T_{(i,k)}$  是分量  $i$  在时刻  $k$  的趋势值,  $Y_{(i,k)}$  是分量  $i$  在时刻  $k$  的值,  $|Y_{(k)}|$  是时刻  $k$  的向量长度。

移项后, 得到趋势值的计算式:

$$\begin{aligned} T_{(i,k)} &= \frac{Y_{(i,k)}/Y_{(i,k-1)}}{|Y_{(k)}|/|Y_{(k-1)}|} \\ &= \frac{\text{证券率}_{(ik)}}{\text{商高指数率}_{(k)}} \end{aligned}$$

趋势值的计算式表示, 证券的趋势值和证券的变化成正比, 而和市场的变化成反比。

根据计算式, 设计 Excel 算表如表 1 所示(详见下节, 实验报告)。

### 三、实验报告

自 1998 年 6 月 19 日到 1999 年 1 月 29 日我们随机选取一

趋势分析 及其在生态股市中的应用



个证券公司 (Liberty) 的 20 个公共基金, 建立了由这 20 个基金组成的基金市场 (20 个基金的英文代码见表 2 第 1 列, 代码和基金名称的对照, 请参考表 1)。超球面模型以 20 个基金为坐标轴建立 20 维空间来代表这个基金市场。每天的市场状况, 用 20 个基金当天的收盘价 (Net Asset Value, NAV) 来表示。

表 2 第 2 列下部的 20 个数字表示 1998 年 6 月 19 日 20 个基金的收盘价格, 20 个数字组成了当天的市场状态。表 2 第 3 列的数字 (斜体) 是对照组 (平均随机回收, Average Random Investment), 平均投资 8000 美元于 20 支基金。第 4 列数字是每 400 美元所买的相应股数。比如, 1998 年 6 月 19 日, 基金 CFGAX, 每股 19.51 美元, 400 美元买 20.50 股。投资实验进行了 7 个月, 于 1999 年 1 月 29 日实验结束。1999 年 1 月 29 日的基金价格列在表 2 第 6 列。第 7 列 20 个数字 (斜体) 表示对照组 1999 年 1 月 29 日的基金的价值, 是第 4 列数字和第 6 列数字的乘积。比如, 第 1 行 CFGAX 的有关数据所表明信息是: 7 个月前对照组买了 CFGAX 基金 20.50 股, 1 月 29 日基金价格涨到了 20.56 美元一股, 所以当初投资 400 美元的基金, 1 月 29 日价值 421.53 美元。依此类推, 可以得到对照组 1999 年 1 月 29 日所有 20 个基金的价值 (第 7 列的 20 个斜体数字), 及其总和 (在 20 个数字上面)。总和是 8210.94 美元, 是初始的 8000.00 元的 102.64%。其中, 第 7 列的 20 个数字中的最大值是 475.93 美元, 是初始值 400.00 元的 118.98%。

与此同时, MDSM 管理的个人退休基金 (IRA) 投资 8000 美元 (去掉手续费, 实际投资 7960.79 元), 根据趋势值购买了 4 个基金 (CFGAX、CFSAX、COLIX 和 CUSGX), 分别投资 \$ 1997.17、\$ 1983.22、\$ 1997.18 和 \$ 1983.22。见第 5 列的数字 (黑体)。在以后的 7 个月里, 根据基金的即时趋势值 (MIT) 进行交易并调整。交易的原则是用趋势值大的基金代

替趋势值小的基金，或用排名榜上部的基金取代下部的基金，时刻保持我们持有的基金的半数以上处于排名榜上部。交易记录见表 3。

表 3 汇总了 27 次交易的基本信息。表 3 的第 1 列是交易的时间记录，分别是月、日、时、分。比如，820136027 代表提出交易的时刻是 8 月 20 日 13 时 60 分，编号（证券公司经纪人编号）27 的交易记录。第 2 列是卖出的基金代码，如 COLIX；第 3 列是卖出的股数，如：308.274 股。第 4 列是当日的收盘价，如 \$6.49。第 6 列是买进的基金代码。第 7 列是买进价，第 8 列是买进股数。第 7、第 8 列的分量的乘积应当等于第 3 第 4 列相应分量的乘积，表示买卖平衡，在整个实验过程中，既没有资金流入，也没有资金流出。在基金售出后，把售出价和当初的购入价相比，增、跌、持平分别用正、负、等号表示，列在增损栏，第 5 列里。

综上所述，第一行数据的交易信息是：8 月 20 日卖 COLIX 308.274 股，每股 \$6.49，用这笔资金买进 CUTLX 102.653 股，每股 \$19.49。由于 8 月 20 日的卖价（\$6.49）低于当初 6 月 19 日的买价（\$6.53），这笔交易赔了，用负号表示。

依此类推，经过 27 次交易，到 1999 年 1 月 29 日，MDSM 账户分别拥有：CFSAX、COLIX、SRSAX 和 STMAX 基金，所持股数分别是：211.841、401.053、165.621、和 155.644 股。按 1999 年 1 月 29 日的价格（参见表二第 6 列数字），它们的价值列在表二第 9 列（黑体数字），分别为：\$2355.08、\$2634.92、\$2636.69 和 \$2389.13，总计 10016.42 美元，是 1998 年 6 月 19 日初始投资的 125.82%。

需要说明的是，按证券公司有关规定，对于基金，无论什么时间提出交易申请，证券公司一律以当天的收盘价（NAV）进行交易。如交易申请的提出时刻晚于纽约时间下午 4 时，则





以下一个交易日的收盘价交易。也就是说，我们并不能根据网络传来的市场即时价格进行交易（而且事实上，网络里只有股票(Stocks)的即时价格，而没有基金(Funds)的即时价格），而是以“未来的收盘价格进行交易”。如果 MDSM 管理的投资能够增值，则 MDSM 必须要有监测预报市场的功能。而这正是我们的实验所要验证的。还有一点需要说明，在实际操作中，MDSM 所管理的资金，因为是退休基金，不能现兑，必须一直留在市场内，无论是熊市，还是牛市。所以，MDSM 管理的基金增值，靠的不是“市场上涨时进场，市场下跌时撤资”的投机行为。特别需要指出的是，在我们的实验中，MDSM 管理的基金甚至超过了市场中最好的基金。这说明，虽然市场行情对实验结果有影响，但实验结果并不受市场的局限。MDSM 管理的基金的高回收，不是靠市场在实验期间的综合状态好（实际上，市场的综合状态并不十分好，仅 2%），而是靠市场内各基金此起彼伏的动态变化。因此，我们有理由设想，如果能引入更多的基金，方差更大的相对变化，也许能带来更高的回报。

MDSM 如何监测预报基金市场的动态？表 1 根据 1999 年 1 月 8 日到 1 月 29 日，最后 15 个交易日的数据说明以基金为单位进行周趋势分析的机制。15 个交易日的数据按时间顺序可以被分为三周：上上周，上周和本周。然后，MDSM 计算每周的平均价格。求平均值的过程，用 MDSM 的术语，被称为中心化。在多维空间中，用一个  $m$ -形心向量代表  $n$  个  $m$ -向量。在我们的投资实验中  $m$  等于 20，20 个基金， $n$  等于 5，求 5 天平均值。20 个基金的三周平均值形成三个 20-形心向量 (Centroid Vector)。每个分量的 3 个值所形成的曲线，表示相应基金过去三周的变化。比如，根据 CFGAX 的 15 天数据，取得三个周的平均值，分别是 \$ 20.22，\$ 20.19 和 \$ 20.25。过这三

点，可以画一条曲线。根据曲线的两端点，曲线可以是上升的、下降的或水平的三种状态。根据中点，曲线可以是凸的、凹的或平直的三种形状。三种状态乘三种形状是9种曲线，基本可以描述该基金过去15天的运动趋势。以此类推，人们可以取得每个基金和高高指数的三个周平均值，分别画出周趋势曲线，再根据周趋势曲线来指导日投资。但是，由于我们人类肉眼的局限，一般很难区别曲线的细微差别，尤其是基金的数目很多时，所以MDSM计算即时趋势值，来代替画曲线。一般计算即时趋势值我们可以精确到小数点后4位。然后根据趋势值的大小来排列基金，如上所述。

表1的下半部就是根据1999.1.8—1999.1.29三周的数据，用第7列“本周”的趋势值排序的结果，称基金排名榜。趋势值大的基金排在榜的上面，小的排在下面。如果投资者能经常保持自己所持有的基金的半数以上排在榜的上部，便可以保证总回收高于市场平均回收。如果某个基金跌落到榜的下部，则应随时调换成榜上半部的基金。

由于惯性的作用，我们用周趋势分析指导日投资，成功的几率较高。当然，突发事件是有的，但就总体来看，惯性是主导的。如果再参照表1第8列二阶趋势值（趋势值的趋势值，MITMIT，表明趋势曲线的拐点），更可以提高趋势分析成功的机率。

退休金一般每年投资一次。由于新的财务年度开始，1999年1月25日星期一有新的退休金投入，第一阶段的投资实验在1月29日星期五告一段落。自1999年1月29日起，基金数目扩大到28支，实验投资额追加到2万美元，实验继续进行。实验结果，每周六在MDSM的网址：<http://www.mdsm.org/in-action>上公布。到2000年1月28日第二期实验结束时，MDSM 12个月的增率是31.99%，而28个基金的平均增率是9.82%

趋势

## 趋势分析 及其在生态股市中的应用



(报告另发, 见 ESA2001)。

### 三、结果与讨论

使用数学工具模拟人的经验, 用机械的或技术的方法监测预报市场动态, 建立买卖证券的技术分析方法 (Technical trading rules), 一直是个引人入胜和很具挑战性的课题。我们的实验表明, 证券的运动规律可能是可以被认识的。通过认真观察, 我们会发现, 任何证券价格曲线的后面, 都隐含着一条趋势线, 而证券趋势线后面又有市场的大势在起作用。超球面模型管理的基金实验表明, 证券价格的波动可以被分解为 4 个部分: 市场的总体波动、个别证券趋势值波动、个别证券日波动和随机误差:

价格<sub>(i,k)</sub> = 市场大势<sub>(k)</sub> + 证券趋势<sub>(i,k)</sub> + 证券日波动<sub>(i,k)</sub>  
+ 随机误差

因此有

趋势值<sub>(i,k)</sub> + 随机误差 = 价格<sub>(i,k)</sub> - 证券日波动<sub>(i,k)</sub> - 市场大势<sub>(k)</sub>

超球面模型使用中心化 (Centralization) 从证券价格波动中滤掉证券的日波动, 用标准化 (Standardization) 滤去市场的波动, 检出趋势, 并用趋势值将证券进行横向比较。在市场上涨时, MDSM 挑选增率高, 相应风险也高的证券。例如前面的情况 2.1.1, 从 +33% 和 +25% 中选择 33%; 而在市场下跌时, 挑选跌率低, 相应风险也低的证券, 如前面的情况 2.1.2 中, 从 -33% 和 -25% 中选择 -25%。在总风险不甚增加, 资金不离开市场的情况下, 取得高回报。超球面模型从  $m \times k$  的股价  $\times$  时间矩阵中提取出  $m \times 1$  的趋势向量, 使  $m$  个证券可

以在一维里排序。在美国证券市场上， $m = 9000 +$ 。由于变量独立的前提假定，MDSM 可以将这 9000 支证券同时进行横向比较，即时排队，这样便为投资者提供了更多的选择，并大大地提高了选择的效率。

超球面模型的重要前提是变量独立。一般认为，在低维空间里无解的股市监测问题，在高维空间里可能有解。而建立多变量空间的前提条件是变量之间独立。变量独立，坐标轴才能互相垂直，我们以上的讨论才能展开。MDSM 用多元向量来表示市场状态的前提是变量之间无关，而忽略变量之间的相互作用。这样虽然要丢失部分信息，引起分析误差，但我们的实验说明，这毕竟提供了一个解决问题的办法。而且，在现实的证券市场中，虽然各个证券互相影响、互相制约，但由于它们之间根本的关系是互相竞争，每个证券都保持自己独特的运动形式和运动规律，所以，直到能够在  $m$ -空间里定量地确定证券之间的关系，而且表示这种关系的值超过某个预定的阈值前（一般地，MDSM 把这个阈值定为 0.7，是 45 度角的余弦值），证券之间无关的假定是应该成立的。这里，“在多维空间”是关键的条件状语。据说有大量的文献支持证券之间相关的论点。我们以为，证券之间的相关，部分可能是从市场传递过来的。从 4.1 式我们可以看到，所有的证券价格都和市场大势相关。那么，如果不排除市场的因素（不标准化数据，不向超球面投影），证券相关的结论也许是不正确的。特别需要指出的是，数学上有条定理： $n$  维空间中最多只能有  $n$  支线性无关的向量。也就是说，如果在二维空间中研究 3 个以上的变量，则结论必定是相关，而不论它们是否真正相关。所以，对于变量独立的问题，我们只接受在  $m$ -空间里的研究结果。如果能够证明两个变量相关，则我们可以归并这两个变量，去掉一个坐标轴，然后，继续在  $m - 1$  维空间里研究系统





和变量的动态。MDSM 假定变量之间无关，只是作为我们研究讨论的出发点，而不是结论。好比，“无罪推定”比“有罪推定”更科学，但假定无罪，并不等于肯定无罪。从另一方面来说，由于向量合成分解的平行四边形法则，任意向量可以被分解为矩形（特殊的平行四边形）。也即，任意向量可以被分解为两个互相垂直的向量。加之， $m$ -空间对坐标轴的数目没有限制，这样，扩展的向量分析就为建立  $m$ -维证券空间提供了依据：即使变量之间有相关关系，只要相关系数没有达到 100%，人们就可以利用其不相关的部分，建立一个坐标轴。

除变量数目多外，股市问题难解的另一个原因可能是，股市变化的因果关系不详，或不确切，不稳定。MDSM 使用的是多变量时间系列方法。它假定变量之间无关，把股市变化的原因归之于时间，把趋势值向量  $T_{(i,k)}$ 、市场状态向量  $Y_{(i,k)}$  定义为时间  $k$  的函数，利用惯性（Inertia）原理求解。物理学常识告诉我们，一个系统，当具有足够的质量时，它便具有惯性。物理学上的惯性在一些文献上相应的提法是“线性增长在邻域”或“算术增长在邻域”（Linearity in a neighborhood。MDSM 把惯性原理推广到了生物、金融等领域，相应的数学描述则被扩展为“指数增长在邻域”或“几何增长在邻域”。打个比喻，对于变量的两个取值：3，4，大数定律认为真值是 3.5；而惯性认为下一个值是 5（算术增长），或 5.3（几何增长）。在技术层面上，在实际操作中，惯性的表现是，趋势值的期望值等于 1，基金在观察期间倾向于保持原态势不变。

超球面模型最大的特点是强调角度。在 MDSM 分析中， $nA = A$ ，其中， $A$  是向量的方向， $n$  是任意实数。它的意思是：角的两边可以任意扩大、延长若干倍，而角度保持不变。也即俗语所说的，“放大镜不能放大角度”。这个命题， $nA = A$ ，粗看似乎和我们已知的知识不很和谐，但是，这个基本原

理却是一切几何制图的基础（对应内角相等的两个三角形相似），因而也是我们认识客观世界的基础。比如，我们都同意，8号铅字的 MDSM 和 12 号铅字的 MDSM 是同一回事。不因为字体的变化而改变它所携带的信息。

超球面模型在揭示证券市场的运动规律，提出一个比较科学的技术分析方法的同时，更重要的是，它为证券监管部门提供了一个中肯全面的指数和一条监管股市的新思路。使用 MDSM，证券监管部门能更好地了解掌握股市的行情和动态。根据 4.2 式，分析个股趋势的必要条件之一是滤去市场大势。而滤去市场大势的重任，似乎非商高指数莫属。超球面模型定义的“商高指数”是向量长度，是所有分量的平方和的算术根，是几何和，是无偏估计。相对应的，虽然道·琼斯指数的历史作用无可否认，但它以偏概全的事实也不容回避。道·琼斯指数是经过人为精心挑选的少数股票的综合，数学上是有偏估计，却往往被错误地宣传为整个股市的指数，被理解为股市的上限。道·琼斯指数下跌，当然表明股市有下跌，但由于它的偏倚作用，以偏概全的结果，道·琼斯的下跌往往被理解为整个股市的，全体股票的下跌。加上媒体的炒作，便可能引发股民的抛售，造成股市的暴泄。相比之下，MDSM 用“商高指数”全面中肯地反映股市的动态。它的涨跌仅反映市场的平均状态。无论商高指数上涨或下跌，总有约一半的证券相对上涨，而另有一半证券相对下跌。投资者总能在风险和回报当中找到平衡点，而不一定要从市场撤出资金。比如，当商高指数下跌时，投资者仍旧可以清醒地将资金转移到跌率较小的基金避跌，如 2.1.2。从而，避免了恐慌性的、滚雪球式的大抛售。另一方面，MDSM 可以用向量方向的变化来表示资金在股市内部的流动。比如，向量  $A = (3, 4)$  和向量  $B = (4, 3)$ ，虽然向量长度都是 5，但向量的空间指向不同。当市场状态

趋势分析 及其在生态股市中的应用



从  $A = (3, 4)$  变化到  $B = (4, 3)$  时，可以定量地“看到”有资金从  $Y$  向  $X$  流动。由于，超球面模型既能提供科学的股市分析手段，加深人们对股市的认识，提高股民的投资信心，又能够增加监管部门的监管力度，这样便为规范化，进而繁荣我们的金融市场，化解金融危机提供了一条可能的新途径。

如果说，当传统的向量分析被局限在三维空间、线性空间的时候，依然在工程力学、量子力学方面发挥了作用，那么，扩展了的向量分析在简单系统的分类，系统动态分析方面（包括，但不局限于植被，股市）必然会起巨大的作用。它的应用前景目前是无可估量的。特别是，根据 M. Crowe 的研究，在人类历史上，人们对数的认识有过几次扩展。数集 (Set) 的每一次扩展，无论是从自然数 (Naturals) 到整数 (Integers) 的扩展，从整数到有理数 (Rationals) 的扩展，或者从有理数到实数 (Reals)，实数到复数 (Complex Number) 的扩展都给科学的发展、人类的认识带来了巨大的飞跃。那么由标量 (Scalars) 到向量的充分扩展，由数轴 (Real Line) - 复平面到多维空间 ( $m$ -Space) 的扩展，也必将为我们人类对环境 (广义的环境) 的认识带来巨大的突破。当然，MDSM 现在仍然很幼稚。它的有关运算规则、前提假定、理论都是很简单的、很基本的。然而，正因为如此，它很可能是属于上游的、基础的科学。它的许多不完善之处，不高级之处，正是有待人们继续认识，深入发掘的未知领域。我们希望本文的发表能够引起其他学科，如数学、金融、经济、计算机等领域的专家学者的兴趣，大家共同来研究，开发这个新的知识宝库。

谨以此文纪念前中国科学院院士，内蒙古大学生物系教授，作者的导师李博先生。本文是李博、雍世鹏、刘钟龄等先

生 20 年前给作者布置的课题：“寻找一种简单易行并且有数学基础的植被分析方法”的延续和扩展。实验结果曾经在美国首都华盛顿召开的“21 世纪的中国和可持续性发展研讨会”暨欧美同学会 99 年年会上报告。会后根据幻灯片整理，并经 (Weatherhead School of Management, Case Western Reserve University)，梁兵教授审阅，复核了数据，提出修改意见。特此表示感谢。

趋

势分析 及其在生态股市中的应用





表 1 (1) 周趋势分析及排序

基金代码	基金名称	趋势分析			MIT1	MIT2	MITMIT
		上上周	上周	本周			
SGI	商高指数	65.4567	65.3057	64.8697	0.9977	0.9933	0.9956
CFGAX	C US Stock	20.22	20.19	20.25	1.0009	1.0096	1.0087
CFSAX	C Federal Securities	11.03	11.08	11.10	1.0061	1.0093	1.0031
CGUAX	C Global Unities	16.18	16.41	16.32	1.0166	1.0012	0.9849
CIFAX)	C Intl Fd for Growth	9.81	9.81	9.81	1.0023	1.0067	1.0044
CNTAX	Tiger	8.06	7.82	7.30	0.9725	0.9396	0.9661
COEAX	C Global Equity	14.62	14.55	14.44	0.9974	0.9991	1.0017
COLFX	C Fund	10.69	10.65	10.67	0.9986	1.0084	1.0099
COLGX	C Select Value	21.58	21.55	21.47	1.0010	1.0026	1.0016
COLHX	C Hi Yld Securities	6.77	6.76	6.79	1.0005	1.0115	1.0109
COLIX	C Income Fund	6.52	6.54	6.56	1.0054	1.0089	1.0035
CONAX	Internatl Horizons FD	12.39	12.35	12.22	0.9988	0.9961	0.9974
COSIX	C Strategic Income	7.09	7.10	7.09	1.0032	1.0064	1.0033
CSBAX	C Strategic Balanced	15.56	15.51	15.48	0.9990	1.0045	1.0056
CSMIX	C Small Cap Value	29.94	29.76	29.25	0.9963	0.9893	0.9930
CTCAX	Tiger Cub	5.91	5.72	5.51	0.9711	0.9694	0.9983
CUSGX	C US Government	6.73	6.74	6.76	1.0038	1.0094	1.0056
CUTLX	C Unities Fund	21.57	21.76	21.49	1.0111	0.9944	0.9835
NGCAX	Greater China	10.47	9.82	9.34	0.9403	0.9577	1.0186
SRSAX	Growth Stock	14.90	15.13	15.50	1.0182	1.0309	1.0125
STMAX	TaxManaged Growth	15.15	14.95	15.06	0.9892	1.0143	1.0253

表 1 (2) 基金排名榜

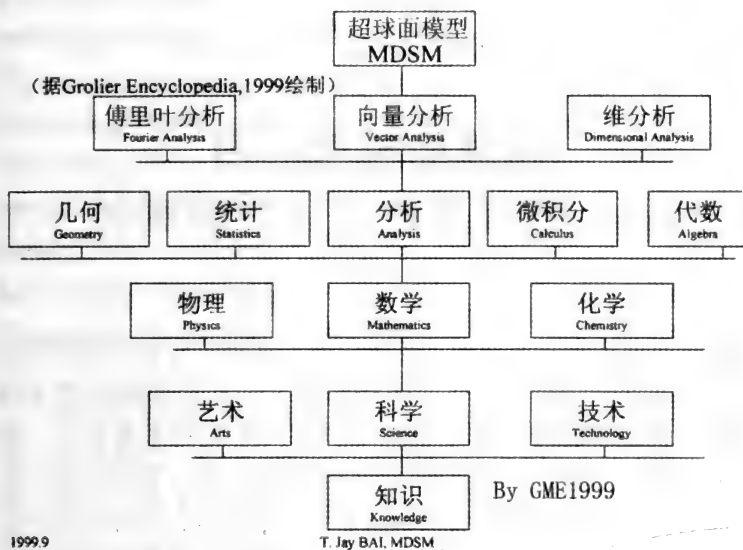
基金代码	基金名称 商高指数	排名			MIT1	MIT2	MITMIT
		上上周	上周	本周			
SRSAX	Growth Stock	14.90	15.13	15.50	1.0182	1.0309	1.0125
STMXX	TaxManaged Growth	15.15	14.95	15.06	0.9892	1.0143	1.0253
COLHX	C Hi Yld. Securities	6.77	6.76	6.79	1.0005	1.0115	1.0109
CFGAX	C US Stock	20.22	20.19	20.25	1.0009	1.0096	1.0087
CUSGX	C US Government	6.73	6.74	6.76	1.0038	1.0094	1.0056
CFSAX	C Federal Securities	11.03	11.08	11.10	1.0061	1.0093	1.0031
COLIX	C Income Fund	6.52	6.54	6.56	1.0054	1.0089	1.0035
COLFX	C Fund	10.69	10.65	10.67	0.9986	1.0084	1.0099
CIFAX)	C Intl Fd for Growth	9.81	9.81	9.81	1.0023	1.0067	1.0044
COSIX	C Strategic Income	7.09	7.10	7.09	1.0032	1.0064	1.0033
CSBAX	C Strategic Balanced	15.56	15.51	15.48	0.9990	1.0045	1.0056
COLGX	C Select Value	21.58	21.55	21.47	1.0010	1.0026	1.0016
CGUAX	C Global Utilities	16.18	16.41	16.32	1.0166	1.0012	0.9849
COEAX	C Global Equity	14.62	14.55	14.44	0.9974	0.9991	1.0017
CONAX	Internatl Horizons FD	12.39	12.35	12.22	0.9988	0.9961	0.9974
CUTLX	C Utilities Fund	21.57	21.76	21.49	1.0111	0.9944	0.9835
CSMIX	C. Small Cap Value	29.94	29.76	29.25	0.9963	0.9893	0.9930
CTCAX	Tiger Cub	5.91	5.72	5.51	0.9711	0.9694	0.9983
NGCAX	Greater China	10.47	9.82	9.34	0.9403	0.9577	1.0186
CNTAX	Tiger	8.06	7.82	7.30	0.9725	0.9396	0.9661

表2 MDSM 指导的投资与平均随机回收的比较

\$1,805.431		98年6月19日				99年1月29日			
	股 价 (元/股)	对照投资 (元)	对照股数	MDSM 投资 (元)	股 价 (元/股)	对照价值 (元)	MDSM 股	MDSM 价值 (元)	
商通指数	65.375	---	---	---	66.136	---	---	---	
增长率	---	---	---	---	---	102.64%	---	125.82%	
总计 (\$)	---	\$8,000	---	\$7,960.79	---	\$8,210.94	---	\$10,016.42	
CFGAX	19.51	\$400	20.50	\$1,997.17	20.56	\$421.53	---	---	
CFSAX	10.90	\$400	36.70	\$1,983.22	11.12	\$408.07	211.841	\$2,355.68	
CGUAX	15.17	\$400	26.37	---	16.31	\$430.06	---	---	
CIFAX	9.25	\$400	43.24	---	9.81	\$424.22	---	---	
CNTAX	6.44	\$400	62.11	---	7.30	\$453.42	---	---	
COEAX	15.11	\$400	26.47	---	14.57	\$385.70	---	---	
COLFX	10.77	\$400	37.14	---	10.78	\$400.37	---	---	
COLGX	21.75	\$400	18.39	---	21.65	\$398.16	---	---	
COLHX	7.25	\$400	55.17	---	6.80	\$375.17	---	---	
COLIX	6.53	\$400	61.35	\$1,997.18	6.57	\$402.45	401.053	\$2,634.92	
CONAX	13.42	\$400	29.81	---	12.28	\$366.02	---	---	
COPIX	7.31	\$400	54.72	---	7.11	\$389.06	---	---	
CSBAX	15.53	\$400	25.76	---	15.57	\$401.03	---	---	
CSMAX	33.23	\$400	12.04	---	29.29	\$352.57	---	---	
CTCAX	4.94	\$400	80.97	---	5.51	\$446.15	---	---	
CUSGX	6.65	\$400	60.15	\$1,983.22	6.76	\$406.62	---	---	
CUTLX	19.97	\$400	20.03	---	21.50	\$430.65	---	---	
NGCAX	9.39	\$400	42.60	---	9.37	\$399.15	---	---	
SRSAX	13.38	\$400	29.90	---	15.92	\$475.95	165.621	\$2,636.69	
STMAX	13.81	\$400	28.96	---	15.35	\$444.61	155.644	\$2,389.13	

表3 MDSM 指导的投资的交易记录汇总

交易代码	基金代码	股数	股价	增损	基金代码	股数	股价
98619		\$1,997.18			COLIX	305.845	6.53
98619		\$1,997.17			CFGAX	102.367	19.51
98619		\$1,983.22			CFSAX	181.947	10.90
98619		\$1,983.22			CUSGX	298.229	6.65
820136027	COLIX	308.274	6.49	-	CUTLX	102.653	19.49
821105179	CFSAX	183.234	10.96	+	CONAX	161.046	12.47
827181084	CFGAX	102.592	17.62	-	CFSAX	163.442	11.06
901143579	CONAX	161.852	11.74	-	COLIX	293.685	6.47
902152958	CFSAX	163.442	11.05	-	CSMIX	69.758	25.89
909144154	CUSGX	302.462	6.76	+	NGCAX	265.538	7.70
911145054	CSMIX	69.814	25.76	-	CFSAX	160.859	11.18
915144254	COLIX	293.685	6.50	+	COLGX	97.895	19.50
917145342	CFSAX	160.860	11.18	=	CNTAX	329.984	5.45
1001141846	COLGX	98.147	18.23	-	CTCAX	412.263	4.34
1007140520	NGCAX	266.272	8.21	+	CFSAX	192.607	11.35
1015142784	CUTLX	103.451	20.56	+	CONAX	182.728	11.64
1021144046	CFSAX	192.607	11.21	-	CSMIX	82.884	26.05
1023111084	CNTAX	330.297	7.13	+	COLFX	232.021	10.15
1027111411	CONAX	183.034	12.10	+	COLHX	343.899	6.44
1103143459	CTCAX	412.263	5.57	+	NGCAX	208.000	11.04
1105144452	COLHX	343.899	6.57	+	CNTAX	294.578	7.67
1110140774	NGCAX	207.998	10.43	-	CFGAX	111.720	19.03
1113140859	CSMIX	83.073	28.19	+	COLIX	360.281	6.50
1119150046	COLIX	360.282	6.50	=	SRSAX	173.855	13.47
1201155818	CNTAX	295.269	7.66	-	COEAX	154.070	14.68
1204	COLFX	232.020	10.90	+	CSMIX	86.404	29.27
1204	COEAX	154.070	14.73	+	COLGX	106.250	21.36
0115141608	CFGAX	116.190	20.22	+	CFSAX	211.841	11.09
0115141608	SRSAX	174.050	15.07	+	COLIX	401.053	6.54
0125	有新资金注入						
0126	CSMIX	86.404	29.5	+	SRSAX	165.621	15.39
0128153790	COLGX	109.42	21.55	+	STMAX	155.643	15.15



1999.9

图1 超球面模型在知识树上的位置

## 超球面模型应用 于草原监测数据分析的探讨

**摘要：**草原监测是必要的和可能的。本文提出分析归纳草原监测多元数据的新数学方法，超球面模型。 $M$ 个物种所组成的草原的状态可以用 $M$ 维变量空间的点， $M$ 元向量来表示，向量的长度用来表示草原的量值，向量的方向表示草原组成。从草原监测的角度，草原的相对组成是更重要的，成分变化的比率更能说明问题。在处理草原监测数据中使用向量等比系列和向量等差系列是可行的。等比系列比较能反映生物几何增长的本质。为了纠正中间过程被忽略的弱点，要坚持逐年监测，逐步分析，或使用滑动平均。二阶趋势值可尝试用于解决时间滞后问题，卡门滤波用于降低预报误差。因为草原监测需要使用分种数据，必须以人工监测记录为主，再辅以其他高科技手段，如遥感。

**关键词：**草原监测 多元向量时间系列 向量分析 等比系列  
超球面模型

**Abstract:** Monitoring grassland is important and possible. Multi-Dimensional Sphere Model (MDSM) is a new multivariate data analysis method to analyze the grassland monitoring data. Grassland can be expressed as a point in multivariate space, or a multi-component vector. The magnitude of the vector expresses the biomass of the grassland, but the direction of the vector expresses the composition of the grassland.

趋势分析 及其在生态股市中的应用



For temporal dynamic analysis, the composition of the grassland is more important than the production of the grassland, and the ratio of the changes can explain more than the differences do. Both the geometrical and arithmetic series can be adopted to analyze the grassland monitoring data. However, MDSM prefers the geometrical series. Moving average and second degree trend may be introduced to improve the model.

**Key words:** Grassland Monitoring, Multi-variate Time Series, Multi-Dimensional Sphere Model, Geometrical Series, Arithmetic Series

草原是中华民族生态屏障。森林、草原、荒漠和人的生态位排列可以一维地表示如下：森林—人—草原—荒漠。草原的生态位介于人和荒漠之间。草原的水土保持功能虽然小于森林，但所占土地面积大，而且处于我国经济发达地区的上游，上风头。草原的状况，直接或间接地影响到下游或下风地区，人口稠密地区的经济发展和人民生命财产的安全。近几年的长江洪水，黄河断流，专家无不认为与上游植被被破坏有关。今年的沙尘暴，更是突出了问题的严重性，但是，沙尘暴的发生不是偶然的，质变往往是量变积累的结果。人们曾经希望在量变的早期就感知环境的细微变化。为此建立了许多生态定位站点，并已经积累了大量的观测数据。现在，我们需要把各自独立的站点连成线，织成网，并在草原监测网上增加时间轴，实现草原的动态监测。本文从数据分析的角度探讨草原观测数据的表述、归纳，和使用超球面模型（Multi-Dimensional Sphere Model, MDSM）处理多变量时间系列的基本原理。超球面模型是基于多维空间（ $m$ -space）使用多元向量（ $m$ -vectors）的多元分析法，可用于草原、股市等资源共享指数增长系统的时间动态分析。在超球面模型的研发初期，曾经使用美国“土地条件

趋势分析 (Land Condition Trend Analysis, LCTA)”课题的数据来开发模型,同时对 LCTA 的数据进行归纳和分析。但是,由于草原监测课题不仅需要大量的资金,而且操作周期长,短期内鲜见经济效益,又很难重复,很难验证,使模型的进一步开发、推广遇到了很大的困难。后来,在中国科学院阳含熙院士和应用数学所王寿仁教授的支持和鼓励下,研发者使用比草原数据公信力强、精度高、更新周期短、潜在经济效益高的证券市场的数据对模型进行了多次测算,进而用指导实际投资来检验、完善超球面模型。多次实验结果表明,超球面模型对多元系统,特别是资源共享指数增长的多元系统有动态监测的功能。本文中,我们在总结归纳使用多元向量监测股市经验的基础上,提出一个草原监测数据的处理模式,争取早日开展草原监测网的工作,以保卫国家的生态安全,造福后代。草原监测数据的处理方法的研究和解决有跨学科的意义。

## 一、草原的数学表述与多元向量

草原植被是由多达上百种植物:如羊草 (*Anurolepidium* spp.)、针茅 (*Stipa* spp.)、铁杆蒿 (*Artemisia* spp.) 等组成的。不同的组成成分可以有不同的生态学意义,比如,一般认为羊草属于中生优质牧草,针茅属于旱生中等牧草,蒿子是一年生适口性差的牧草。每种植物都可能携带着草原演替的重要信息。为了全面地描述草原,我们用  $m$  个物种 (这里  $m$  特指大于 3 的自然数) 为坐标轴建立  $m$ -物种空间,并用  $m$ -物种空间的点或  $m$ -元向量 (连接  $m$ -空间的点和原点,我们得到多元向量, Position Vectors) 来表示  $m$  个植物所组成的草原的状态。

$$X = (X_i) = (X_1, X_2, X_3, \dots, X_{m-1}, X_m) \quad i = 1, 2, \dots, m \quad 1$$

趋势分析 及其在生态股市中的应用





其中，我们用  $X$  表示草原观测样本向量，用  $X_i$  表示向量的第  $i$  个分量，下标  $i$  用来区别分量，即植物种。我们使用  $m$ -元向量来描述草原植被的重要原因是“向量是既有量值 (Magnitude) 又有方向 (Direction) 的变量”。而草原是我们面临的一个新的特殊的研究对象，不仅因为它是多元的，更主要的是因为它既有量的属性，又有质的属性。草原的“量”指的是它的生物总量的问题、产量的问题；草原的“质”指的是它的组成成分的问题。草原的状况从根本上来说是由它的组成来决定的。草原的总产量可以随年降水量有很大的波动，但它的相对组成却应该是相对稳定的。比如，对于上面所列举的（羊草 + 针茅 + 蒿子）草原，向量  $A_1 = (3, 1, 0)$  和向量  $A_2 = (6, 2, 0)$ ，虽然量值不等，却都是优质的羊草草原。它们之间在量值上的不同可能是由不同年代（早年或润年）、不同生长季（夏季或秋季）、不同取样面积（10 平方米或 20 平方米）造成的。而另一方面，产量相同的草场，却可能是组成完全不同的植被。比如，对于上面所列举的（羊草 + 针茅 + 蒿子）草原，向量  $A = (3, 1, 0)$  和向量  $B = (0, 1, 3)$ ，虽然量值相等，却是不同的草原： $A$  是优质的羊草草原，而  $B$  却是退化了的蒿子地。草原的这种特性，既有量的属性，又有质的属性，可以，而且只能用向量来描述。向量既可以用量值表示草原植被所含物质、能量、信息的总量，又可以用方向来表示这个总量在  $m$  个分量间的分配。进一步，在时间动态分析 (Temporal Dynamic Analysis) 中，向量是草原植被中不同组成成分的升降向我们传递着植被演替的重要信息，而且只有这种成份升降才携带着植被变化的信息。比如对于上述的（羊草 + 针茅 + 蒿子）草原，羊草比重提高，可能表示草原在向中生的、改良的方向发展；针茅比重提高，可能表示草原在向早生的方向演变；而蒿子比重增加，可能表明草原在退化。这样，我们用向量不仅

能够表示草原的量,更重要的是能描述草原的质及其变化。

为了充分表示草原的成分,也为了使来自不同地区、不同年代、不同面积的观测数据可以互相比较、归并,以便充分利用所有不同来源的样本,超球面模型首先要把数据标准化,样本的每个分量除以样本的向量长度:

$$Y = X' = (X_i / |X|)$$

$$i = 1, 2, \dots, m \quad 2-1$$

$$\text{或}(Y_i) = (X_1 / |X|, X_2 / |X|, X_3 / |X|, \dots, X_{m-1} / |X|, X_m / |X|) \\ = (Y_1, Y_2, Y_3, \dots, Y_{m-1}, Y_m)$$

$$i = 1, 2, \dots, m \quad 2-2$$

其中,多元向量  $Y = X'$  代表样本在“单位超球面”上的投影。式 2-1 是向量表示法,式 2-2 是分量表示法。标量  $|X|$  是样本向量的长度,等于各分量的平方和的算术根:

$$|X| = \sqrt{\sum X_i^2}, i = 1, 2, \dots, m \quad 3$$

由于向量长度是  $m$  个分量的“向量和”,也可以把向量标准化,各分量除以向量和,类比于多维的百分比。标准化较百分比优越之处在于,标准化向量  $Y$  的坐标  $Y_i$  是向量与相应分量所代表的轴的夹角余弦值,  $\cos\theta_i = X_i / |X|$ 。 $m$  个余弦值组成余弦向量(状态向量),表示向量在多维射影空间( $m$ -Projective Space)的指向。它是多维空间中单位超球面上的一个点。通过数据标准化,模型把草原状态向量的角度和植被的组成联系起来。这样便突出了植被组成所携带的信息,并可以通过监测状态向量在单位超球面上的投影的运动,来监测植被的变化。所以说,标准化是超球面模型的关键、核心。在以后的讨论中,若非特别说明,我们所涉及的都是标准化了的向量。

为了讨论草原随时间的变化,我们需要给草原变量再增加一个时间下标  $j$ 。这样,  $K$  年的草原状况可以用  $m$ -物种空间的  $K$  个点或  $K$  个  $m$ -元向量所组成的“多元向量时间系列”来表示。

趋势

势分析 及其在生态股市中的应用



$$Y_0, Y_1, Y_2, \dots, Y_j, \dots, Y_{K-1}, Y_K$$

$$j=0, 1, 2, \dots, K \quad 4$$

式 4 中,  $m$ -元向量  $Y_j = (Y_{1,j}, Y_{2,j}, Y_{3,j}, \dots, Y_{m-1,j}, Y_{m,j})$  表示  $j$  时刻的草原状态, 下标  $j$  表示时间。如  $Y_0$  表示草原的初始状态,  $Y_k$  表示草原开始监测后第  $k$  年的状态。这  $m \times K$  矩阵形式的数据, 我们特称为“多元向量时间系列”。它是  $m$ -空间的  $K$  支向量, 或  $m$ -空间的  $K$  个点, 表示草原在多维物种空间  $K$  年的运动轨迹。面对草原监测的历史数据, 多元向量时间系列 4, 我们的第一个任务是对草原监测的历史数据进行归纳, 第二个任务是使用已知的数据, 对  $K+1$  时刻的草原状况进行预报。

## 二、多元向量时间系列

自联合国教科文组织 20 多年前提出“人与生物圈计划”以来, 各国政府和科研单位, 纷纷制订长期生态定位研究计划。但草原监测的数据分析问题, 目前尚没有比超球面模型更成功的解决办法。我们这里试用“向量等差系列”和“向量等比系列”来解释超球面模型的工作机制。为了简便, 我们首先试用等差系列法, 然后导出等比系列法。对于式 4 所表示的“多元向量时间系列”, 我们可以把相邻两年的草原状况的差别用“多元差向量”  $= \Delta_j$  来表示,

$$\Delta_j = Y_j - Y_{j-1} \quad 5-1$$

$$\Delta_{i,j} = Y_{i,j} - Y_{i,j-1}$$

$$i=1, 2, \dots, m; j=1, 2, \dots, K \quad 5-2$$

式 5 用黑体表示的向量  $\Delta_j$ ,  $Y_j$  和  $Y_{j-1}$  都是  $m$ -元向量。第一下

标  $i$  用来区别分量, 取值从 1 到  $m$ , 第二下标  $j$  用来区别时间, 取值从 1 到  $K$ 。其中, 5-1 是多元差向量的定义式, 相邻两年的草原状态的差别, 式 5-2 是计算式。“多元差向量”  $\Delta_j$  的分量是后前两相邻状态向量  $Y_j$  和  $Y_{j-1}$  对应分量的差, 表示两状态的差异。“多元差向量”的分量  $\Delta_{i,j}$  的取值可以大于 0、等于 0 或小于 0, 分别表示相应分量增加、持平或减少。有了多元差向量  $\Delta_j$  后 ( $j=1, 2, \dots, K, k < K$ ), 第  $k$  年的草原状态向量  $Y_k$  便可以用初始状态向量  $Y_0$  与  $k$  个多元差向量  $\Delta_j$  的和来表示:

$$Y_k = Y_0 + \Delta_1 + \Delta_2 + \Delta_3 \cdots + \Delta_j + \cdots + \Delta_{k-1} + \Delta_k \quad 6-1$$

$$Y_{i,k} = Y_{i,0} + \Delta_{i,1} + \Delta_{i,2} + \Delta_{i,3} \cdots + \Delta_{i,j} \cdots + \Delta_{i,k-1} + \Delta_{i,k} \\ i=1, 2, \dots, m \quad 6-2$$

式 6 中, 向量  $Y_k$  是草原开始监测后第  $k$  年的草原状态向量,  $Y_0$  是草原的初始状态向量,  $\Delta_j$  是多元差向量。其中, 式 6-1 是向量表示法, 式 6-2 是分量表示法, 表示第  $k$  年的草原状态向量的第  $i$  个分量的计算法。当然, 每年的差向量之间不一定相等, 空间的  $K$  个点不一定等距, 因此, 4 式不一定是一个“向量等差系列”。但在没有其他更好的办法的情况下, 我们用等差系列的想想法来处理它、模拟它。也就是说, 对上列数据, 我们可以在求出“多元平均差向量” =  $\Delta_{\text{平均}}$  后回归:

$$Y_k = Y_0 + k \times \Delta_{\text{平均}} \quad 7-1$$

$$Y_{i,k} = Y_{i,0} + k \times \Delta_{i,\text{平均}}, i=1, 2, \dots, m \quad 7-2$$

比较式 6 和式 7,  $\Delta_{\text{平均}}$  = “多元平均差向量”是  $K$  个“多元差向量”的代数平均 (见式 8)。

$$\Delta_{\text{平均}} = (\Delta_1 + \Delta_2 + \Delta_3 + \cdots + \Delta_{k-1} + \Delta_k) / K \quad 8-1$$

$$\Delta_{i,j} = (\Delta_{i,1} + \Delta_{i,2} + \Delta_{i,3} + \cdots + \Delta_{i,k-1} + \Delta_{i,k}) / K, \\ i=1, 2, \dots, m \quad 8-2$$

趋势分析 及其在生态股市中的应用



式 8 说明“多元平均差向量”的求法：分量的平均值作平均差向量的分量。式 8-1 是定义式，式 8-2 是计算式，而  $K$  是时间链的长度。如果用图来表示，把式 7 的  $k$  换成  $K$ ，在时间  $j$  和分量  $i$  所构成的平面上，式 7 是连接始终两点， $Y_{i,0}$  和  $Y_{i,K}$ ，且过其他  $K-1$  个点的截距等于  $Y_{i,0}$ ，斜率等于  $\Delta_{i,平均}$  的直线（关于忽略中间过程的问题，详见讨论与结论）。对应于常规的在二维平面的标量线性回归，我们称式 7 为“多元向量线性回归”，或“在多维空间的线性回归”。该式对  $m$ -草原  $K$  年的历史数据进行了归纳，解决了前面提出的任务一。我们的任务是已知数据，对时刻  $K+1$  的草原状态进行预报。很显然，对于一个渐变过程或连续函数，我们可以根据惯性原理和式 8，用初始值向量  $Y_0$  与平均差向量  $\Delta_{平均}$  的  $K+1$  倍的和来预报草原在时刻  $K+1$  的状态向量：

$$Y_{K+1} = Y_0 + (K+1) \times \Delta_{平均} \quad 9-1$$

$$Y_{i,K+1} = Y_{i,0} + (K+1) \times \Delta_{i,平均}, i=1, 2, \dots, m \quad 9-2$$

式 9-1 给出了根据多元向量等差系列，使用历史数据预报草原未来状态的公式，9-2 表示对应分量的计算。以上是利用向量等差系列的思路来处理草原监测数据的推导过程。下面，我们根据同样的思路，在定义了  $m$ -向量乘法和除法的基础上，推导使用比值来处理草原监测数据的方法。对于上面讨论的式 10-1 表述的“多元向量时间系列”，我们既可以把它看作是  $m$ -空间的  $K$  个点，也可以把它看作是始自原点的  $K$  支向量，或一支  $m$ -元草原状态向量在多维空间连续偏转的  $K$  个状态：

$$Y_0, Y_1, Y_2, \dots, Y_j, \dots, Y_{K-1}, Y_K$$

$$j=0, 1, 2, \dots, K \quad 4$$

其中，多元向量黑体  $Y_j = (Y_{1,j}, Y_{2,j}, Y_{3,j}, \dots, Y_{m-1,j}, Y_{m,j})$  表示  $j$  时刻用余弦值表示的草原状态向量在空间的指

向，下标  $j$  表示时间。我们可以把两相邻状态向量指向的变化用“多元比向量” =  $\lambda_j$  来表示：

$$\lambda_j = Y_j / Y_{j-1} \quad 10-2$$

$$\lambda_{i,j} = Y_{i,j} / Y_{i,j-1},$$

$$i = 1, 2, \dots, m; j = 1, 2, \dots, K \quad 10-3$$

式 10-2 是“多元比向量”的定义式，相邻两年的草原状况的比（在超球面模型的一些文献里，多元比向量也被称为“趋势向量”（Multivariate Instantaneous Trend，简称 T）。式 10-3 是比向量的计算式，“多元比向量”  $\lambda_j$  的分量是后前两相邻状态向量  $Y_j$  和  $Y_{j-1}$  对应分量（余弦值）的比，表示向量在空间的偏转趋势。“多元比向量”  $\lambda_j$  的分量的取值范围可以大于 1、等于 1 或小于 1，分别表示状态向量向相应的坐标轴的靠拢、保持或远离（或相应分量的运动趋势在上升、持平或下降）。有了“多元比向量”  $\lambda_j$  后，第  $k$  年的草原状态向量  $Y_k$  便可以用初始状态向量  $Y_0$  与  $k$  个多元比向量  $\lambda_j$  的连乘积来表示：

$$Y_k = Y_0 \times \lambda_1 \times \lambda_2 \times \lambda_3 \times \dots \times \lambda_{k-1} \times \lambda_k \quad 11-1$$

$$Y_{i,k} = Y_{i,0} \times \lambda_{i,1} \times \lambda_{i,2} \times \lambda_{i,3} \times \dots \times \lambda_{i,k-1} \times \lambda_{i,k}$$

$$i = 1, 2, \dots, m \quad 11-2$$

式 11 中，向量  $Y_k$  是草原开始监测后第  $k$  年的草原状态， $Y_0$  是草原的初始状态， $\lambda_j$  是第  $j$  年的“多元比向量（趋势向量）”。其中，式 11-1 是向量表示法，式 11-2 是分量表示法，表示第  $k$  年的草原状态向量的第  $i$  个分量的计算法。虽然，每年的比向量不尽相同，式 4 不一定是多元等比系列，但我们仍然可以求出  $K$  个比向量的“多元均匀比向量”或“多元几何平均比向量”，简称“均匀比向量” =  $\lambda_{均匀}$ ：

$$\lambda_{均匀} = (\lambda_1 \times \lambda_2 \times \lambda_3 \times \dots \times \lambda_{k-1} \times \lambda_k) \quad 12-1$$

$$\lambda_{i,均匀} = (\lambda_{i,1} \times \lambda_{i,2} \div \lambda_{i,3} \times \dots \times \lambda_{i,k-1} \times \lambda_{i,k})^{1/K}$$

趋势分析 及其在生态股市中的应用



$$i = 1, 2, \dots, m$$

12-2

式 12 说明“均匀比向量”的求法：分量的“均匀比值”作均匀比向量的分量。式 12-1 是定义式，式 12-2 是计算式，而  $K$  是时间链的长度。其中， $\lambda_{\text{均匀}}$  是  $K$  个比向量的连乘积的  $K$  次根。在求出“均匀比向量” $\lambda_{\text{均匀}}$  后，我们可以用初始的草原状况和“均匀比向量”的  $k$  次幂的积来表示第  $k$  年的草原状况：

$$Y_k = Y_0 \times \lambda_{k, \text{均匀}} \quad 13-1$$

$$Y_{i,k} = Y_{i,0} \times \lambda_{i,k, \text{均匀}} \quad i = 1, 2, \dots, m \quad 13-2$$

式 13 中，向量  $Y_k$  是开始监测后第  $k$  年的草原状态， $Y_0$  是草原的初始状态，“均匀比向量” $\lambda_{\text{均匀}}$  也是  $m$ -元向量，是  $K$  个“多元比向量”的连乘积的  $K$  次根（见式 12）。同样，根据惯性原理，我们可以使用已知的数据，对  $K+1$  时刻的草原状态进行预报。用  $K+1$  取代式 13 的  $k$ ， $Y_{K+1}$  可以用初始状态向量与“均匀比向量”的  $K+1$  次幂的积来表示：

$$Y_{K+1} = Y_0 \times \lambda_{K+1, \text{均匀}} \quad 14-1$$

$$Y_{i,K+1} = Y_{i,0} \times \lambda_{i,K+1, \text{均匀}}$$

$$i = 1, 2, \dots, m \quad 14-2$$

式 14-1 给出了使用历史数据预报草原未来的公式，式 14-2 给出了对应分量的计算。对应于前面的“多元向量线性回归”，我们称式 14 为“多元向量几何平均回归”。与线性回归侧重距离不同，几何回归侧重的是比值（角度）。与线性回归比较，在草原监测中我们宁愿使用几何回归。因为，根据我们以上的讨论，当草原向量的各分量成等比例增长时，草原的状况保持不变。向量  $A_1 = (3, 1, 0)$  所代表的草原等于向量  $A_2 = (6, 2, 0)$  所代表的草原。但是，当草原向量的各分量成等差增长时，草原状况不一定能保持不变。向量  $A_1 = (3,$

1, 0) 和  $C = (4, 2, 1)$  比较, 各分量的增量都等于 1, 但它们所代表的草原却是不同的。用草原学的术语来说, 建群种和濒危种都增加一个多度单位, 两者的效应是不同的。因此, 我们的结论是, 草原状态要用向量来描述, 草原的变化要用向量的商来表示。况且, 使用几何回归, 在增长时, 分量增率取值从 1 到无穷大, 能够解释种群爆炸的现象; 在减少时, 增率取值从 1 到无穷小, 可以解释生命的顽强性。符合生物种的指数增长的本质。

### 三、结论与讨论

从以上的讨论, 我们可以看到, 超球面模型无非是把每个分量分别处理, 然后用所有分量的“向量和”(向量长度)来表示系统。模型的特殊之处在于思考的角度不同。使用向量的前提假定是组成草原的变量之间独立或变量之间的关系可以被忽略。这与传统的多元分析方法不同。在传统的多元分析中, 我们一般是通过随机取样, 假定样本独立, 然后在样本空间分析变量之间的关系, R-分析。但是, 在草原监测问题上, 我们需要假定变量独立, 然后在变量空间研究样本(观察值)的空间分布和时间分布(Q-分析)。鉴于草原上的每个物种都有可能携带草原演替的信息, 而且我们认为物种间的“互相关”要小于“自相关”, 加之种间的相关关系目前还不能定量描述, 超球面模型忽略植物种之间一对一的关系, 而令样本的动态为时间的函数。模型提出变量独立所根据的事实是草原植物种间的生殖隔离。在草原动态监测中, 我们把“生殖隔离”的意思延伸为: 种群的增长只和物种本身及资源这两个因素有关, 而和其他的与之分享资源的另一个物种无关。其他

趋势

势分析 及其在生态股市中的应用





物种的影响只有联合起来通过资源竞争，通过环境压力起作用。我们说过，超球面模型的关键是标准化。标准化把各分量的平方和的算术根放在分母，表明变量的增长与环境压力（所有物种的集合）成反比；同时使各分量的向量和等于1，这样来描述资源共享这个事实，刻画草原各组成成分之间的关系。

用这里介绍的等比系列法解决多元系统的时间动态问题，一般来说，虽然强调了初始和结束的状态，却把中间过程忽略了。对于逐年都要观察取样的草原监测来说，由于比向量产生于连续两个时刻的草原状态： $\lambda_j = Y_j/Y_{j-1}$ ，没有中间过程，问题不大。只要我们坚持逐年监测，逐步分析，超球面模型应该是可行的。对于其他领域的多变量时间系列，比如，证券数据，一个比较被认可的解决办法是使用滑动平均，让每一个时刻都有作始作终的机会。在我们的投资实验和模拟投资实验中，我们使用滑动平均，逐日对周趋势进行分析，取得了很好的效果。然而时间系列法也有它的弱点，它会带来时间滞后的问题，因为我们假定系统有惯性，使用时间  $K$  的状态去预报时刻  $K+1$  的状态，这是时间系列法的先天不足。对此，我们提出了一个可能的补救办法：使用“二阶比向量” $(\lambda\lambda)_j$  描述空间3点  $Y_j, Y_{j-1}, Y_{j-2}$  的关系：

$$\begin{aligned}\lambda\lambda_j &= \lambda_j/\lambda_{j-1} = (Y_j/Y_{j-1}) / (Y_{j-1}/Y_{j-2}) \\ &= (Y_j \times Y_{j-2}) / (Y_{j-1} \times Y_{j-1})\end{aligned}$$

如果二阶比向量的分量  $\lambda\lambda_{i,j}$  大于1， $(Y_{i,j} \times Y_{i,j-2}) > (Y_{i,j-1} \times Y_{i,j-1})$ ，两个端点的比例中项大于中点，那么它是凹型的；对于上升的种表明上升的速率在增加，对于下降的种表明下降的速率在减慢。如果  $\lambda\lambda_{i,j}$  小于1，它是凸型的；对于上升的种表明上升的速率在减慢，对于下降的种表明下降的速率在增加。这样便能起到预警的作用，弥补了时间滞后的不足。

为了缩小预报误差，也可以考虑在草原监测中使用卡门滤波。

$$E_{k+1} = \alpha \times D_{k+1} + \beta \times E_k \times \lambda_k \quad 15-1$$

$$E_{i,k+1} = \alpha \times D_{i,k+1} + \beta \times E_{i,k} \times \lambda_{i,k}$$

$$i=1, 2, \dots, m; j=0, 1, 2, \dots, K \quad 15-2$$

其中， $E_{k+1}$ 是我们对草原状况在  $k+1$  时刻的期望值， $D_{k+1}$ 是我们在  $k+1$  时刻的取样观察值，而  $E_k \times \lambda_k (= P_{k+1})$ 是我们使用已知数据在时刻  $K$  对时刻  $k+1$  的草原状态所作的预报值。两个新的标量参数， $\alpha$  和  $\beta$  分别是我们给观察值  $D_{k+1}$  和预报值  $E_k \times \lambda_k$  的权重。卡门滤波把历史数据和观察数据结合，能有效地降低预报误差。关于二阶比向量（二阶趋势值）和卡门滤波的工作我们做得还不很多，有待继续研究。

在这里，我们需要指出草原监测的一个误区。由于现代科技的高度发展，许多人误以为我们可以从几千公里以外对草原实行监测。如果我们要监测的仅仅是草原的生物量，或草原的面积，草原变化的结果，这种想法是对的。但是，如果我们需要监测的是草原组成成分的变化，目的是要在量变的早期就感知草原的质的细微变化，探索草原变化的原因，则一定要使用人工监测。因为，草原成分的变化，最终依赖于人们对植物繁殖器官的形态的判断，而这却是任何基于光谱的分析手段都无法胜任的。由于草原监测要使用草原分种（或生态种组）多度，所以人们只能使用最原始，然而也是最可靠、最本源的、依靠当地群众的人工监测为主要手段；要建立类似于气象监测网、水文监测网的统一的草原监测网，而辅之以遥感等现代科学技术。

（2001年发表于《植物生态学报》25卷第6期）

趋势

势分析 及其在生态股市中的应用



## 关于《超球面模型 应用于草原监测数据分析的 探讨》一文中几个问题的讨论

**植声：**在结论与讨论部分，“使用向量的前提假定是组成草原的变量之间线性无关”。无论从种间竞争还是互惠共生的角度来看，种间都不可能是无关的，就拿生物量来说，各个种的生物量之间负相关的可能性很大。

**白·图格吉扎布：**变量独立的假定是超球面模型的基础，它引发了很多争论。我们在使用股票市场数据的文章中曾对此展开了比较充分的讨论。超球面模型的重要前提假定是变量独立。一般认为，在低维空间里无解的股市监测问题，在高维空间里可能有解。而建立多变量空间的前提条件是变量之间独立。变量独立，坐标轴才能互相垂直，以上的讨论才能展开。MDSM用多元向量来表示股市状态的前提是变量之间无关，而忽略变量之间的相互作用。虽然这样要丢失部分信息，引起分析误差，但我们的实验表明，这毕竟提供了一个解决问题的办法。而且，在现实的证券市场中，虽然各个证券互相影响、互相制约，但由于它们之间根本的关系是互相竞争，每个证券都保持自己独特的运动形式、运动规律，所以，直到能够在  $m$ -空间里定量地确定证券之间的关系，而且表示这种关系的值超过某个预定的阈值前（一般地，MDSM把这个阈值定为 0.7，是 45 度角的余弦值），证券之间无关的假定是应该成立的。这里，

“在多维空间”是关键的条件状语。我们认为，证券之间的相关，部分可能是通过市场传递过来的。从式 3-1（见《超球面模型及其在股市分析中的应用》）我们可以看到，所有的证券价格都和市场总趋势相关。那么，如果不排除市场的因素（不标准化数据，不向超球面投影），证券相关的结论也许是不正确的。特别需要指出的是，数学里有条定理： $n$  维空间中最多只能有  $n$  支线性无关的向量。也就是说，如果在二维空间中研究 3 个以上的变量，则结论必定是相关，而不论它们是否真正相关。所以，对于变量独立的问题，我们只接受在  $m$ -空间里的研究结果。如果能够在  $m$ -空间里证明两个变量相关，则我们可以归并这两个变量，去掉一个坐标轴，然后，继续在  $m-1$  维空间里研究系统和变量的动态。MDSM 假定变量之间无关，只是作为我们研究讨论的出发点，而不是结论。从另一方面来说，由于向量合成分解的平行四边形法则，任意向量可以被分解为矩形（特殊的平行四边形）。也就是说，任意向量可以被分解为两个互相垂直的向量。加之， $m$ -空间对坐标轴的数目没有限制，这样，扩展的向量分析就为建立  $m$ -维证券空间提供了依据：即使变量之间有相关关系，只要相关系数没有达到 100%，我们就可以利用其不相关的部分，建立一个新的坐标轴。

我们也可以草原上马和牛的关系来解释。草原上的马群数量基本上是由年轻骡马的数量和草场资源来决定的。如果草场资源不够，无论是由牛造成的，还是由羊、鹿造成的或火灾、人类开垦造成的，对马群来说是一样的。把牛、马共享草场资源的现象理解为马和牛负相关是值得商榷的。马和牛的貌似负相关是由“有限资源”造成的，并不是马和牛的自然属性。

从模糊数学的观点看，种间相互关系是个模糊概念，在定量分析中需要确定一个 0—100% 之间的值参加运算。如果无

趋势分析 及其在生态股市中的应用



法提供准确的量值，则我们要在 0 和 1 两个值中取一个值作为近似。我们初始取值 0，没有考虑种间关系。但据式 15-2，

$$E_{i,k+1} = \alpha \times D_{i,k+1} + \beta \times E_{i,k} \times \lambda_{i,k} \quad i = 1, 2, \dots, m; j = 0, 1, 2, \dots, k$$
在每一步预报后，我们都用观察值来校正由我们（变量独立）的假定和方法所带来的误差，所以，从整个过程来说，模型也包容了变量之间的可能的相关。

**植声：**“可以把向量的标准化看作是多维的百分比”，这句话不确切。尽管这里的标准化与百分比都是相对量，但百分比时和为 1，而标准化的各分量之和并不等于 1。你可能会说它们的几何和为 1，但上述比喻容易引起误解。

**白·图格吉扎布：**为了消除可能的误解，我们把标量的百分比和向量的标准化做一番比较，比较两者的区别与联系。两个同名数 3 和 4 相加，和是 7。究竟 7 是怎么来的，是  $3+4=7$ ， $1+6=7$ ，还是  $4+3=7$ ？我们无法知道。为了说明和的组成，我们用百分比来表示： $3/7=43\%$ ， $4/7=57\%$ 。而且，各分部的和等于总体： $43\%+57\%=1$ 。在超球面模型里，当不同的分量相加时，为保持组成成分的信息，两分量垂直相加： $3^2+4^2=5^2$ （勾股定理）。并且用空间不同的点来表示组成不同的 5（ $3^2+4^2=5^2$  和  $4^2+3^2=5^2$  是空间不同的点）。对应于一维的百分比，向量标准化（除以向量长度）的实质是求余弦值： $3/5=0.6$ ， $4/5=0.8$ 。余弦值表示向量与相应坐标轴的关系。在多维射影空间（ $m$ -Projective Space），余弦值既可以用来表示向量的指向，也可以用来表示系统的相对组成。对应于一维的“各百分比的和是一”，在多维空间，所有的余弦值的平方和是一。 $0.6^2+0.8^2=1$ 。综上所述，百分比和标准化，一个除以代数和，一个除以向量和，分别用来表示一元系统和多元系统的组成。

**植声：**从本文后面的分析看，标准化并不是关键的，因为

预测得到的变量并不是标准化的，那么对初始向量的标准化有什么意义呢？

白·图格吉扎布：在第2部分开始不久我们便强调，对初始向量的标准化，使不同来源、不同尺寸、不同指标、不同年代的样方可以比较、可以归类。使向量（1，2，3）、向量（2，4，6）、向量（10，20，30）可以归并到一起（以上3点在多维射影空间共线），并且使盖度、多度、产量等不同数量指标都转换成相对量。这样，我们才有可能建立草原样方数据库，实现数据共享，推动草原科研走向社会化生产。

我们在文章里没有强调标准化的另一个功能是滤波器（Filter）。标准化滤去了年降水量等外部条件对植被的影响，仅保留了植被内部各物种对降水量的反应，突出了物种对应系统的相对运动。在另外一个层次上来说，标准化使我们得以认识多维空间，否则我们很难捉摸多维空间中无着无落的点。好比星空投影到穹隆，是我们认识星空的基础，外部世界投影到我们的眼球底，是我们认识外部世界的基础。标准化把多维空间的点投影在单位超球面上，是我们认识多维射影空间的基础。

植声：“使用几何回归，在增长时，增率取值从一到无穷大，能够解释种群爆炸的现象；在减少时，……可以解释生命的顽强性”，由于向量已标准化，所以这里的比率不再是种群增长速率，故这里的解释是不成立的。

白·图格吉扎布：一方面，并不是唯有种群增长速率才能描述种群增长。在我们的文章中已经提到多个量，如增量  $\Delta$ 、增率  $\lambda$ 、增率的增率  $\lambda\lambda$  都可以用来描述种群增长。另一方面，我们觉得用指数增长、几何增长来描述种群增长更贴切，更符合生物增殖，复制（Replicate）自身的本性，而不在意它是相对数还是绝对数。

植声：在结论与讨论部分的第一段，作者讨论了超球面模型与

## 势分析 及其在生态股市中的应用



传统思维的不同，在于“在物理世界，化学世界里，科学研究的目的是发现变量之间的关系”，而球面模型提出变量独立，这一点我认为是作者理解上的偏差，或者是把两个相比较事物的内涵和外延都没搞清楚，事实上分量独立在欧氏空间等其他正交坐标系都有，作者所研究的也是变量之间的关系，只不过是草原的状态与时间的关系而已。

**白·图格吉扎布：**谢谢您指出初稿中的不妥提法。根据您的意见，我们已经在定稿中作了相应修改。我们认为把草原状态作为时间的函数来处理实在是没有办法的办法，是无法明确多元系统因果关系时的一个补救的办法。我们同意分量独立、正交坐标系、多维变量空间、射影空间等在数学上确实都早已有之，似乎接受变量独立的假定不应当有很大困难。但事实上，我们遇到最大的挑战正是变量独立的假定。我们自己是花了将近13年的时间（1984年发表系统状态转移矩阵的通解是对角矩阵，1997年发表向量除法乘法的定义）才完成从对角矩阵到多元向量的跨越的。

鉴于变量空间和变量独立的问题始终是超球面模型的最有争议的环节，我们下面从变量空间和样本空间对照的角度进一步阐述一下变量独立的假定。传统的多元分析方法是在样本空间里研究变量。人们通过随机取样实现样本独立（顺便问一句，样本之间真的独立吗？），得以建立样本空间。但是，草原监测的问题有其特殊性。它要求我们在变量空间研究样本（即所谓的Q-分析），研究样本在变量空间的时间分布和空间分布（草原演替的问题是样本在变量空间的时间分布问题，草原分类的问题是样本在变量空间的空间分布问题。两者都要求我们建立变量空间）。建立变量空间要求变量独立。我们的处理方法是把变量作为资源共享系统的元素，而把所有元素的“向量和”作为系统。在系统的大背景下，元素之间的关系被削弱

了,使我们得以假定:在资源竞争系统中,一个元素只与系统(所有元素的向量和)相关,而与另一个元素的关系有待指证。在指证前,默认值是零。当特别必要时,我们可以对物种的相关做预处理:在样本空间里研究物种之间的关系。当物种间相关系数大于一个预定的阈值时,我们认定相关关系成立,小于阈值时相关关系不成立。

从统计的观点,2元素之间相关的概率和系统的元素数目(另一个比较准确的说法是和2的元素数目次幂,  $2^M$ )成反比。例如,当系统的元素数目增加到10时,2元素间相关的概率降低到0.1(或2的10次幂的倒数,约0.001),小到可以忽略不计。

**植声:**最后一段指出的草原监测的一个误区也是值得讨论的,如果说不用遥感等手段,大尺度的问题和空间动态就很难实现,在这一点上作者应该考虑不同时空尺度,必须有尺度外推(Scaling up)。

**白·图格吉扎布:**我们文章的主要意思,不是说不用遥感,而是说不能轻视、畏难或放弃人工地面监测,草原成分的监测。草原退化的根本原因是过度利用,而不是什么“厄尔尼诺现象”,或气候反常。我们同意,草原监测应有不同的层面或尺度,量的监测是必要的,但我们认为应当更重视质的监测、草原成分变化的监测。因为量是果,质才是因。所以草原的退(进)化,首先体现在成分的变化,然后才是产量、面积的变化。如果我们想在发生质变(无论进化或退化)的早期,检测到草原的细微变化,就必须侧重于成分的分析,必须进行现场监测。我们也同意,与人工地面监测网比较,遥感速度更快,效率更高,成本更低,更容易操作,但它是末,人工地面监测才是本。我们主张两者结合,而不同意本末倒置。

(《植物生态学报》25卷第6期)

20  
趋势分析 及其在生态股市中的应用





## 发展骆驼，保护草原， 加固中华民族的生态屏障

2002年8月4日，部分在美国亚利桑那州图桑市参加美国生态学会年会 ESA02 的中华海外生态学者聚集在哥伦比亚大学亚利桑那生物圈2号 (BioSphere-2) 会议室，对中国北部的沙尘暴 (Dust Storm) 问题展开了研讨。

会议由中华海外生态学者协会 (SinoEco) 主席缪世利博士主持，亚利桑那生物圈2号科研部负责人林光辉博士做了中心发言。与会者在会上展开了讨论，学术交流在会后继续进行。

部分学者从生态位的角度对沙尘暴现象的原因——草原退化进行了有意义的探索。“生态位 (nich)”是说，每个生命现象，无论物种或植被，都在变量空间占有特定的位置和空间。按生态位理论，如果把森林、草原、荒漠以湿润度从100到0排列的话，它们的生态位可以简单图示如下：森林以75为中心，分布在100—50，草原（草本植被）以50为中心，分布在70—30，荒漠以20为中心，分布在40—0。人类的生态位当在森林和草原之间，中心大约在60，分布在80—40间。在这种模式下，人类和荒漠的生态位基本不重叠，在人类和荒漠之间，有草原作为生态屏障。但是如果把人类活动的范围扩大，生态位向上、向下各扩大10，扩大到90—30，而由于人类活动加剧，草原的生态位缩小10，仍以50为中心，分布缩小到60—40，荒漠的生态位相应地向上扩大10，分布在50—0。则

人类和荒漠的生态位将发生重叠，人类社会将感受到荒漠化的冲击和压力，如沙尘暴。这种学术观点认为，沙尘暴现象的原因在于草原退化，草本植被退化，因此防止沙尘暴的重点在于保护草原，而不只是植树造林。绝大部分的草原地区并不适于植树。这个观点可以解释植树造林防治沙尘暴效果不很理想的部分原因。

持有这种观点的学者认为防治沙尘暴的重点在于保护草原，而草原退化主要是由于畜草不平衡，过度放牧造成的。因为从讨论会上显示的图片看，围栏封育的草场，远离水源的无水草场都没有严重退化，或基本没有退化。草原，作为生态系统，一个很重要的原则是畜草要平衡。“畜草平衡”是说畜和草要保持一定的比例。这个比例不是固定不变的，但要相对稳定，是动态平衡。宏观的畜草平衡可以简单地用下式表示：

$$S(a) = S/N$$

其中， $S$  是草场资源总量， $N$  是消费者（家畜）总量。 $S(a)$  是畜均资源，它等于资源总量除消费者总数。对特定的草原生态系统，它接近一个常数，或围绕一个常数波动。当草场资源匮乏时，人们自然会考虑投入种植业以增加资源总量  $S$ ，但事实上，由于草原畜牧业的分母基数很大，削减总头数  $N$  才是保持畜草平衡更简单、更有效的办法。而且由于牲畜有指数增长的趋势，如果人们不主动有意识地限制头数，则资源的有限增长，很快就会被消费者的增殖抵消掉。也就是说，人们种草的速度永远赶不上千百万牲畜增殖的速度，永远赶不上千百万牲畜日夜啃食的速度。因此，他们认为，从畜草平衡的角度看，“种草养畜”的提法不科学。用“种草养畜”的思想做指导，无法掌握草原畜牧业生产的主动权。生态学中的食物链理论也不支持“种草养畜”的提法。从生态系统能量传递的角度，草原畜牧业的生产流程图可以表示如下：



趋势分析 及其在生态股市中的应用



日光能→草场→家畜→畜产品→市场。根据食物链理论，能量在传递过程中有“十一律”。也就是说，在食物链的能量传递过程中，传递到下一级的能量只有10%左右，其他90%的能量都用于维持本能量级的休养生息了。换句话说，如果人们在市场经营管理上的投资效率是1:1的话，在家畜改良，畜群结构上的投资效率则是10:1；而在草原建设，种草养畜上的投资效率就只有100:1了（当然种草的生态意义和草产品直接进入市场的情况不属于我们现在的讨论之列）。

所以发展草原畜牧业的科学原则应该是：“以草定畜”。养畜要“少而精”，不追求数量，而以提高质量为主，提高畜均产值为主。为了更好地贯彻“以草定畜”的原则，建议中国政府比照气象、水文的作法，建立全国的草原监测网（这是一个很大的系统工程，有待进一步深入讨论），了解自己的家底，了解自己家底的动态。

与会者对人民日报（海外版）7月16日的报道——“沙漠之舟面临搁浅，草原骆驼哀鸣声声”，表示了极大的忧虑。在牛—马—绵羊—山羊—骆驼（从湿润到干燥的生态位排列）中，骆驼忍饥，耐渴，善走，放牧半径大，能充分利用无水草场，是最适合中国西北草原（荒漠）区发展的家畜。骆驼最重要特性是抗灾能力强。可谓是“大风3天刮不跑，大雪3天饿不倒”。现在，在短期经济效益的驱动下，竟遭大规模屠宰，令人痛心。牲畜所能利用的草场资源不仅和草原的草场资源密度有关，而且和牲畜的最大放牧半径有关。家畜的资源利用公式可以简单表示为：

$$S_{(i)} = D \times R^2$$

其中有效资源  $S_{(i)}$  是资源密度  $D$  和放牧半径  $R$  平方的乘积。当资源密度低时（如退化草原、荒漠），扩大放牧半径，以致游牧，是发展畜牧业并保护草原的有效措施。单从市场短

期利益出发，发展放牧半径小（按放牧半径排列，山羊—绵羊—牛—骆驼—马），啃噬草场严重的山羊头数是饮鸩止渴。只会加速草场退化，加剧沙尘暴的肆虐。

他们建议中国政府重视草原退化问题的研究。使用税收杠杆，比照石油输出国组织（OPEC）的榜样，对羊绒产业施行限量发展。同时提高骆驼绒的收购价，保护发展骆驼。拨款对骆驼产品（主要是驼绒）进行开发和深加工、配套市场的研究。他们相信，经过千百万年的严寒、酷暑、干旱环境育成的双峰驼，肯定保有极大的科学价值和经济价值，有待人们去进一步探索、开发。

## 后 记

SinoEco 没有讨论滥垦（草原）、滥挖（药材、黄金）等破坏土壤结构，从而直接导致沙尘暴的问题。因为那已属于违法行为，属于罪与罚的范畴，不在海外学者讨论之列。

（2002 年发表于《植物生态学报》）





## 论蒙古包在生态学方面的意义

### ——为蒙古包申请世界 文化遗产而呼吁

在蒙古、藏、印第安、阿拉伯等游牧民族中，唯有蒙古族发明使用蒙古包（*ger*）。游牧民族的可移动民居有共同点：木头支架，轻便，不用浆糨结，减轻自重，且摆脱了对水的依赖，容易装卸，驮运。此外，蒙古包有它自己的特点。最明显的特点就是它的外形：包，包形。下部圆筒形，上部圆锥形，顶为蛋壳形（其剖面为拱形）。我们现在知道，在材料相同的条件下，球形有最大的容积和最小的散热面积。其次是圆筒形。而锥形最稳定。如果同时考虑到稳定性、坚固性、高度、容积和表面积，则圆筒形、圆锥形，加蛋壳形所形成的包形是最佳的结构：稳定性最好，散热面积最小，包内可利用容积最大。而且最坚固：由于蒙古包的侧壁呈圆筒形结构，没有平面，没有棱角。受外力撞击时，很容易把力传递出去。蒙古包的顶呈蛋壳结构，能够承担比较大的正压力，所以蒙古包的结构也是最坚固的结构。各种形式的帐篷，虽然都能满足自重轻、驮运搭建方便、稳定性好特点，但即使是比较现代的帐篷也多数为锥形、方形，而鲜有扣碗形，也就是包形的。

## 一、蒙古包的扣碗形结构 及其特点功能分析

屋顶，也称天窗，蒙语称“套努”（ᠰᠠᠨᠨᠠ）。 “套努”的外形好似一个倒扣的大箩筐。正面看，像轮子，轮子的四根辐条呈十字排列。辐条之间的空间相当于天窗，不仅用来透光、通气，并用来通烟筒。“套努”从侧面看是个拱顶，立体看似蛋壳结构。蒙古包主要由“套努”形成的蛋壳（拱形）结构来体现它的优越性。“套努”的作用相当于房子的大梁、托柱，支撑椽子（蒙古包的“乌尼”杆（ᠤᠨᠢ））。一般建筑物对大梁都要求很高。椽子折几根没有问题，但如果大梁断了，整个建筑物就倒塌了。由于大梁有强度要求，一般有较大的自重。而蒙古包的“套努”是圆形的，是拱形的。像个反扣的大箩筐，没有什么特别的强度，自重也很轻。关键是它采用了蛋壳结构，所以它在强度不很高的情况下仍能稳固地支撑起蒙古包。

“套努”的风帽作用：由于蒙古包顶的蛋壳结构，风无论从哪个方向来，都会在蒙古包的上空形成一个低压区，所以蒙古包的炉子特别抽，特别好烧。即使是比较复杂的、烟道回转270度的回龙灶（蒙语称“哈伦灶和”（ᠬᠠᠯᠠᠨᠵᠠᠬᠡ/ᠰᠠᠨᠨᠠ）），烧比较有潮气的羊砖，也绝对没有倒烟的现象，保证了供暖的可靠性。我们现代人一般用鼓风或抽风的方法来解决炉灶的供氧问题。鼓风法就是用风箱或吹风机向炉灶内鼓风供氧。但由于鼓风造成的正压，如果烟道不严密、不通畅，则容易漏烟、溅火星，有安全隐患。另外一个供氧的办法是抽风法。现代人利用抽风原理的一个重要形式就是高烟囱，用高大烟囱内的热空气柱的升力造成负压，为炉灶抽风供氧。因而，高大的烟囱曾经

趋势分析 及其在生态股市中的应用



一度成为现代工业的象征。抽风的另一个形式是风帽。在烟囱上装一个伞形的装置。这个装置不是用来防雨的，而是利用风力来产生吸力的。古代蒙古族游牧民们把自己的居所构造成包状，将烟囱从包顶伸出来，巧妙地将抽风功能糅合在蒙古包的构架中，从而摆脱了累赘的鼓风设备和高大的烟囱，也避免了使用鼓风带来的火灾隐患。蒙古包为古代牧民提供的使用火、驾驭火的能力，让我们现代人也佩服。再分别介绍乌日和、乌尼杆和哈那。

**乌日和** “套努”的上面有一块方形的幔毡子，蒙古语称“乌日和（ $\text{Уриг}$ ）”。在四边形的四个顶点，有四条绳索分别搭向左、右、后、前面，其中搭向后、左、右的三条是拴在蒙古包的横绳上相对固定的。而搭向前面的绳索是活动的，用来控制采光和通风。白天，把绳索拉向后面，使“乌日和”对折，露出半个“套努”，可以采光通气；夜晚，把绳索拉向前面，用“乌日和”幔罩住“套努”。关闭天窗，以便保温。

**乌尼杆** 相当于房屋的椽子。相同长度的木棍，比大拇指稍粗，一端稍尖，插入“套努”的母眼，另一端有细绳套，套在哈那头上。多根乌尼杆支撑屋顶（套努），由于“套努”的作用，乌尼杆受的是顺长轴的压力，而不是剪力。俗话说“立木顶千斤”，即使乌尼杆不粗，蒙古包的特殊结构使它能承受较大的压力。在雨天，包顶的毛毡全湿的情况下，蒙古包也不会被压塌。冬天大雪也不怕。在有限的材料下，蛋壳结构为蒙古包提供了最大的抗压性能。蒙古包的抗压性能是由蛋壳结构来实现的。所以，使蒙古包顶保持正常的蛋壳结构是保证蒙古包挺立的关键。如果由于某种原因，蒙古包的“套努”的一个角陷进来一些，这变形的蛋壳结构会造成应力分布不匀，严重影响蒙古包的坚固性。遇到这种情况，牧民们会用蒙古包内专有的一根立柱，蒙语称“图力古尔”（ $\text{Тугилгур}$ ），把陷进来的一

角顶上去，让“套努”回复水平状态，蒙古包顶保持正确的蛋壳形状。把“图力古尔”翻译为“立柱”纯粹是为了汉语行文的方便。“图力古尔”的功能是纠偏，纠“套努”之偏，而不是承重。它既不立（它的姿势往往是斜的），也不柱（粗不及手腕）。

**哈那** 菱形的木框架，可以方便地调节蒙古包的高度和周长。哈那眼近方形为佳。夏季高耸，图凉快，冬季略低矮，保温，多容纳过夜的人。蒙古包的哈那（ $\text{hāna}$ ）成双数，4或6，便于驮运。

**哈那、乌尼杆和套努** 构成了蒙古包的骨架。保持蒙古包的蛋壳结构，不但保证蒙古包有最大的抗撞击能力，而且具有很强的抗风性能。荒漠草原上的风灾是很频繁的。草原上经常发生狂风把土木结构房子的屋顶掀翻，把帐篷刮跑的事故，但我们却从来没有听说蒙古包被狂风吹塌、掀翻的故事。千百年来，草原上多风的生态环境为蒙古游牧民们淘汰了备选的居所，而保留了扣碗结构的蒙古包。首先，由于蒙古包是圆形的，无论风从哪个方向刮来，蒙古包都没有迎风面。圆形化解了各方向来风的正压力。更重要的是，对于一般的房屋，当狂风吹过屋顶的时候，由于流体力学中的柏努利原理，会对屋顶产生向上的吸力。一般人们用增加屋顶自重的方法来固定屋顶，或用牵拉的方法来固定帐篷。蒙古包另有自己的抗风结构：蒙古包的“套努”是露天的（当“乌日和”掀开时），不但有透光、通风的效果，而且有抗狂风的功能。狂风产生的对包顶的上升的吸力会被“套努”泄出的气流所抵消、破坏。如果狂风更加肆虐，包棚顶（蒙语称“德布日”）显得有点轻飘时，使用蒙古包内从“套努”上垂下的一条绳套，蒙语叫“恰哥塔格”（ $\text{chagotag}$ ）的，吊上一个重物，如米袋子、水桶即可。这是老人或孩子在包内就可以从容完成的操作。由于蒙古包的





扣碗结构、“套努”的外泄气流，加上“恰哥塔格”，有个百来斤的米袋，就能抵抗并化解狂风能够摧毁其他土木建筑的千钧力量。独特的结构使蒙古包扣住大地不放松，任凭大、强、狂、烈的风。

## 二、蒙古包作为世界文化遗产的现实意义和生态意义

“逐水草而居”的游牧方式是历史形成、自然选择的结果，是维护草原、延续民族的重要条件，现在看来甚至可以说是必要条件。生态学的一条重要规律是“十一律”（又有称金字塔、食物链的，但笔者以为“十一律”更传神、更量化、更准确地阐明了“等比级数”而非“等差级数”的关系。金字塔表示的是“等差级数”），说的是能量在自然生态系统中不同能量量级之间传递的比例大约是 10:1。根据“十一律”，为保证草原生产力可持续利用，在自然状态下，只可以把十分之一的草原生产力转化为草食动物。换句话说，为保证草原的永续利用，牲畜只应该利用草原生产力的十分之一。古代的蒙古牧民是深谙自然保护的精意的。他们称自己的游牧为走“敖特尔”，是为了让牲畜“采草尖”（请注意：不是吃，更不是啃，而是采，采蜜的采。据说，英语的 GRAZING 也有轻擦的意思）。他们之所以搬家，不是因为牲畜把草吃光，没有草吃了，为了“逐草”而搬。而是游牧文化的传统和风俗习惯使然。当然，从农耕文化的角度来看，草原上仍有 90% 的生产力没有被收割、没有被利用，是很大的浪费。农学院培养的草原工作者依照颗粒归仓的原则，是通过把草剪光来估算草原载畜量，以期“尽草而牧”，是造成畜草严重失衡的技术原因。生态学原理和

游牧民族千百年的实践告诉我们，保留 90% 的生产力是使草原得以休养生息，从而达到永续利用的根本保障。而使牧民定居、家畜定牧却是草原退化和沙尘暴的一个主要原因：定牧后家畜的可利用草场资源被“放牧半径”局限在水源和定居点附近。因此造成过度放牧，引起草场退化，畜质退化，进而引发、导致“沙尘暴”。古代美索不达米亚肥沃的平原如何会演变成今日的伊拉克荒漠，学术界普遍认为是土地过度利用的恶果（革命导师恩格斯对此已早有论述）。而从电视上看，伊拉克并不是没有树，也不是没有水，“美索不达米亚”的原意就是“两河平原”。

在形而上学者看来，“逐水草而居”的游牧生产方式、生活方式是古老的、粗放的，因而是落后的。他们很乐意看到游牧被定居所取代、畜牧业被种植业取代、农业被工业取代。诚然，工业的单位产值（人均、单位土地面积、单位时间）高于农牧渔业。但问题是，无论工业如何发达、如何现代，在能够人工合成食物之前，我们仍然要靠农、牧、渔业来提供人们的最基本的生活需求——食物。我们现在不应该考虑如何取代牧业，而应该考虑如何用现代工业来武装农、牧、渔业，用农牧渔业的现代化来解决持续的食物问题。

牧业的现代化要两条腿走路。现代化牧业要由草原游牧和近水垦殖（种草养畜）共同组成。好比近海养殖不能取代远洋捕捞一样，种草养畜不能代替草原游牧。中国的基本国情是，由于历史的原因和人口的压力，所有能种的、该种的地已经都种了，许多能种但不该种的地也已经种了。剩下从事草原畜牧业的地域，科学的分类是荒漠、半荒漠或干旱草原，是非常脆弱的生态系统。比较科学的主要的利用方式唯有轻度放牧或游牧。游牧现象有很丰富的科学道理和文化内涵，需要我们去挖掘、探索、认识。游牧现代化，蒙古包有重要的参考价值。也

趋势分析 及其在生态股市中的应用



许我们的牧业现代化可以不用蒙古包。好比现代的远洋捕捞不用舢板，而用装备了现代化通讯、导航、捕捞、加工设备的巨轮。现代化的游牧可以使用装备了动力、通讯、家庭电器、现代炊具、餐具、寝具，装备了履带的房车来代替蒙古包，让用科学技术武装了的现代化游动牧民在放牧的同时兼起防火、防盗、守土的责任。由定居的牧工完成畜产品的深加工来增值。用“十一律”来指导“畜草平衡”，进而施行游牧现代化是防治沙尘暴的根本措施，因而是关系到国家生态安全的重大研究课题。而科学合理、简易轻便、坚固保暖的蒙古包作为中华大家庭古代游牧民族合理利用自然资源、自觉保护生态环境的历史的民居形式，是重要的世界文化遗产，是中华民族对世界文明的重大贡献，是人类合理持续利用资源方式的奇葩，值得我们自豪。我们要通过申请世界文化遗产把它的文化内涵发扬光大。

(2005年发表于《呼和浩特科技》)

## 中国沙尘暴起因之我见

### ——在草原生态与游牧文明研讨会上的发言

2005年9月11日，星期日，由国家民委铁木尔委员主持，民间人士马晓力女士赞助的《草原生态与游牧文明研讨会》在中央民族干部学院第四会议室召开。参加研讨会的有专程从内蒙古呼和浩特市赶来参加会议的草原文化专业的学者、美国福特基金会官员、美国 MDSM 数据分析公司人员、民族报记者、中国民族杂志社编辑、民族出版社编辑、中国青年报记者、苍狼白鹿文化传播公司人员、中国社科院科研人员、中央民族大学教师及研究生、国际广播电台播音员、当年曾经在内蒙古牧区插队的知识青年等多人。由美国 MDSM 数据分析公司白·图格吉扎布先生在会上的发言。整理如下：

谢谢主持人，铁木尔委员。

很高兴认识这么多朋友。听说大家都很关心中国的草原和草原文化，并做了很多有意义的工作。你们中有的人，自己出钱把《草原法》翻译成蒙文，发给牧民，帮助牧民——一个社会弱势群体，用法律武器保卫自己的家园，保卫他们祖祖辈辈赖以生存的草原，并取得了胜诉。可以想象，那是很不容易的。对方既有地方政府的默许，甚或公开的支持，又趁着市场化大潮的势头。对方既有钱、又有势，既能说、又会道，赤手

趋势

势分析 及其在生态股市中的应用



空拳的牧民能打赢官司，我想全靠了他们的知青朋友。虽说法律是公正的，但法律是不会自动发生效力的，还有那个“曾经草原”网站。我一听就知道可以引为同道。我非常高兴在这里与大家交流草原生态和游牧文明，探讨草原退化和沙尘暴产生的深层、文化层面的原因，与大家分享我多年在这方面的科研成果。

首先，让我们来共同探讨国家生态安全这个概念。中国是有13亿人口的大国，环境保护，自然资源等生态问题是我们必须要面临的重大课题，重大挑战。一位前中国领导人曾经对生态的重要性说过一句话，大意是，当今的时代是民族觉醒的时代。外敌入侵，用武力征服中国，奴役中华民族，已经不可能了。如果中国将来出问题，很大可能是内部问题，如路线问题、民族问题或生态问题。其中尤其以生态环境问题最危险、最严重。如果我们不与大自然和谐共处，走可持续发展、自然资源永续利用的道路，而只考虑当前，不考虑子孙后代，搞掠夺式的经营，我们将遭到大自然的报复和惩罚。而如果真的招来了大自然的惩罚，我们人类将无法自救。历史上许多文明、许多民族的衰落，就是由于违背自然规律，耗尽了他们赖以生存的自然资源。无产阶级革命导师恩格斯对美索不达米亚（古伊拉克）过度地利用土地造成土地荒漠化有过很精辟的论述。在座的有记者、政府官员、杂志编辑。许多人都是学社会科学，搞文的，可能知道恩格斯原话的出处。他的大意是说，人们对大自然的每一次胜利，都遭到了大自然的报复。搞掠夺式的经营，必然会遭到大自然的惩罚。美索不达米亚过度地盘剥土地，因此造成了土地的荒漠化。对圣经或对西方文化有所了解的人都知道，伊拉克在古代是十分富饶的。即使现在，也不缺水，有两条河（美索不达米亚就是希腊文“两河平原”的意思），也有树。但与古代相比，伊拉克现在的自然条件令人们

十分震惊。历史不可重演，使伊拉克恢复到历史上所记载的原状，几乎是不可能的了。而我们也不能让伊拉克的历史进程在中国重演来检验草原退化和荒漠化之间的联系。空间换时间，我们主张以美索不达米亚（古伊拉克）为戒，防患于未然，或如铁委员所说，防大患于已然。中国北方土地荒漠化的问题已经越来越明显，沙尘暴愈演愈烈。我个人认为这场正在发展的生态灾患的成因是复杂的。其中，用农耕文化的思想指导草原畜牧业生产，致使草原植被退化是重要的原因。下面，分几个方面，分别说明。

草原是人类的生态屏障。这是根据生态位提出的一个很抽象的命题，需要进一步阐述。如果相信进化论，认为人是由猿进化而来，从树上下来，来到地面，来到草原。那么，人的生活环境应当介于树与草之间。其实，即使不借用进化论，把森林、人、草原和荒漠的生态位按湿润度排列，我们也会发现，人的生态位介于森林和草原之间。由于现代物质文明，我们可以生活在森林，也可以生活在荒漠。但我们一般都接受，最适宜人类生活的环境还是介乎于森林和草原之间的环境。这样，人的生态位和荒漠之间本来有草原隔开。如果草原萎缩，功能退化，则人类将受到荒漠化的冲击。这完全是根据生态位的概念，对土地荒漠化，对沙尘暴的抽象解释。当然，其他人有别的解释，比如什么厄尔尼诺现象、气候变暖、过度砍伐树木等。

我的这个观点，草原植被退化导致生态条件恶化，土地荒漠化是1984年主持《草原退化趋势预测及对策研究》课题时形成的。当时的背景是，美国发表了《美国2000年》的报告。中国领导人邓小平提出了翻两番的著名论点。为了落实翻两番的战略目标，国家启动了《中国2000年》的课题。各个省、市地区政府各部门都作翻两番的预、测、规划。当时农业部找

趋势分析 及其在生态股市中的应用



到中国农业科学院草原研究所，要求我们承担 2000 年草原畜牧业产业翻两番的预测、预报工作。我答应参加承担这一重大课题，但对农业部的指令，为畜牧业生产翻两番做沙盘推演、提供科学依据提出了不同意见。我根据自己在牧区的经验，提出了草原正在退化的观点。因为，“两个退化”（畜质退化、草原退化）当时已经初显端倪。老牧民反映，现在羊的个头越来越小，马越来越不能跑路；畜群组成上，大畜（改良畜）比例越来越低。放牧场正在从居民点向远处退去，而且，牧草产量、质量也都在下降。我以为，科学工作者因为占有，消耗了比常人更多的社会资源，就有责任为人民坚持科学，追求真理。参加课题论证会的各位老师（包括当时内蒙古省委常委，科委主任许令妊教授的书面意见）支持我的设想，通过了《草原退化趋势分析及对策研究》的论证。在课题进行中，我们搜集了大量第一手的、让我们触目惊心的数据和资料，说明“我国牧区畜牧业的基本生产条件或说生态环境条件正在发生严重的、缓慢而持续的变化；草场有明显的退化趋势，水源、居民点附近退化现象十分严重”。为了说明问题，更为了引起决策人的注意，我们纳入了八级以上大风天数为指标，进行回归分析。我们当时的假设是：作为植被，草原会影响风产生的原动力（太阳能对地表的加热）和风运动的阻力（改变大地和大气界面的摩擦系数）。而且，进一步的分析表明，草原退化和大风天数高度相关。气象数据回归分析的结果表明不但锡林郭勒草原上大风天数有增加的趋势，而且，北京的大风天数也有增加的趋势，并定量列出回归公式。回归公式预测 15 年后，到 2000 年，北京的大风天数将达到每年 60 天。当时，也有其他人员指出了这个趋势。如人民日报当时有一篇文章——“风沙逼近北京城”。但是，当时的人们普遍不能接受。为此，还有人专门上课题论证，说明北京的尘降不是来自内蒙古，而是来自北

京郊区的大兴。意图否定中国北方，北京地区存在土地荒漠化的趋势。但是，事情却不幸被我们的研究报告言中。报告发表15年后，2000年，中国北方发生了特大的沙尘暴，并袭击了北京。除了中国，近代历史上世界范围内还发生过几次规模较大的沙尘暴。比如，美国历史上发生过黑风暴，被认为是由于过度开垦所致。前苏联也发生过沙尘暴，被认为是由于过度砍伐所致。中国的沙尘暴的原因，现在还没有定论，我们认为这是由于错误地开垦和过度放牧造成草原退化所致。滥垦草原的恶果在学术上、法律上已经有了定论，这里就不再重复。我的专业是草原生态学，所以仅就超载过牧、畜草失衡、定居轮牧发表几点看法。

比较直观地说，草原退化就是草原产草量下降。但这还不全面，草原不仅有产量的问题，还有质量的问题，而且，更重要的问题在于质量。听说你们在座的人中，就有人在做草群的组成工作，试图以草群组成来说明草原是否退化。我觉得这很有意义。草原产草量有时高有时低，降雨量高，产量就高；降雨量低，产量就低，随降雨量的变化而变化。但草原组成则是相对稳定的。比如一块羊草草原，被开垦撂荒后，长起了蒿子，产量成倍增加，但牧民并不认为，草原恢复了，草原变好了。对于牧民来说，重要的问题是，它是什么草原，是羊草草原呢，还是冷蒿草原。为了描述草原成分的变化，我把多元向量应用于草原定量分析。我们知道，草原是由多种植物组成的。我把草原看作是多维物种空间的点，好像我们学过的直角坐标系。在平面直角坐标系中，一个点由两个数字，即X坐标、Y坐标决定；在多维物种空间中，草原状态由组成草原的牧草来决定。每种草提供一个数字，有多少个物种便有多少个数字。这多个数字组成一个数组，称为多元数组（或多元向量）。每个数字表示该牧草的量值，整个多元数组表示草原的

趋势分析 及其在生态股市中的应用





状态。两片各组成成分比例相同的草场，称同比草场，两者互为缩影，在多维射影空间共线。可是多维空间在自然界是不存在的，只存在于我们的头脑中，我无法举例来说明，也不准备在这里多谈。

现在，好多科学工作者已经达成了一个共识，草原退化的重要原因是长期过牧超载造成的。那么，草原为什么超载呢？有人说，是牧民有追求牲畜头数的本性：“在游牧民族传统的畜牧业观念中，头数占有非常重要的地位。这是由于在严酷的自然环境下，人们几乎没有能力同自然抗争，较多的牲畜数量是维持生存和生产所必需的”。听起来，似乎有道理，小农经济思想是造成追求头数的政策原因。但是什么人持有这种小农经济思想呢？说游牧牧民有小农经济思想似乎有点牵强，而且，这套说法无法解释，为什么违背自然规律的“追求头数”的游牧民族在草原上生存了几千年，而没有早早被大自然所淘汰？草原为什么只是在最近几十年才急剧退化？我在牧区劳动时发现运动中的畜群是无法计数的，要依赖棚圈，人们才有可能对牲畜计数，而过去个体的游动牧民是没有棚圈设施的。他们是把自己的牲畜赶到一个特定的洼地，根据牲畜的疏密确定当年的牲畜的丰收或歉收（蒙语称：浩特格日都仍）。历史上，游牧牧业一直是自给自足的自然经济，牧民没有理由追求过多超过草原和自己承受能力的头数。而且，根据我做马背教师的经验，大多数牧民至今仍然是文盲，不擅计数。说牧民不擅计数，没有小看他们的意思。牧民其实是很聪明的。他们管理经营自己的牲畜靠“认”，认识的认（蒙语称：究斯冷），而不是靠计数。我们都有这样的经验。班主任老师一进教室，就会发现：“王小毛今天怎么没来学校，谁知道他家出什么事了？”她不是靠点名，靠计数来发现有人缺课。靠计数虽然能知道有人缺课，但却不知道是谁缺课。班主任一进教室就发现王小毛缺

课，靠的就是“认”。说不擅计数，自给自足的牧民追求牲畜头数，是不确实的，是打错了板子。因为这里牵扯到造成草原退化的责任问题，所以一定要说清楚。我个人以为，牲畜超载、草原退化是中国政府政策失误造成的，是忽视草原文化而用农耕文化思想指导牧业生产的结果。20世纪50年代，内蒙古自治区主要领导明确提出，“‘千条万条，发展牲畜头数第一条’，还要‘人畜两旺’”。所以说在牧区发展牲畜头数，其实是中国政府的政策。这不是某个个人的责任，也不是地方政府的责任，而是全局性的，属于中央政府政策的不当，或说是文化层面的差别造成的。政府把农业生产中追求单位土地面积产量最大化的政策搬到草原畜牧业生产，并解释为牲畜存栏数最大，成为指导牧业生产的基本政策。当时，在政府的“头数挂帅”政策推动下，基层领导追求存栏数，争当牲畜头数超百万的大旗大县。为此大搞“抗灾保畜”，给畜群形成了一个向乏弱者看齐的生态压力：哪个牲畜乏弱，就给好吃好喝，哪个肥壮，就杀就卖。哪个畜种繁殖得快（如山羊、绵羊），就大力繁殖。几代下来，牲畜整体质量就自然下降了。这实际上，完全不符合“适者生存，劣者淘汰，优者优饲”的牧业生产规律，也完全违背了游牧牧民传统的做法。在牲畜质量退化的同时，几十年的时间，把草原牲畜头数推到了草原载畜量的极限，甚至逼近或达到了草原生物量。而且，这种“头数挂帅”政策一旦形成，便有极大的惯性。以至于当大家都已经认识到畜草失衡是制约草原畜牧业的主要障碍后，仍然有人坚持发展头数，甚至提出“种草养畜”这样本末倒置的策略。我以为现在还在主张“种草养畜”的人，如果不是蹲在城里的书呆子，就是顽固的农耕主义者，在用“发展畜牧业”的招牌，贩卖“开垦草原”的狗肉。因为稍加匡算，就可以算出“种草养畜”是赔钱的买卖。如果以内蒙古恩格贝的水、热、土壤条件，以

趋

势分析 及其在生态股市中的应用



鄂尔多斯羊绒集团的财力，尚不能维持“种草养畜”，要广大牧区“种草养畜”则纯属欺人之谈（城市郊区、条件较好的农村又当别论）。

与上面的说法相反，我认为，是游牧牧民和他们的草原文化保护了草原。游牧牧民视草原为母亲，本能地保护草原。牧民采用游牧的生活方式和生产方式让牲畜采草尖，不但保证牲畜得到最好的营养，更重要的是使草原得到最适度的利用。用牧民的话说是，“俄布斯乃素日古格，乌苏乃通拉嘎”，我把它译作“采草之精，饮水之清”。牧民跟着牲畜在草原上漫游，让牲畜采食牧草最精华的一小部分，被中原人描述为“逐水草而居”，实际上很不恰当。游牧牧民之所以搬家，不是因为草被吃完了不得不搬家到有草的地方，而是为了保护草原，使草原得到最适度的利用，而主动离开，同时也是为了让牲畜拣吃最好最嫩的草。两种文化的差别，在“盲流”的身上表现得最充分。五六十年代，草原牧区上有从农区来的所谓“盲目流入人口”。他们非常不愿意搬家。到了倒场的时候，总是和生产队的领导争论，“草坡上还有草呢，牲畜还有吃的呢，为什么要搬”。因为他们习惯了农耕文化的“颗粒还家”，认为只有把草吃光了再搬家才不浪费。有句老话，形容一个地方富饶：“棒打獐子，瓢舀鱼，胖胖的野兔钻锅底。”遇到这种地方，有识之士首先是要把它保护起来，以便我们自己 and 我们的子孙后代永续利用。但另外一些有殖民心态的人就不同。他们要来“开发”，直到把獐子都打完了，把鱼都舀光了，把兔子都炖着吃了，才歇心。否则，心里就总有个事似的。

游牧牧民的做法，让牲畜只采食草原产草量的很小的一部分，约十分之一，便主动离开是符合生态学原理的。我们现在讲科学决策，探讨草原退化的原因，就要好好挖掘探索草原文化的内涵及其科学性。比如金字塔、食物链，这些生态学基本

概念大家可能都已经比较熟悉了。今天给大家发的材料里，有世界图书出版公司翻译出版的“英汉双解生态与环境词典”中的一页，解释 PYRAMID，说的是“每一个营养级的生物量大约是其下面营养级生物量的 10%”。虽然按字面，PYRAMID 可以译作金字塔，但我建议按意译作“十一律”比较妥当。因为在这里说的生态系统上下营养级生物量之间的数量关系其实是等比关系，10%，10%，再 10%，往上逐级递减，而不是等差关系。金字塔表现的是等差关系。10 块砖，9 块砖，8 块砖，……2 块砖，1 块砖，每级差 1 块，一直擦到顶上，就形成一个金字塔。生态系统、食物链这些生态学概念传入中国时，我也参加了部分工作。大约是 1979 年，美国科罗拉多州立大学草原生态学实验室创建人，范·达因教授及其夫人访问中国，并到内蒙古大学开展学术交流。已故中国科学院院士李博先生主持接待工作。当时，我是李博先生的研究生。先师让我把范·达因的文章翻译成中文，以便大家了解他的学术思想。范·达因在他的著作中的提法类似剩余价值：自留 90% 用于维持自己的生存，把大约 10% 提供给下一个营养级别。他文章中的示意图看起来根本不像金字塔。比较形象的描述是：一个大办公桌上放着一本书，书上立着一小截粉笔头，分别代表第一性生产力（草原产草量）、草食动物、肉食动物的生物量。它们之间的比例是按照 1000、100、10 画的，而不是 3、2、1。当然，有人会争辩说，把等比级数指数化后可以作为等差级数处理。1000、100、10 是 10 的 3 次、2 次、1 次方，也就是 3、2、1，成等差级数了。当然，这样做，数学上看起来和谐了，但用来指导生产会造成混乱。我以为还是称“十一律”好。

离开“十一律”，就无法解释中国草原退化的问题，就找不到原因，就只有到老天那里找原因，到国外找原因，什么厄尔尼诺现象，什么气候变暖呀，什么外蒙古原因，反正不是我



们自己的错。我们是计算了产草量，计算了草原载畜量的，是按照畜草平衡的思想做的呀，怎么可能超载呢？问题的关键就在于我们没有遵照“十一律”，把草原生产力的90%留给草原供其休养生息。这个问题是个普遍存在的问题，并不是我个人在这里杜撰的。我在做研究生时，曾随中国科学院自然资源综合考察委员会和中国科学院北京植物研究所的科学家们在国家锡林郭勒草原生态定位站实习，后来到中国农业科学院草原研究所工作。在多年的工作中，我和我的导师们、同学们、同事们一直都是用“净草法”来估算草原载畜量的：随机取若干1平方米的样方，贴地面把草剪净，称鲜重烘干，再秤干重。1平方米的干草产量克数就是1平方公里草原产草的吨数。然后，把草原产草量乘以草原可利用面积除以单位牲畜年食量便得出该片草原的载畜量。这是经典的草原载畜量估算法，是能在教学参考书和实验手册上找到的唯一的草原载畜量估算法。然而，大家可能已经注意到，从草原产草量到草原载畜量的转换过程中没有任何扣除，是百分之百地转换。也就是说，我们的“畜草平衡”是建立在“颗粒还家”、“竭草而牧”的基础上的。这样，我们便实际上过高几倍、近10倍高地估计了我们的草原家底，因此繁殖了过多的、近10倍多的牲畜。这样“杀鸡取蛋”、“竭泽而渔”的直接结果便是造成畜草非常的不平衡、草绝对地不够吃。牲畜吃不饱，营养不够，便啃灌木、刨草根，不但造成了草原的退化，而且影响了牲畜的生长发育。我个人认为，草原长期的（几十年）、大面积的、高强度超负荷过牧（900%）导致草原植被退化，才是沙尘暴的真实原因。与这样长期、高强度的超负荷过牧相比，其他因素的影响（如所谓“厄尔尼诺”现象的解释）就显得微不足道了。而且，“厄尔尼诺”现象、“世界气候变暖”等说辞无法解释，既然整个世界气候变暖，何以只有中国北方沙尘暴肆虐、世界其

他各地却相对的平静，也无法解释，何以同样位于中国北方，宁夏沙坡头的风沙多年来呈下降的趋势，而其他地区，如北京呈上升趋势。我本人愿意就此和国内专家们本着百家争鸣的精神，共同探讨。

同时，我认为草原退化的另一个重要原因是中国政府在牧区实行的“定居轮牧”的政策造成的。游牧生活是很艰苦、很简陋的。世界上尚没有一个游动民族实现了现代化的先例。因为财富的积累需要时间、需要稳定，所以，让游牧民定居是牧业现代化的先决条件。让牧民定居是好事，但如何在定居的条件下保持发扬草原文化的传统，保证荒漠、半荒漠草原的永续利用呢？这是个世界历史的难题，至今还没有确定的答案。我看过中央电视台的节目，似乎牧业现代化就是房子、车子、票子。牧民定居了，住上房子了，骑上摩托车了，有钱了，似乎牧业现代化也就实现了。可是草原的状况如何呢？在定居点附近，甚至几里之内，地面光溜溜的，什么也不长。在卫星照片上，一个定居点、一口机井就是一个白点。而且定居点的历史越久，白点就越大，颜色就越白，显示草原退化和定居轮牧是高度相关的。进一步分析说明两者还有因果关系。从定居点早上出牧晚上归来的牲畜都必定有一个最大的放牧半径。最大放牧半径以外的草场，家畜就利用不上了，是为“不可利用草场”。如果以定居点为圆心、以放牧半径为半径做圆，则圆周以外的草场是“不可利用草场”，放牧强度等于零。圆周上的草场，家畜刚刚够得着，放牧强度最轻。从圆周到圆心，放牧强度（家畜停留时间）和距原点的距离的平方成反比。原点处，即定居点附近，距离等于零，放牧强度最大，理论上趋于无穷大，会被牲畜啃为裸地。牲畜啃噬和践踏的双重作用，会导致定居点周围植被退化、土壤变性，以致最终不可恢复。随着时间的推移，裸地会呈同心圆向外逐步扩张，草场会退离定

**趋势分析** 及其在生态股市中的应用



居点越来越远。最后，不得不放弃旧定居点，盖新的定居点，开始新一轮的退化过程。所以“定居轮牧”必定导致草原退化。世界上几大草原都退化了，多个当初的游牧民族后来都衰落了或者转业了，也许就是他们经不住定居的诱惑的结果。而欧亚草原之所以成为世界上“保护得最好的草原”（李博语），可能是和蒙古民族一直坚持游牧，有完整科学合理的草原文化有关。这个课题值得我们继续深入发掘、探讨。比如，蒙古牧民宁愿在野外寻捡牛粪背回来，也不肯砍伐灌木作燃料。再比如，蒙古牧民特有的蒙古包，便是最坚固、最保暖，兼有抗风、防火功能，科学合理的可移动民居。草原文化是中华民族的骄傲，是我们对世界文明的贡献。我认为，发扬光大草原文化，才是草原畜牧业现代化的方向。以房子、车子、票子为标志的牧业现代化，以牺牲生态环境为代价换取人民生活水平的提高是不可取的。因为那很可能会给整个中华民族带来一场生态灾难。如何实现草原牧业现代化，提高牧民生活，发展畜牧业生产，同时保持草原的生态环境生态功能，是我们这代人面临的一个重大课题。如果不定居，草原牧业如何才能实现现代化。如果定居，如何避免放牧强度分布不均匀。如果定居且游牧，牧民家属定居，牧工跟随畜群在放牧场上漫游，则需要具备哪些条件？我以为关键还是要有足够的放牧场。要压缩牧区的牲畜和人口，在“十一律”的基础上，建立畜草平衡。而要使限量的牲畜产生足够的产值，牲畜的质量、畜产品的深加工产业链、农牧区之间的分工（牧区繁殖、农区育肥）等配套措施都要跟上。

和谐的民族关系是国家繁荣发展的重要基石，是中华民族各民族的共性。个人以为，现阶段民族关系的最高境界是“和而不同”。在讲究民族团结，共同进步的同时，我们也需要了解“农耕文化”和“草原文化”的个性。两者的对立斗争、转

化、妥协、和解是中国历史的重要组成部分。比如，长城几乎成了中华文化的标志。但当初为什么要修长城呢，不是为了民族团结，而是为了防御侵略，更正确地说是防游牧民族的骑兵的。城墙不能阻挡步兵，人员可以翻越城墙，但马匹不能。最近，看到一部有趣的书《狼图腾》，说到牧业文化是中国文化中重要的、进步的成分，我有同感。我觉得，那种以为牧业生产，游牧生活原始落后，无视草原文化，用“先进的”农耕文化思想指导牧业生产是个战略失误，其严重后果不可低估。草原文化，是中华文化的一个重要组成部分，是黄河、长江以外的又一个中华文化的源头。草，草原，除了我们上面提到的环境保护、生态屏障的意义外，在人类历史发展过程中，起过特别的作用，留下了特别深的痕迹。例证就是，我们的主要家畜，驼、牛、马、羊等多是草食动物。我们的主要食物，主要农业作物，都是禾本科植物，是草原植物。

作为人类文明摇篮的草原，其实是很神奇的。草原是自然形成的原生植被，是自然资源，既不同于作为放牧对象的“草场”，也不同于作为耕作对象的“草地”，当然更不是农耕文化所定义的待开垦的“荒地”。草原是可更新的自然资源。如果我们能够与草原和谐相处，适度利用，草原可以让我们的子孙后代永续利用，而且能保证我们天蓝、地绿、水清，保证我们的可持续发展。我们大家都知道森林有水土保持功能。同为植被，草原也有水土保持的功能，而且有和森林不同的、更独到的功能。根据宁夏沙坡头，据说已经得到联合国教科文组织认可的防沙治沙的经验，在沙地表面用人工布草格子是非常成功有效的方法，是至今为止唯一成功有效的方法。沙坡头的人们最先曾经试种树，但失败了。沙坡头的经验从一个侧面说明了草本植被防止沙尘暴的功能。草本植物耗水量小，而且它的生物量主要分布在地表，在大气和大地的界面。个人认为，草

趋势分析 及其在生态股市中的应用





原植被旱生禾草的强大浅层根系才是保护土壤，防止风蚀的根本保障。相比之下，树木的耗水量大，而生物量却主要分布在空中和地下。所以，对防止沙尘暴来说，植树的作用不如种草。尚且不说，草原地带降雨量不够，钙基层浅露，基本上种不活树。可是却听说有林业上的人乐观的报告说，种了多少多少树，因此沙尘暴已经被有效地遏制了。我觉得这样的说法是不负责任的。或者他是个外行，在说外行的话。按我们上面的讨论，虽然宏观上说植树可以改善生态环境，但落实到草原具体到防治沙尘暴，植树却几乎无能为力，是白白地浪费草原上宝贵的水资源。而且生态环境的变化，气候的变化是以年代、世纪，甚至地质年代为单位来表现的。一两年不出现沙尘暴，并不能说明土地荒漠化趋势的转变（其实，要想知道气候变化的趋势，把气象资料作回归分析即可）。我们已经欠了太多的生态债，在短期内是根本无法偿清的。（在我们的会议几个月后，2006年春，沉寂了两年多的沙尘暴再一次袭击中国北方，又一次引起了世界人们的注意。）草原同时又是不可再生的自然资源。草原本身有一定的恢复功能。局部的、轻度的、短期的破坏可以恢复。但如果超过它的恢复能力，一旦毁灭，草原是不可再生的。据报道，许多大跃进时期的大片的撂荒地，现在仍然没有恢复植被，从而成为沙尘暴的沙源地。美索不达米亚是不可再生的自然资源被破坏后的典型例子。我的观点是：草原是自然资源，而不是我们的劳动对象。我们可以认识它、利用它，但不可以改造它、建设它。否则，将会遭到大自然的惩罚。举例来说，海洋是自然资源，不是我们的劳动对象。我们可以近海养殖，可以远洋捕捞，可以修港口、建码头。但我们不能“建设海洋”。因为海洋是一种自然存在，是不以人的意志为转移的自然存在。与大自然相比，我们人类太渺小了。有人说，人定胜天，我们要搞“草原建设”，我们要“种草养

畜”。我但愿我自己错了，他们能“种草养畜”。但我建议先冷静冷静，反思一下《人口论》的教训。即使再伟大的人物，也不能违背自然规律。在自然规律面前，我们都是小学生。希望我们认真挖掘探索草原和草原文化的含义，按生态规律办事，维护好草原这个中华民族的生态屏障，否则中华民族将面临一场严重的生态灾难，绝不可以掉以轻心。

谢老师（在看了中国青年报的报道后）问我 2001 年以来的近况和研究课题的进展。2000 年，我从美国农业部辞掉工作，回国内呆了 18 个月。归国不得归，在经济上和心理上都遭受重大挫折，中国青年报已经报道了。返回美国后，沉寂了两年多。因为是坚持辞职回国，坚持研究成果属于内蒙古大学。因此和原单位、和学校关系不太和谐（一般来说，与大学签雇佣合同后，研究成果属于岗位发明，自动属于雇主，除非能证明除外。我向学校证明以夹角余弦值做相似系数是我硕士研究生论文的成果，而超球面模型是我硕士论文的延续）。因此，我没有回原单位、原学校，在得到妻子和家人的理解和支持后，和几个朋友合作，东山再起，自己创业开公司。秀才做生意，而且是老秀才，学究做生意，困难很多，挑战也很多。今年 3 月份，终于在律师朋友帮助下注册了两个公司：一个数据分析公司，一个投资俱乐部。而且，我们的材料有幸地吸引到了有关部门的眼球。在我这次回国前，收到联邦政府证券委员会中部地区办公室来信，约我面谈，了解我们材料中所宣称的 6 年投资实验回报率超过道·琼斯、纳斯达克标准、普尔 500，是否属实，是否有数据支持。我的超球面模型研究，曾经得到包括中国科学院资深院士、中国生态学会名誉主席阳含熙先生，中国科学院应用数学研究所前所长王寿仁教授等多位科学家个人的签名支持，但从来没有找到肯为我们做鉴定的权威单位。因此，我把美国证券会的来信看作是一次机会。虽然



他们是市场规范办公室，是纠察机关，不是科研鉴定单位，但对我来说仍然是一次难得的机会。我与律师协商后给他们回信，说明我是科研工作者，不是诈骗犯。我从中国回来后，会亲自到他们办公室提供我们6年投资的文字记录。所以，如果我回美国后3个月，你们听说我还自由的话，我的模型就成功了。

既然大家有兴趣，我就顺便介绍一下我的科研工作。大家可能已经注意到我的名片上的勾股定理，3平方加4平方等于5平方，说的是向量加法。据文献记载，中国最早提出勾三股四弦五的是商高。我所定义的商高指数是全部分析股票的平方和的算术根。个人认为，商高指数优于道·琼斯指数。铁委员开玩笑说，不知道我是不是在吹牛。西游记里唐僧说，出家人不打诳语。我们搞科研的也不打诳语，而是靠数据说话，靠事实说话。为什么我敢说商高指数要比道·琼斯指数好？不仅我们六年的实验结果远远优于道·琼斯，而且在理论上我们能够证明这个判断成立。道·琼斯是取30个样本，用样本的参数来作总体的参数，理论上是“有偏估计”。例如，中国有13亿人口，如果我们想知道中国人的平均身高（这个问题听起来有点蠢，但这里仅为举例），对13亿人口都量身高是不可能的或者太费时间、太费钱。我们找1.3万中国人的身高的数据，取其平均，作为中国人的平均身高。如果取的样很有代表性，1.3万人的平均身高可以非常接近13亿人的平均身高，但它确实不是13亿人的平均身高，和13亿人的平均身高肯定有差别。用1.3万人的平均身高来代表13亿人的平均身高因此被称为“有偏估计”。如果我们能迅速、经济地得到13亿人的平均身高，我们当然宁愿用真正的数据，而不用样本来代表。商高指数是所有被分析的股票的正平方和的算术根，是几何和（它可以衍生出几何平均，类似代数平均）用它来表示股市状态在理论

上是无偏估计。所以，商高指数要比道琼斯指数好，要比从道琼斯理论衍生出来的任何指数都合理。由于计算机的使用，计算所有股票价格的平方和的算术根很方便。所以，商高指数迟早要取代它们，不管它们现在是如何的流行、如何的庞大，不管它们将来如何抵抗、如何反对。

我的专业是数量生态学。研究课题是（生态）系统动态监测，通俗的说法是趋势分析，是介于线性代数、向量分析、统计之间的边缘科目，在科研文献方面是个空白。联合国教科文组织发起的生态系统监测网已经有 30 多年的历史了，但生态系统动态监测的数据分析问题似乎仍然没有比超球面模型更好的答案。为了描述系统，我们使用向量（数组）来表示系统状态。为了使用历史数据预报未来，我们用回归分析。为了进行回归分析，我们在超球面模型里定义了向量的乘法和除法，用以处理指数增长。区别和认识指数增长很要紧。马尔萨斯的历史功绩，不仅在于提出了人口问题，而且还指出了线性增长和指数增长的区别。由于其细胞分裂的本质，生物的增长是指数增长：2，4，8，16，…。生物增殖（以及其他具有自身复制功能的现象，如资本生息、病毒细菌繁殖、爆炸等）有时候，或者在一定条件下看起来好像线性增长，用线性模型也能解决一部分问题，但其本质是指数增长，只有用指数增长的观点才能真正解决问题。我曾经给内蒙古自治区领导写信，提出单纯植树种草不是解决内蒙古问题的办法，因为人类植树种草的速度根本赶不上牲畜增长的速度，赶不上千百万牲畜日夜啃噬的速度。他听了可能挺不受用，或者以为我是在讽刺挖苦他。他宁愿花自治区人民的血汗钱远去南美的阿根廷考察“学习”畜牧业，也没有在他的办公室里见我一面。而我是在美国拿到草原生态学博士回来报效的内蒙古自己培养的理学硕士，而且有牧业生产的实际经验。

趋势

## 趋势分析 及其在生态股市中的应用



再回到这次会议的出发点——草原生态和游牧文化。感谢铁委员为我们提供了一个讨论的平台。希望大家共同努力，争取最后能够影响推动民众和决策者，实行科学决策、民主决策，调整现存的牧业生产政策，进一步加强草原保护政策。更具体地说，我们要在“十一律”的指导下，建立新的畜草平衡，使草原得以被我们永续利用。如果世界上最可宝贵的人，尚且需要控制增长，则更何况牲畜。让我们把草原文化真正作为中华民族的文化遗产的重要组成部分，全民动员，挖掘、抢救、振兴、发扬草原文化，建立和谐的人与自然的关系，回答历史的挑战，迎接中华民族的复兴。

我们现在正处于草原科学的转型期，我们对草原的认识正在逐步深化。过去（据1954年宪法），除了森林、农田以外的都是荒地。现在，草原已经有了法律地位，是国土资源，可更新自然资源，受国家法律保护。未来，由于生产发展，农区和半农半牧区的畜产品的数量质量的提高，草原畜产品的比重将下降。从而，草原作为畜牧业的物质基础的意义将有所下降，而生态意义将有所加强。人们最终将认识到草原是中华民族生态屏障。草原科学将从农业建设型转到生态保护型。草原、大漠将成为高质量的生态旅游地，为都市人们提供休憩地。极而言之，草原之与国土，好比城市绿地之与城市。城市绿地是城市之肺，没有人要在市内草坪上养宠物。草原是国土之肺，将来草原牧畜也只是为了保护国土之肺功能的需要。

## 多维空间与向量分析

Two and Multi-Dimensional Spaces, multi-component vectors, communities of plants. Sample space and variable space,  $R$  analysis and  $Q$  analysis. Magnitude and direction of the vectors. MDSM on Knowledge tree. Three subscripts data,  $D_{(i,j,k)}$ , variable-sample-time  $m$ -data analysis

### 一、小游戏

草原植被时间动态分析 (Temporal Dynamic Analysis, TDA) 的问题, 或植被演替监测的问题属多元指数增长系统的动态分析问题。本讲义将向读者介绍多元指数增长系统动态分析及其研究方法, 超球面模型 (Multi-Dimensional Sphere Model, 简称 MDSM) 的原理及其应用。在进入多元系统动态分析之前, 我们来做个数学游戏:

用笛卡尔平面直角坐标系  $XOY$  表示两个植物种所组成的植被: 横轴  $X$  代表植物种甲, 纵轴  $Y$  代表植物种乙, 原点是  $O$ , 单位刻度是克/平方米。植被状态用平面上的点  $Z = Z_{(x,y)}$  来表示。另有时间变量  $k$ 。  $X$ ,  $Y$  和  $Z$  都是时间的函数,  $X = X_{(k)}$ ,  $Y = Y_{(k)}$ ,  $Z = Z_{(k)} = (X_{(k)}, Y_{(k)})$ , 在游戏开始时,  $k = 0$ , 物种甲的地面生物量是每平方米 30 克, 物种乙是每平方米 40

趋势分析 及其在生态股市中的应用



克:  $X_{(0)} = 30, Y_{(0)} = 40, Z_{(0)} = (X_{(0)}, Y_{(0)}) = (30, 40)$

下面我们分 A、B、C 三种情况来讨论植被的变化。

### 1. 情况 A

假如第二年,  $k = 1$ , 草原植被发生变化, 由  $Z_{(0)} = (30, 40)$  变成  $Z_{(1)} = (40, 50)$ 。也就是说, 物种甲由每平方米 30 克, 上升到每平方米 40 克, 同时物种乙由每平方米 40 克上升到每平方米 50 克。增幅是  $+10$ ,  $\Delta X = \Delta Y = +10$ 。作为数量生态学工作者, 我们感兴趣的问题是, 从第一年到第二年, 在我们的观察期间, 植被成分是否有变化, 有变化的话, 在向哪个方向变化?

解: 物种甲地面生物量由 30 克上升到 40 克, 而同时植被地面生物量由 70 克 ( $30 + 40$ ) 上升到 90 克 ( $40 + 50$ ); 所以物种甲在植被中的份额由  $30/70 = 42.8\%$  增加到  $40/90 = 44.4\%$ 。与此同时, 物种乙的地面生物量由 40 克上升到 50 克, 所以物种乙在植被中的份额由  $40/70 = 57.2\%$  减少到  $50/90 = 55.6\%$ 。

答案: 植被不仅产量发生了变化, 而且植被成分也发生了变化, 植被成份在向植物物种甲增加的方向变化。

### 2. 情况 B

假若第二年,  $k = 1$ , 草原植被发生变化, 由  $Z_{(0)} = (30, 40)$  变成  $Z_{(1)} = (20, 30)$ 。也就是说, 物种甲由每平方米 30 克下降到每平方米 20 克, 同时物种乙由每平方米 40 克下降到每平方米 30 克。增幅是  $-10$ ,  $\Delta X = \Delta Y = -10$ 。我们感兴趣的问题同样是, 植被成分是否有变化, 有变化的话, 在向哪个方向变化?

解: 物种甲地面生物量由 30 克下降到 20 克, 而同时植被地面生物量由 70 克下降到 50 克; 所以物种甲在植被中的份额由  $30/70 = 42.8\%$  下降到  $20/50 = 40\%$ 。同时, 物种乙的地面生物量由 40 克下降到 30 克, 所以物种乙在植被中的份额由  $40/$

$70 = 57.2\%$  上升到  $30/50 = 60\%$ 。

答案：在 B 情况下，两植被组成成分物种甲和物种乙地面生物量各下降 10 克，植被不仅产量有变化，而且植被成分也发生了变化，植被成分在向物种乙增加的方向变化。

### 3. 情况 C

假若第二年， $k = 1$ ，草原植被发生变化，由  $Z_{(0)} = (30, 40)$  变成  $Z_{(1)} = (60, 80)$ 。也就是说，物种甲由每平方米 30 克跃升到每平方米 60 克，同时物种乙由每平方米 40 克跃升到每平方米 80 克，地面分种生物量分别各增加了一倍。我们感兴趣的问题仍然是，植被成分是否有变化，有变化的话，在向哪个方向变化？

解：物种甲地面生物量由 30 克上升到 60 克，而同时植被地面生物量由 70 克上升到 140 克；所以物种甲在植被中的份额前后保持不变，都是  $30/70 = 60/140 = 42.8\%$ 。

同样，物种乙的地面生物量由 40 克上升到 80 克，植被地面生物量由 70 克上升到 140 克；所以物种乙在植被中的份额也前后保持不变，仍然是  $40/70 = 80/140 = 57.2\%$ 。

答案：在 C 情况下，植被仅产量有变化，但植被成分没有变化，处于稳定平衡状态。

以上，我们讨论了增量是正负自然数的情况，现在，我们要把分析推广到一般情况，对于任意的数对（两个数），怎样确定植被向哪个方向变化？

解：过坐标原点  $O$  和代表初始状态的点  $Z_{(0)} = (30, 40)$  做射线  $OZ$ ，射线  $OZ$  将  $XOY$  平面划分为 3 个区域：以  $OX$  和  $OZ$  为边界的平面  $XOZ$ ，以  $OY$  和  $OZ$  为边界的平面  $YOZ$  和射线  $OZ$ 。这三个区域分别是：物种甲上升区域，物种乙上升区域，物种甲乙保持初始比例发展的区域。可以分别表示为： $T_{(x)} > T_{(y)}$ ， $T_{(x)} < T_{(y)}$  和  $T_{(x)} = T_{(y)}$  区域（见图 1-1）。

**势分析** 及其在生态股市中的应用



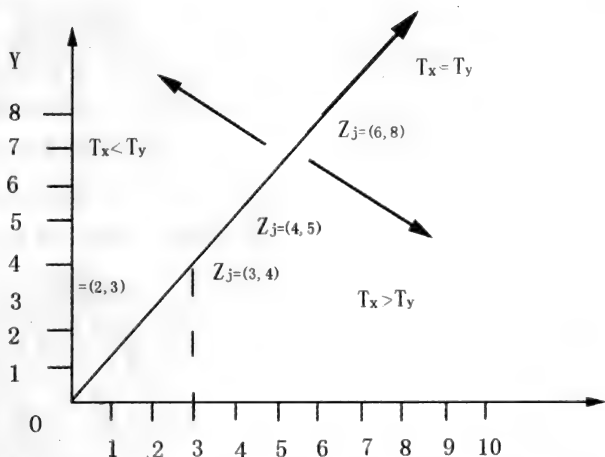


图 1-1 MDSM 应用于股市分析的二维表示

其中  $T$  是变化趋势值， $x$  和  $y$  是表示物种的下标。我们将在第 5 章（本书中有讲义的前二章）里正式给出趋势值的定义。

当代表植被的点，沿初始向量的方向  $OZ$  做径向运动时，植被是在按初始的组成比例在增长，我们认为此时的植被是处于稳定平衡状态。

如果代表植被的点横向运动，或说位置向量（position vector，连接空间的点和原点的向量）在平面上发生偏转，则表示植被的组成发生了变化。这种组成变化，可以用向量在空间的偏转来度量：如果向量偏向  $X$  轴，则植被在向物种甲增加的方向变化；如果偏向  $Y$  轴，则植被在向物种乙增加的方向变化。

为了使我们的设想真正能够为实践服务，我们需要把这个结论从二维空间推广到多维空间（Multidimensional space,  $m$ -空间），把笛卡尔（Descartes, Rene 1596—1650）平面直角坐标系

扩展到多维空间直角坐标系。当植被上有  $m$  个植物种时，或者说，当植被是由  $m$  个植物种组成的时候，这片植被可以被看作是以  $m$  个植物种为轴所建立的多维空间中的一个边界模糊的超椭球体，或超体积 (hypervolume)。当我们用多维空间来表示植被，用多维空间的点来代表植被状态时，空间的点或以空间的点为端点的多元向量向某个坐标轴偏转，则表明对应的物种在植被中的比重在增加。而向量在多维空间的偏转，可以用向量和相应坐标轴的夹角，正确地说是用夹角余弦值  $\cos$ ，或向量在该坐标轴上的投影来定量地度量。当状态向量向  $X$  轴偏转时，夹角  $\angle XOZ$  变小，而  $\cos \angle XOZ$  增大，空间点在  $X$  轴上的投影向离开原点的方向移动，而这个位移是可以度量的。

从这个小游戏中我们可以看到，由于组成植被各物种的基准点，在这里指的是初始的地面生物量不同，虽然组成系统的各分量增幅（减幅）相同，但植被动态分析的结果说明植被组成变化了。另一方面，即使各分量增幅（减幅）不同，但如果增率相同，则植被动态分析的结果显示植被是稳定且平衡的。所以，植被动态分析和差（增幅）无直接关系，而只和增率有关。比如，草原建群种羊草的地面生物量每平方米减少 1 克和弱势种仙人掌的地面生物量减少 1 克的生物学效果，以及其向我们传递的信息，肯定是不同的。羊草地面生物量一般在每平方米 100 克左右，减少 1 克，变化率仅是 1% 左右，很大程度上可归于随机误差，没有什么特殊意义，但一个弱势种的地面生物量每平方米可能只有 2—3 克，减少 1 克，相对变化幅度很大，约 30%—50%，无法归于自然误差，而需要分析其非自然原因。比如其原因也许是野战军车的碾压，毁坏了肉质植物。所以，植被动态分析的基本出发点应当是比率、是乘除法，而不是加减法。也就是说，描述植被动态增长的模型，监



测植被的模型应当是基于指数增长的多维模型。

如果您对以上的数学小游戏感兴趣，并已基本掌握了游戏规则，那么，这本讲义的深度就比较适合于您。

## 二、基本定义和概念

超球面模型 (Multi-Dimensional Sphere Model, 简称 MDSM) 是由草原监测发展起来, 能够对多元系统进行动态分析的新的数学工具。前面的小游戏中已经包含了超球面模型的几个重要前提假定: 变量独立、资源共享、时间惯性、指数增长。使用多维直角坐标系表示 MDSM 接受变量独立的假定; 使用向量描述植被表示 MDSM 接受多维变量空间的点和植被一一对应的假定; 使用夹角表示 MDSM 接受指数增长的假定。用趋势值预报未来, 表示 MDSM 接受惯性的假定; 下面我们开始进入正题, 逐个地阐述超球面模型的基本定义和概念。在这本讲义中, 植被的时间动态分析 (Temporal Dynamic Analysis, TDA) 被归于系统动态分析。系统的变化决定于系统组成的变化。系统动态分析就是系统组成变化的分析。在做系统动态分析前我们首先要明确系统, 系统的组成元素, 系统的定性定量描述 (第一章) 及其基本运算法则 (第二章)。

### 1. 系统和元素

我们在这本讲义里取上海辞书出版社 1979 年出版的辞海中的说法: 系统是指多个相同或相类的事物按一定的秩序和关系所组合而成的总体。按这个定义, 植被是一个系统, 是由植物组成的系统, 组成系统的个体我们称之为系统的元素或分量。植物种是植被的元素。系统和元素的关系是总体和个体的关系。总体是由个体组成, 并由个体来体现的。认识了个体,

才有可能了解总体。但认识了每个个体，如果不进行归纳和综合，不了解个体之间的关系仍然不一定了解总体。这本讲义所研究的着眼点在于个体和总体的关系以及系统和元素之间的关系，讲解如何从元素的运动状态来解释系统的运动状态，如何从草原组成各物种的消长，来解释草原的动态，并预报未来的状态。

### 竞争系统和互惠系统

我们上面所说的系统的例子：植被的一个重要特征是资源共享，所有组成系统的元素分享，竞争共有的资源。

组成植被的各植物种之间互相竞争水分、阳光和养料。由于植物之间的竞争，加之人类活动影响了这种竞争，影响了水分、阳光、养料在不同植物种之间的分配，导致了植被的变化。我们认为草原监测的问题，不仅是产量的监测，更重要的是草原组成成分变化的监测，是物质能量信息在不同植物种之间分配的变化的监测。也就是说，草原分种产量的变化才是草原变化的实质问题。由于竞争共有资源这个特性，我们在进行系统动态分析时，有必要假定组成系统的元素之间的关系是平等的，不存在互相依赖，互相制约，或特惠于某个特定元素的关系。系统中或许存在有不同元素之间的互惠关系或相克现象，但就总体来看，竞争是主要矛盾。特别是当我们侧重于研究系统和元素之间的关系的时候，研究生物与其环境的关系的时候，我们认为元素和系统的关系是主要关系，是我们研究的重点，而个别元素之间的关系是次要的、从属的、相对的，是在进行宏观研究和系统动态分析中所必须忽略的。另一方面，变量无关的假定，又为各物种间可能存在的各种各样的关系预留了足够的操作空间。而且，只有变量无关的假定，才能够为各物种间的各种各样可能存在的关系预留足够的空间，好比笛卡尔平面直角坐标系为研究两个变量的复杂关系提供了有



力工具，多维空间直角坐标系也为多个变量的关系的研究提供了一个有力工具。

下面，我们提出一个命题来描述竞争系统：在竞争系统中， $1+1 \leq 2$ 。我们的读者也许听说过在一些系统中一加一大于二的。论证一加一大于二的一个比较浅显的例子说，一个瞎子和一个瘸子住在同一家旅馆里。当旅馆失火时，瞎子和瘸子独自都很难逃脱。但当瞎子背起瘸子时，情况就变了。由瘸子指路，由瞎子走路，两个人都脱离了险境。类似的情况，一加一大于二的例子可以举出许多，如汽车、飞机、航空母舰都属于这类系统。这样的系统是互补系统或互惠系统，多数的物理系统、工程系统都属于互惠系统。互惠系统不属于我们这本讲义的讨论范围。在这本讲义中，我们只讨论竞争系统。竞争在社会科学领域或生物学领域是非常普遍的现象，或者说是规律。比如，有个老板，在社区里开了个饺子馆。年收益 10 万元。如果他因此想在同一社区里开第二家同样的饺子馆，人们可以预料，他两家饺子馆的收益加起来不会超过 20 万元，因为客源有限。而我们使用超球面模型可以预计，他的两个饺子馆的收益的期望值是一加一的平方根，约 14 万元。

超球面模型用向量的加法来表述这种竞争系统中元素量值增加和系统量值增加的关系。按向量加法的平行四边形法则，两个向量相加时，和向量的长度是以两个加向量为边的平行四边形的对角线的长度，如公式 2-1 和图 1-2 所示。

$$r^2 = (q \times \sin \langle p, q \rangle)^2 + (p + q \times \cos \langle p, q \rangle)^2$$

其中， $p$  和  $q$  是加向量， $r$  是和向量， $\langle p, q \rangle$  是两向量之间的夹角。正弦  $\sin \langle p, q \rangle = R'R$  和余弦  $\cos \langle p, q \rangle = Q'Q$  分别是两个加向量的夹角的正弦和余弦。图 1-2 显示，当两向量的夹角  $\langle p, q \rangle$  为 0 度时，和向量的长度是两向量的长度的和，当两向量的夹角为 180 度时，和向量的长度是两向

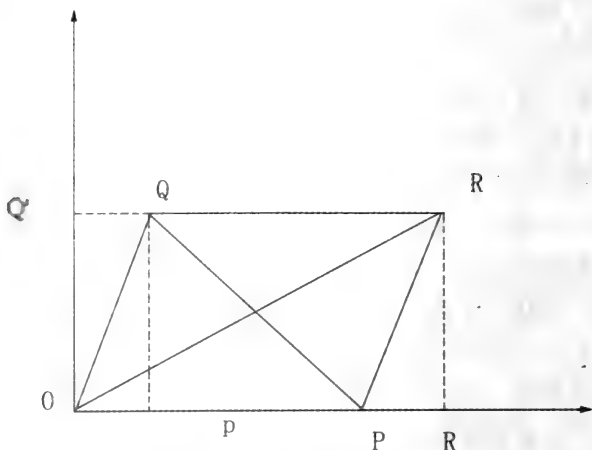


图 1-2 向量加法的图示

量的长度的差，当两向量的夹角为  $90$  度时，和向量的长度是以两向量为边的矩形的对角线的长度。从这里我们可以看出，标量加法是向量加法的特例（当夹角等于  $0$  或  $180$  度时），而向量的加法是标量加法的扩展。由于任意向量可以被分解成一对互相垂直的向量，我们在这本讲义中只讨论向量垂直的情况。虽然在客观世界里，向量垂直的情况少而又少，但研究向量垂直的情况却更有普遍意义。因为任何指向的向量都可以由两个互相垂直的向量来合成。这也是我们可以接受变量独立的假定的基本考量。

### 指数（几何）系统和线性（算术）系统

植被的特性之二是组成系统的所有元素呈几何增长（Geometric Growth），或指数增长（Exponential Growth）。我们的读者过去接触更多的可能是线性增长（Linear Growth），或算术增长（Arithmetic Growth）。一般认为任何曲线都可以用折线去无限逼近，所以似乎任何增长都可以用线性增长来描述或模拟。但

趋

势分析 及其在生态股市中的应用



近来有些研究，特别是生态学的研究，表明一些现象不能够用线性模型来满意地解释。例如生态学中一个最基本的现象，种群爆炸，似乎就不能用线性增长来解释。数学上认为按比例增长的事物，可以而且应该作指数增长来处理。生物学对指数增长的解释是，指数增长的实质在于生物具有通过细胞分裂以及有性或无性繁殖来复制 (Replicate) 自身的特殊性质。从本质上讲，凡是复制自身的现象都应当用指数增长来表述。生物繁殖、资本产生利息、信息拷贝都是自身复制的典型例子，都属于指数增长的范畴。物理学中的原子弹爆炸、高温物体的冷却、化学中的元素衰变也都是指数增长 (衰减) 的例子。三者的区别是，前者是指数增长，而后二者是指数衰减的例子。处理指数增长的最基本运算是除法、乘法和乘幂。没有定义除法运算的系统不能很好地处理指数增长。而指数模型可以用来处理线性增长：当幂指数在 0 附近波动，增率围绕 1 振动的指数模型呈现平稳的线性增长。我们认为，生态系统中的线性增长是指数增长的特例，是增率等于 1，幂指数等于 0 的指数增长。

我们在后面要说明，对于一个多元指数增长系统，当组成系统的元素都保持初始的比例增长时，表达系统的运动轨迹在多维空间是一条具有确定指向的射线，如本章开头的游戏 1.1.3 中的例子。让我们设想，如果我们用不同坐标轴来表示国民经济的不同部门，由于不管计划指导或自由竞争，国民经济总是要按比例发展，因此，在多维空间里，国民经济的运行轨迹应该近似于一条射线。当国民经济失调时，这条射线在空间偏转摆动。原则上，我们可以监测这条射线偏转摆动的程度，来调整指导宏观经济的发展。从这里，我们似乎可以看到超球面模型的重要理论意义和实践意义。

### 随机系统，确定性系统，多变量时间系列

一些生态学者们认为生态变量的变化是随机的、不确定

的，因此是无法测量的。然而，我们对超球面模型的研究和检测实践表明生态变量的变动虽然看起来似乎是随机的，但元素在系统中所占的份额似乎是确定的。比如，在草原样方中，羊草的株数似乎是个随机变量，但经验告诉我们：羊草在特定草原植被中的比例是相对稳定的。还有一些生态学者们认为生态系统的变化过程是随机的、不确定的，因此是无法预测的。然而我们对超球面模型的研究和实践表明系统的变动在一定程度上似乎是可以预测的。这里我们所说的预测当然不像对确定性模型，如出膛的炮弹一样，可以随时准确无误地预报它的位置和速度，但植被也决不是不可预测的，因为它的运动是很有惯性的，也就是说，草原的现状和它的历史是有联系的。我们可以把它比作一个具有能量和自我控制机制的生物体在多维空间的游走。对于一个主动运动的物体，我们不能从初始位置和初始速度来预测它在任意时刻的位置和速度，因为它不是一个被动运动的炮弹。但是，如果我们能够知道物体在时间  $k$  的位置和速率，我们便有可能预测它在时刻  $k + \epsilon$  的位置和速率，特别是当  $\epsilon$  和时间单位相比很小（极小量）的情况下。在这种情况下，也就是当客体在时刻  $k$  的位置和速率已知，而且时间增量很小的情况下，我们把植被监测问题和系统监测问题归结为时间系列问题，利用惯性原理来处理系统动态分析问题可能比较合适。由于植被的惯性比较大，度量植被变化的时间单位，年，与植被自然变化的地质年代相比很短，推动植被变化的动力变化不是很剧烈，植被的变化趋势应该是可以预测预报的。中国北方的沙尘暴，内蒙古植被退化很可能是政策层面的失误，而不是人们没有早早感知它的变化。事实上，我们早在 1985 年，《中国 2000 年的预测》课题中就预报了中国草原的退化，及其对气候变化，特别是大风天数增加的影响。我们甚至定量地预报了北京地区 8 级以上大风天数增加的趋势。现在看

趋势分析 及其在生态股市中的应用





来，结果不幸被我们言中了。而我们当时在预测中使用的状态转移对角矩阵就是现在的超球面模型趋势向量的雏形。

## 2. 多维空间，空间的点，和多元向量

在前面的小游戏中，我们曾经用二维空间中的点来表示二元植被的状态，并对二元植被进行了动态分析。我们现在需要把这个思维方法推广到任意多元植被和任意多维空间。在这本讲义里我们用  $m$ -维空间来表示我们所要研究的任意多维空间。按惯例， $m$  是小写斜体字母。在本讲义中， $m$  特指大于三的自然数， $m$  维空间是超过三维的空间。我们现在翻译 MDSM 为超球面模型，代替过去的“多维球面模型”。因为，二维、三维空间也是多维空间。但二维、三维空间的研究已经很透彻了，而植被中只有两种或三种植物的例子却很少见。

引申哈钦松为生态位的定义 (Hutchinson, 1957)，我们取草原植被的数量生态学描述为：草原植被是多维植物种空间的超体积 (hypervolume)。如果一个植物群落有  $m$  个植物种，则草原植被的状态可以而且应该同时用  $m$  个数来描述。当一个植物种的数量指标可以取任何实数时，这个植物种可以用一个数轴来表示。当  $m$  个植物种之间的关系可以忽略不计时，我们用组成草原植被的  $m$  个植物种为坐标轴来建立  $m$  维直角坐标系。这  $m$  个植物种的所有可能的取值组合和加减乘除四则运算构成  $m$  维空间。由于是以变量为坐标轴建立空间，所以又称  $m$  变量空间，以便与样本空间相区别。在这本讲义中，我们用  $R^m$  来表示  $m$  维变量空间。其中幂  $m$ ，表示维数，底  $R$ ，表示元素的取值范围，是实数。

这样，植被是多维物种空间。一块有  $m$  个植物种的植被可以用一个  $m$ -向量来表示。当我们定义了加减乘除四则运算，并解释了它们的和差积商的植被学意义时，植被就是我们要研究的  $m$ -空间。

需要强调的是，这里定义的  $m$ -空间和其他的多维空间，如线性空间有不同。它的元素之间要进行数学运算，而且不仅是加减法运算，还要能做乘除法运算。能否做向量的除法运算是  $m$ -空间和线性空间的根本区别。根据线性空间的定义，在线性空间里是不能做向量除法运算的。

多维空间是个数学概念。当  $m > 3$  时， $m$  维空间在现实世界是不存在的。为了研究  $m$  维空间的性质，我们要借用二维空间和三维空间的知识。我们认为高维空间是由低维空间组成的。也就是说，高维空间也是由点、线、面组成的。而且，在发现新的规律之前，我们假定低维空间的规律也同样适用于高维空间。在《数值分类学》一书中，Sneath 氏也指出，“可在代数上证明，常规三维空间的几何定理能够扩展到多维的欧几里得空间中。”例如，我们认为多维空间中的多元向量是一个点，各分量之间比例相同的点组成一条射线，两个相交向量确定一个平面。我们可以在这个平面上研究这两个向量的关系：它们之间的夹角、投影、垂线。因此，二维的平面几何和三维的立体几何是多维空间、多元向量分析的重要工具，也是主要工具。

### 投影

由于分量之间独立的假定，我们可以分别研究  $m$  元向量的  $m$  个分量，也就是向量在各个坐标轴上的投影，然后归纳  $m$  个分量的结果来合成结果向量。空间  $m$ -向量和第  $i$  个坐标轴在 origin 相交，这两条发自原点的射线确定一个平面。我们可以在这个平面上，从  $m$ -向量的端点向第  $i$  个坐标轴作垂线，连接原点和垂足的线段是向量在第  $i$  个坐标轴上的投影。这样，原点，向量，垂足三点构成一个直角三角形。我们利用这个直角三角形，在空间的第  $i$  个平面上，向量和它的投影所确定的平面上研究第  $i$  个分量的性质。

势分析 及其在生态股市中的应用



## 变量空间与样本空间，Q 分析与 R 分析

变量空间。以变量为坐标轴建立的空间称变量空间。变量空间的点、向量，分别代表样本或系统的状态。我们建立变量空间是为了对样本，对客体 (Objects) 进行分析，找出样本之间，或样本所代表的系统状态之间的关系。这种对客体进行的分析被称为 Q 分析。我们这本讲义主要讲的是 Q 分析。

与变量空间相对应的是样本空间。样本空间有两种解释：一种是指样本的所有可能的集合，另一种是指用样本为轴建立的空间 (Gauch, 1982 年)。两者都假定样本之间独立，这里的独立是指被选取的机会相等。通过随机取样，每个样本被选的几率是相等的。样本空间的点，代表变量，或属性 (Descriptors)。人们在样本空间里研究变量之间的关系，属性之间的关系。在样本空间研究变量之间关系被称为 R 分析。一般来说，常见的统计分析属于 R 分析。

## 多元向量

我们在介绍空间的同时曾经简单介绍了多元向量。现在我们进一步来阐明空间的点、向量和数组三者之间的恒等关系。首先， $m$ -维空间的点和  $m$ -元数组是一一对应的。从  $m$ -空间的一个点分别向  $m$  个坐标轴投影，我们得到  $m$  个数，这  $m$  个数按一对顺序排列构成  $m$ -元数组 (array)。反过来， $m$ -元数组和空间的点也是一一对应的。对于任意一个  $m$ -元数组，以数组的元素为坐标， $m$ -空间中有一个点和这个数组相对应。人们可以用  $m$ -数组来表示  $m$ -空间的点，也可以用  $m$ -空间的点来表示  $m$ -元数组。其次，如果把  $m$ -空间的点看作是从原点  $O$  运动发展来的，则  $m$ -空间的点是发自原点的一支  $m$ -元向量的端点。或以空间的点为端点的位置向量 (Position Vector)。这样， $m$ -空间的点和  $m$ -元向量是一一对应的。对于空间任意的点，连接该点和原点，我们得到一确定的向量。对于任意向量，它

的端点，在多维空间是惟一确定的。这样，我们可以用  $m$ -空间的点来表示  $m$ -元数组，表示  $m$ -向量。反过来，也是一样的。

向量赋予数组一层新的意义。向量是有指向的，向量通过它在空间的指向对系统的组成，其整体结构，组成成分的相对比例，做唯一的刻画描述。这里的唯一，也是一一对应的意思。每一个  $m$ -向量刻画了  $m$ -系统的一个状态，而  $m$ -系统的每一个状态都可以用一个  $m$ -向量来表示。我们这本讲义的目的就是向读者介绍如何使用向量来描述系统的状态和不同状态之间的关系。

### 向量的量值和方向

向量的概念是超球面模型的重要基础。有必要对向量做一个严格的定义。我们取格若莱尔百科全书的定义（Grolier, 1995年）：向量是既有量值（Magnitude）又有方向（Direction）的量。

细心的读者可能已经注意到，格若莱尔百科全书对向量的定义和一些线性代数教科书（如 Leon, 1994年）的定义不同。在 Leon 的线性代数教科书中，向量被定义为： $N$ 个数所组成的数组，而没有提到方向。而格若莱尔百科全书则很强调方向，并把方向作为向量的两个基本要素（量值和方向）之一。在这本讲义里，我们取格若莱尔百科全书的定义。向量就是要有方向，否则就不成其为向量了。如果把向量仅作为数组，或  $M \times 1$  的矩阵，则完全忽略了向量的方向性。事实上，具有方向性是向量非常重要的性质和功能，也是我们这本讲义阐述发挥的基本点。

### 向量方向重要性的实例：

三维空间中的向量  $A = (3, 1, 0)$  和向量  $B = (0, 1, 3)$  是不同的。 $A$  和  $B$  分别是三维空间中不同的点，是不同的



向量，代表不同的系统状态，虽然  $A$  和  $B$  的量值同为根号 10。在植被学中， $A$  和  $B$  可以代表不同的植被。如果这个三维空间的三个坐标轴分别是：树、灌木、草，则  $A = (3, 1, 0)$  是林地，它的组成是 (3 林 + 1 灌 + 0 草)。而  $B = (0, 1, 3)$  是草地。它的组成是 (0 林 + 1 灌 + 3 草)。对于向量来说，不仅量值，方向也是很重要的。对于植被，不仅组成植被的总量 (生物量、能量、信息量的总和) 是重要的，这个总量在各个植物种间的分配也是重要的。总生物量在不同植物上的不同分配，形成不同的植物群落。超球面模型认为发自原点的同一射线上所有的点 (共线的点，Colinears) 代表相同的植被， $B = (0, 1, 3)$  等同于  $B' = (0, 2, 6)$ 。按以上的分析， $B$  和  $B'$  都是草地，而且是具有相同组成的草地。我们把植被组成作为植被的质的指标，而把植被的量值 (Magnitude) 作为量的指标。换句话说，我们认为向量的方向代表植被的质，而向量的长度代表植被的量。我们把上例中  $B$  和  $B'$  在量值上的差异主要归因于可能的取样面积的差异，取样季节的差异，丰年歉年的差异，取样手段的差异。并把它们作为植被动态分析中需要滤去的噪音 (Noise) 来处理。

### 3. 向量分析

#### 超球面模型在知识树上的位置

我们前面说过，超球面模型是为草原监测设计的，是为竞争性资源管理服务的。但超球面模型不属于生态学，虽然它源发于植被分析，源发于数量生态学。超球面模型属于数学的一个分支——向量分析。同时，MDSM 又对传统的向量分析有发展。

图 1-3 是根据 1999 年光盘版的 Grolier 百科全书所绘制的知识树的局部，显示超球面模型在知识树上的位置。依据百科全书，知识树先分为艺术、历史、技术 (Technology)、科学

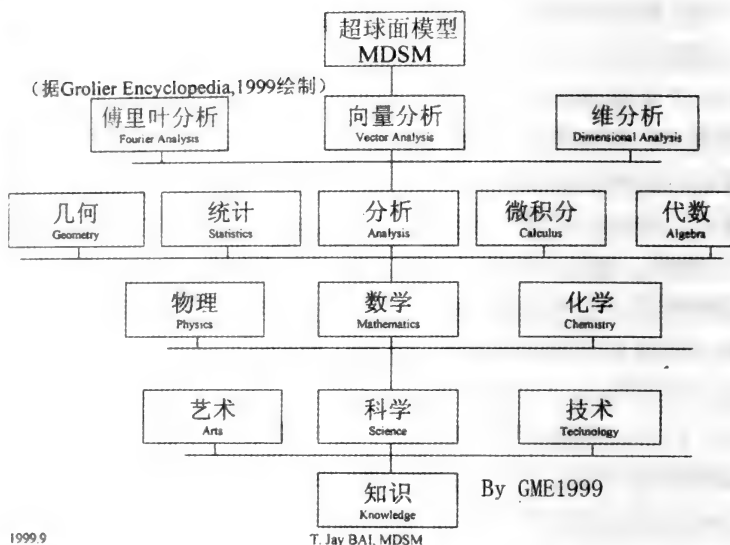


图 1-3 超球面模型在知识树上的位置  
(据 GROLIER Encyclopedia, 1999 绘制)

(Sciences), 科学再分为物理、化学、生物、数学 (Math) 等, 数学再分为代数、几何、数论、微积分、统计、概率、分析 (Analysis) 等, 分析再分为维数分析、傅里叶分析、向量分析 (Vector Analysis)。

传统的向量分析在物理工程上有广泛的应用; 但是:

(1) 传统的向量分析一般限于三维空间 (Davis, 1979 年); 而植被的元素数目不能仅限于 3。超球面模型把向量分析扩展到任意多维空间, 使向量分析可以有更广泛的应用。

(2) 传统的向量分析一般限于线性空间。线性空间没有定义向量除法, 而且, 线性空间的向量乘法的定义对空间不封闭。向量乘法不封闭的意思是说, 两个  $n$ -向量的乘积不再是  $n$ -空间的点, 因此传统的向量乘法没有逆运算。超球面模型作



为扩展了的向量分析，它定义了向量的乘法除法，而且这样定义的向量乘法除法对空间封闭，可以有逆运算，可以处理指数增长，因此特别适用于诸如生物、金融和信息等新兴的科学领域。

(3) 相对于传统的向量分析，MDSM 更侧重于向量的角度。MDSM 认为向量的角度携带着系统组成的重要信息。MDSM 进行分析的主要途径是投影到单位超球面，从而保留向量的方向，保留向量间的夹角。

### 三、三下标变量

我们一般要通过取样来认识辽阔的草原。为了更准确地了解认识草原，我们要做多个观察，取多个样品。草原数据通常都是多样本的，有时一块草地的样本数目可以达到几百个。另一方面，草原植被是由多种植物构成的，样本中对每一个植物种都要有描述、记录。因此，草原数据又是多变量的，一般草原上的植物可以多达上百种。草原数据通常是一个样方  $\times$  变量的矩阵。为了进一步研究草原的动态，人们往往需要回到同一地点，重复观察，称为定位观察，对同一块草地的数据进行多年的积累。自联合国教科文组织推出《人与生物圈计划》以来，世界各国的科学家都建立了各自的定位观察站，并已经积累了成吨的数据。由于这些数据是多样方、多变量、多时间的，一般被称为三向数据 (Three Way Data)。三向数据需要用一种比较特殊的数据形式——三下标变量，来表示：

$$D_{(i,j,k)}, i = 1, 2, \dots, m, j = 1, 2, \dots, n, k = -1, 0, 1, \dots, o;$$

其中第一下标  $i$  用来区别变量或属性，第二下标  $j$  用来区别样本 (观察值)，第三下标  $k$  用来区别时间。三下标变量对

于生态学工作者是新的数据形式。在这本讲义中，我们称其为“晶阵”或“立体矩阵”，以便和二下标变量矩阵相对应，相区别。

至此，我们认识到，数组（Array）是个比较大的概念，它包括向量、矩阵和晶阵，这三者分别是一下标变量，二下标变量和三下标变量。其中，向量具有一个特殊的性质——方向性，从而使它有别于其他的数组。

根据大数定律，当样本数目增多时，样本的平均值逼近总体的真值。在设计取样时，为了更全面准确地描述总体，人们倾向于大量重复取样。特别是计算机的广泛应用，使人们在取样的时候，往往遵循宁滥勿缺的原则，这样就造成了很大的数据库，而且难免形成数据中大量的冗余信息。如何从这些庞杂的数据中，拣出我们所需要的信息，滤去冗余和糟粕，发现数据内部关系，需要特殊的技能和训练以及专门的知识。这本讲义就是指导多元数据分析工作者利用超球面模型对三下标数据进行处理，使用特定的滤波器：中心化（Centralization），标准化（Standardization），卡门滤波（Kalman Filter）来滤去噪音，从大量的植被数据中提取有用的信息，以向量为基本单位，通过对向量的加减乘除运算和角度分析，对向量所代表的系统的状态进行监测和预报，帮助我们认识植被的动态，预报植被未来的状态。

（《超球面模型基本知识》第一章多维空间和多元向量）

趋势分析 及其在生态股市中的应用





## 多元向量基本运算

Basic VECTOR Arithmetic: Addition and Subtraction, Zero vector and negative vectors. Introduction of multiplication and division of  $m$ -vector, unit vector, inverse vector. Transition matrix and diagonal matrix.

超球面模型的最根本的特点是用多元向量,  $m$ -向量来描述多元系统,  $m$ -系统。它不仅用  $m$ -向量的各个分量描述  $m$ -系统对应的元素, 而且, 更深入一步, 用  $m$ -向量的量值——向量长度表示  $m$ -系统的总量, 用  $m$ -向量的方向表示  $m$ -系统的组成。这样,  $m$ -向量就为我们提供了系统的更全面的信息。由于定义了  $m$ -向量的四则运算, 特别是乘法和除法运算, 超球面模型可以对多元系统, 特别是指数增长系统的状态进行分析综合, 对  $m$ -系统的动态进行监测。我们这里所说的  $m$ -系统动态是指系统组成成分的变化。这种变化, 超球面模型用  $m$ -向量在空间的角度偏转来表示, 来度量。

上一讲, 我们介绍了空间、向量等基本概念。这些基本概念是超球面模型的前提, 我们不经证明就接受它们。在这一章里, 我们再定义  $m$ -向量的基本运算。然后, 我们从基本概念出发, 展开逻辑推理和运算。

## 一、向量的表示

按一般惯例，我们用大写字母表示  $m$ -空间的点，如点  $P$ 。用两个端点来表示线段，如用  $PQ$  表示以点  $P$  和点  $Q$  为端点的一条线段。用两个斜体大写字母表示有方向的线段、向量。如  $PQ$ ，表示起点为  $P$ ，终点为  $Q$  的向量。由于向量是有方向的， $PQ = -QP$ 。在这本小册子中，我们也用一个大写斜体字母表示发自原点的向量。如用  $P$  表示  $OP$ ，原点  $O$  省略不写，因为本讲义中所有的向量都发自原点。我们也用黑体小写字母来表示向量，如  $\mathbf{p}$ 。所以，在这本小册子里，斜体大写和黑体小写可以互相通用， $P = \mathbf{p}$ 。我们讨论的  $m$ -向量，读作埃木元向量，是能进行加减乘除四则运算的有  $m$  个元素的向量。在上下文清楚，或在句子开头， $m$ -可以省略不写，直接称向量，或冠之以多元二字，称多元向量。在这本小册子中，多元向量、向量、 $m$ -向量指的是同一个概念。

### 向量的长度和方向的表示

多元向量是有量值和方向的量。向量的量值，或向量长度，我们用绝对值的符号来表示，（在一般说来模和绝对值不是一回事。但在这本书里，我们用绝对值来表示向量、线段的长度。） $|OP| = |\mathbf{p}| = OP$ 。向量的长度是标量。

多元向量在  $m$ -空间的方向用该向量和  $m$  个坐标轴的夹角，或反余弦来表示。在向量与第  $i$  支坐标轴所确定的平面上，或由原点，空间的点，和空间点在坐标轴的垂足为顶点的直角三角形中，根据余弦的定义， $m$ -向量与坐标轴的夹角余弦值可以用向量在坐标轴上的投影与向量长度的比来表示，

向量与第  $i$  轴的夹角余弦值  $\cos\theta_{(i)} = P_{(i)}/|P|$ ，

趋势

势分析 及其在生态股市中的应用



$$i = 1, 2, \dots, m$$

因此，多元向量的方向

$$\theta_{(i)} = \arccos (P_{(i)} / |P|), \quad i = 1, 2, \dots, m$$

这里  $\theta_{(i)}$  是向量和第  $i$  个坐标轴的夹角。多元向量和  $m$  个坐标轴有  $m$  个夹角，有  $m$  个余弦值，所以向量和坐标轴的夹角余弦值也是  $m$ -向量，简称余弦值向量。余弦值向量在植被学中表示植被组成成份的相对重要值 (Importance Values) 在系统分析时表示系统状态，是非常重要的概念。

两个向量  $p$  和  $q$  的夹角 (用  $\langle p, q \rangle$  来表示。两个向量的夹角余弦值用  $\cos \langle p, q \rangle$  来表示，它是标量。

### 单位向量

在某个坐标轴上的投影是一个单位，而在其他轴上的投影是 0 的向量称在指定轴方向上的单位向量。例如，在三维空间， $A = (1, 0, 0)$  和  $B = (0, 1, 0)$  都是单位向量。 $A$  是第一轴上的单位向量， $B$  是第二轴上的单位向量。

### M 向量

始自原点，在所有  $m$  个轴上投影都是 1 的向量称 M 向量。 $OM$  的长度  $|OM| = OM = \sqrt{m}$ ，和每个坐标轴的夹角都相等，都等于  $\arccos (\sqrt{\frac{1}{m}})$ 。

例如，在 3-空间里，M 向量是  $OM = (1, 1, 1)$ 。3-M 向量的长度是， $|M| = \sqrt{3} = 1.732$ 。3-M 向量和任一轴的夹角的余弦值是  $\frac{1}{\sqrt{3}} = 0.577$ 。3-M 向量和任一轴的夹角是  $\arccos 0.577 = 54.74$ 。

在 5-空间里，M 向量是  $OM = (1, 1, 1, 1, 1)$ 。5-M 向量的长度是， $|M| = \sqrt{5} = 2.236$ 。5-M 向量和任一轴的夹角的余弦值是  $\frac{1}{\sqrt{5}} = 0.447$ 。5-M 向量和任一轴的夹角是  $\arccos 0.447 =$

63.43。我们在这里介绍  $M$  向量是为了以后引入逆向量的概念。

## 二、向量的基本运算

在进行  $m$ -向量运算前，首先要定义  $m$ -空间。定义  $m$ -空间指的是首先明确我们在多少维空间里讨论问题，组成空间的变量是哪些。例如，在植被分析前，我们要作本底调查，明确植被中有哪些植物种。在系统分析前，要首先明确我们要监测分析哪些元素。超球面模型所要研究的主要对象是  $m$ -系统的动态，研究同一系统在不同时刻的不同状态之间的关系。如植被是如何依时间演替的。它所研究的是只有量变而没有质变的系统。也就是说，在研究过程中， $m$ -系统始终保持维数相同，向量始终是同一空间的向量。超球面模型只能处理在同一空间的向量，在同一空间的向量之间进行比较和运算。

我们这里将介绍的  $m$ -向量运算是把基本的标量四则运算扩展到向量。在做  $m$ -向量加法、减法、乘法和除法时，把对应分量相加、相减、相乘、相除，即分别得到和向量 (Sum)，差向量 (Difference)，积向量 (Production) 和商 (Quotient) 向量。这样运算的结果，和、差、积、商向量都仍然是  $m$ -空间的向量，而且有逆运算。下面，对向量的加减乘除四则运算逐一加以说明。

### 1. 向量加法

向量加法的定义

以分量的和作和的分量

$$A_{(i)} + B_{(i)} = C_{(i)}$$

$$i = 1, 2, \dots, m \quad [2-1]$$



向量  $X = (0, 3)$ , 向量  $Y = (0, 4)$ , 它们的和向量  $Z = (0, 3) + (4, 0) = (3, 4)$ 。这是例子 1.1.1 中的草原初始状态。

我们上面介绍的例子是向量互相垂直的情况。当向量互相不垂直时, 我们使用平行四边形法, 如图 1-2。按向量加法的平行四边形法则, 两个向量相加时, 和向量的长度是以两个加向量为边的平行四边形的对角线的长度, 和向量的方向是对角线的方向。平行四边形法的实质是把一个向量, 如图 1-2 中的  $OQ$ , 向另一个向量  $OP$  做投影, 分解 (Decompose) 成两个互相垂直的向量, 分别用夹角余弦值  $QQ'$  和夹角正弦值  $RR'$  来表示。余弦值  $QQ'$  和被加向量方向相同, 正弦值和被加向量垂直。然后再分别相加, 取得和向量  $OR$ 。

很显然, 这样定义的  $m$ -向量加法满足交换律 COMMUTATIVE PROPERTY

$$\begin{aligned} A + B &= A_{(1)} + B_{(1)}, A_{(2)} + B_{(2)}, \dots, A_{(m)} + B_{(m)} \\ &= B_{(1)} + A_{(1)}, B_{(2)} + A_{(2)}, \dots, B_{(m)} + A_{(m)} \\ &= B + A \end{aligned}$$

向量加法交换律的几何表示见图 1-2: 由于平行四边形的对边平行相等,

$$OP + PR = OR = OQ + QR = PR + OP。$$

同时,  $m$ -向量加法的定义满足结合律 ASSOCIATIVE-PROPERTY

$$\begin{aligned} A + (B + C) &= A_{(1)} + (B_{(1)} + C_{(1)}), A_{(2)} + (B_{(2)} + C_{(2)}), \dots, A_{(m)} + (B_{(m)} + C_{(m)}) \\ &= (A_{(1)} + B_{(1)}) + C_{(1)}, (A_{(2)} + B_{(2)}) + C_{(2)}, \dots, (A_{(m)} + B_{(m)}) + C_{(m)} \\ &= (A + B) + C \end{aligned}$$

向量加法结合律的几何表示见图 1-2:

$$OP + PR + RQ = OR + RQ = OP + PQ = OQ$$

和向量，中心向量与同簇向量

向量  $OP$  和向量  $OQ$  的和向量  $OR$  的二分之一， $OO'$  “称中心向量 (Centroid Vectors)”。 $O'$  是平行四边形两条对角线的交点，是  $OP$  联线的中点。中心向量以和向量的方向为方向，以和向量的量值的二分之一为量值。以中心向量为中心，向量  $OP$ ， $OQ$ ， $OR$  形成同簇向量。这个向量加法法则可以由两支向量推广到  $N$  支向量。 $N$  个  $m$ -向量可以叠加，产生和向量。如图 1-2 中， $OQ = OP + PR + RQ$ ， $OQ$  是  $OP$ ， $PR$ ， $RQ$  的和向量。因为向量  $QO$  使折线  $OPRQ$  回到原点，向量  $QO$  和向量  $OQ$  互为负向量，向量  $OQ$  是折线  $OPRQ$  的和。和向量的  $N$  分之一称中心向量。中心向量以和向量的方向为方向，以和向量的量值的  $N$  分之一—— $N$  个向量的平均值，为量值。叠加构成中心向量的  $N$  个  $m$ -向量以中心向量为中心构成同簇向量。 $OP$ ， $PR$ ， $RQ$  是以  $OQ/3$  为中心向量的同簇向量。这里需要指出的是，在中心向量的定义中，有权的概念。虽然， $OP$  和  $RQ$ ，互为负向量，两者的和为零，但它们的权仍在，表现为中心向量是和向量的三分之一。同簇向量的概念在植被分析中非常重要。同簇向量与中心向量的差别属于取样误差。在宏观分析时可以忽略不计（我们在第三章里将详细讨论）。

## 2. 向量减法

向量减法的定义

以分量的差作差的分量。

$$A_{(i)} - B_{(i)} = C_{(i)}$$

$$i = 1, 2, \dots, m \quad [2-2]$$

$OP$ ， $OQ$  两向量的差向量是平行四边形  $OPRQ$  的另外的一条对角线。

如图 1-2，由于  $OP + PQ = OQ$ ，移项得  $OP - OQ = QP$ ，

趋势

势分析 及其在生态股市中的应用



或  $OQ - OP = PQ$

### 零向量

多维空间有一个非常特殊的没有方向没有量值的向量，它是原点，用  $O$  表示。

### 负向量

方向相反，量值相等的两个  $m$ -向量互为负向量。两个互为负向量的  $m$ -向量的和为  $O$  向量。 $PQ = -QP$ ， $PQ$  和  $QP$  互为负向量。在植被学中，负向量的经验很少。

多维空间的任意向量可以用发自原点的两个向量的差来表示。向量

$$PQ = PO + OQ$$

相当于  $P$  点运动回原点  $O$ ，再运动到终点  $Q$ 。可以表示为：

$$PQ = PO + OQ = -OP + OQ, \text{ 相应的,}$$

$$QP = QO + OP = OP - OQ = OP - OQ$$

## 3. 向量的标量乘法

### 多元向量的标量乘法

量值相等，方向相同的两个  $m$ -向量是相同向量。

相同的两个  $m$ -向量相加，量值加倍，而方向不变。由此可以引出  $m$ -向量的标量乘法：

$$A + A = 2A = A \times 2$$

### 标量乘法

$$\lambda A = \lambda A_1, \lambda A_2, \dots, \lambda A_m \quad [2-3]$$

标量乘法把向量量值放大，却不改变向量的方向。

方向相同的向量是同向向量，同向向量共线 (Colinears)，位于同一射线上，同向向量所代表的植被是同比植被。同向向量在单位超球面上有同一的投影。

超球面模型最大的特点是定义了  $m$ -向量的乘法和除法。

#### 4. 向量乘法

##### 多元向量的乘法的定义

对应分量的积作积的分量。

$$A_{(i)} \times B_{(i)} = C_{(i)}$$

[2-4]

如此定义的向量积是  $m$ -元向量。

很显然，如此定义的向量乘法满足交换律：

$$A \times B = A_{(i)} \times B_{(i)} = B_{(i)} \times A_{(i)} = C$$

满足结合律：

$$A \times B \times C = [A_{(i)} \times B_{(i)}] \times C_{(i)} = A_{(i)} \times [B_{(i)} \times C_{(i)}]$$

满足对加法的分配律 [DISTRIBUTIVE PROPERTY]：

$$A \times (B + C) = A \times B + A \times C$$

##### 多元向量积的几何表示

积的几何表示，以乘向量与被乘向量的对应分量为边的  $m$  个矩形，见图 2-1。图 2-1 两个三元向量  $a$  与  $b$  的积的图示。图显示的是以两个三元向量为边的三乘三的大矩形，包括 9 个小矩形。其中，对角线上分列的 3 个小矩形  $a_1b_1$ ,  $a_2b_2$ ,  $a_3b_3$ ，表示三元向量的积，仍是三元向量。

##### MDSM 定义的向量乘法和传统的向量乘法的比较

传统的向量分析有两种向量乘法，点乘和叉乘，也称求内积，外积。三种向量乘法可以分别用点乘  $a \cdot b$ ，差乘  $a \times b$ ，和向量乘  $a \times b = A \cdot B$  表示。

点乘的定义是：

$$a \cdot b = a_{(1)}b_{(1)} + a_{(2)}b_{(2)} + \cdots + a_{(m)}b_{(m)}$$

点乘的积是个标量，不再是  $m$ -维空间的向量。点乘没有逆运算。

点乘的几何表示是图 2-1 中对角线上 3 个小矩形的面积的和。它是一个数，标量。

叉乘的定义是





b3			
b2			
b1			
	a1	a2	a3

$$\mathbf{a} \times \mathbf{b} = \begin{pmatrix} \mathbf{a}_{(1)}\mathbf{b}_{(1)}, \mathbf{a}_{(1)}\mathbf{b}_{(2)}, \dots, \mathbf{a}_{(1)}\mathbf{b}_{(m)} \\ \mathbf{a}_{(2)}\mathbf{b}_{(1)}, \mathbf{a}_{(2)}\mathbf{b}_{(2)}, \dots, \mathbf{a}_{(2)}\mathbf{b}_{(m)} \\ \mathbf{a}_{(m)}\mathbf{b}_{(1)}, \mathbf{a}_{(m)}\mathbf{b}_{(2)}, \dots, \mathbf{a}_{(m)}\mathbf{b}_{(m)} \end{pmatrix}$$

叉乘的积是  $m \times m$  矩阵，不再是  $m$ -维空间的向量，叉乘也没有逆运算。

叉乘的几何表示，是图 2-1 中分列的 9 个小矩形。从直观比较，超球面模型定义的向量乘法比点乘携带更多的信息，比叉乘更简练，没有冗余。最重要的是，超球面模型定义的向量积仍然是同一空间的点，是多元向量。这样定义的向量乘法有逆运算——向量除法。

### 5. 向量除法

#### 多元向量的除法定义

对应分量的商作商的分量。

$$A_{(i)}/B_{(i)} = C_{(i)} \quad [2-5]$$

### 逆向量

积是  $OM$  向量的两个向量互为逆向量。

以分量的倒数为对应分量的两个向量互为逆向量,  $A^{(-1)} = (1/A_{(i)})$ ,

商向量等于  $OM$  向量的两个向量相等: 方向相同, 量值相等。

### 6. 多元向量乘法除法的意义

由于定义了乘法和除法, 使我们能够在  $m$ -空间作乘法除法, 把多维空间由线性空间扩展到了非线性空间, 增加了我们认识处理多元系统的能力, 特别是增加了处理指数增长系统的能力。具体讲, 向量除法的定义, 使描述系统动态的方程

$$AX = B \quad [2-6]$$

有解, 使监测多元系统的动态成为可能。

在式 2-6 中,  $A$  和  $B$  是连续时刻的系统状态向量,  $X$  是状态转移向量, 有  $X = B/A$ 。

### 7. 用运算符号表示的系统监测流程

在定义了多元向量的乘法和除法后, 我们可以把使用超球面模型进行系统动态监测, 比如, 草原监测的思路表达如下:

如果草原上有  $m$  个植物种有待我们监测。我们用第一下标  $i$  表示植物种, 用第二下标  $j$  表示样本, 第三下标  $k$  表示时间。用  $k-1$ ,  $k$ ,  $k+1$ , 分别代表过去、现在和将来三个时间, 则过去、现在和将来的植被状态可以分别表示为  $Z_{(i,k-1)}$ ,  $Z_{(i,k)}$  和  $Z_{(i,k+1)}$ 。为了研究它们三者之间的关系, 我们再引入趋势向量  $T$  表示趋势, 预报向量  $P$  表示对未来的预报值, 样本中心向量  $S$  表示观测值, 期望向量  $E$  表示期望值, 且令所有的权重都相等, 则多元系统监测的主要流程, 运算可以简单地表示如下:

# 趋

## 势分析 及其在生态股市中的应用



中心化 (见第三章) 可以表示为:

$$Z'_{(i,k)} = \sum Z_{(i,j,k)} / n$$

标准化 (见第四章) 可以表示为:

$$Z_{(i,k)}' = Z_{(i,k)} / |Y_{(k)}|,$$

其中  $|Z_{(k)}|$  为向量长度, 或商高指数, 等于各分量的平方和的算术根。如果把向量长度  $|Z_{(k)}|$  理解为“向量和”, “几何和”, 则标准化可以被理解为多维的百分比。

植被  $k$  时刻的趋势值  $T_{(k)}$  可以表示为现在比过去的增率 (见第四章):

$$T_{(k)} = Z_{(k)}' / Y_{(k-1)}',$$

趋势值是多维的商。

而未来植被状态的预报值可以表示为现状和增率的积 (见第五章):

$$P_{(k+1)} = T_{(k)} \times Z_{(k)} \text{—预报值是多维的积或幂;}$$

未来植被状态的期望值可以表示为预报值和观察值的加权平均 (见第七章):

$$\begin{aligned} E_{(k+1)} &= P_{(k+1)} + S_{(k+1)} \\ &= Z_{(k)}^2 / Z_{(k-1)} + S_{(k+1)}; \end{aligned}$$

期望值是预报值和实测数据的加权平均; 预报误差可以表示为期望值和预报值的差 (见第八章):

$$\begin{aligned} R_{(k+1)} &= E_{(k+1)} - P_{(k+1)} \\ &= \text{SIN} \langle E, P \rangle, \end{aligned}$$

预报误差是期望值和预报值的夹角正弦。如果读者掌握了以上的基本思路, 便已经基本了解了本书所介绍的超球面模型的内容。理解超球面模型的关键在于接受它的前提假定。超球面模型的逻辑运算部分则基本上是一维标量的四则运算向多维空间的扩展。

### 三、讨论

#### 植被状态转移矩阵是对角矩阵

植被是我们人类最重要的自然资源和生态屏障之一。在传统的植被科学中，人们用样本 - 变量矩阵来表示植被。植被状态变化的原因被抽象为植被状态转移矩阵。下面，我们从矩阵求解的角度讨论植被状态转移矩阵的解的存在形式，来阐明使用向量表示植被状态，使用向量表示状态转移矩阵，提出向量除法及超球面模型的必要性和必然性。

**命题：**植被状态转移矩阵 (State Transition Matrix) 如果有解一定采取对角矩阵 (Diagonal Matrix) 的形式。

设有连续的植被状态  $A$  和  $B$ ， $B$  不等于  $A$ 。令  $AX = B$ ，则  $X$  是状态转移矩阵。

求解  $X$ ， $X$  有  $m \times m$  个元素。

**解：**

因为  $AX = B$

一般有  $X = B/A$ ，问题归结为对植被状态数组 (Array)  $A$  求逆

##### 2.3.1 $A$ 是矩阵的情况

在  $A$  是矩阵的情况下， $A$  逆存在的充要条件是：

###### 2.3.1.1 $A$ 是方阵

###### 2.3.1.2 $A$ 满秩，组成矩阵 $A$ 的 $m$ 个向量之间线性无关。

先来讨论 2.3.1.1  $A$  是方阵的情况。若  $A$  不是方阵，变量数不等于样本数，则  $A$  无逆，状态转移矩阵  $X$  无通解。如  $A$  是方阵，也就是说，样本数等于变量数，则  $A$  是方阵，可能有逆。但我们可以看到这个结论是多么勉强，偶然，多么不合



逻辑。状态矩阵有无逆的条件本应由植被状态来决定，而在这里却可以轻易地为人们的取样数目来改变。

下面我们讨论 2.3.1.2，即使样本数等于变量数， $A$  是方阵， $A$  仍然不一定有逆，理由如下。我们是借用动力学的方法来研究植被动态的。在动力学中，研究对象是被作为质点来看待的。一个运动的汽车是个质点，一个飞行的飞机是个质点，一个运动的天体，如地球也是质点，虽然它们可能是很复杂，甚至庞大的系统。同样的，在植被动态研究中，植被也必须被作为质点来看待。用植被学的术语讲，植被必须是匀质的 (Homogeneous)，否则，一片植被在多维空间不可能有统一的运动轨迹。匀质的植被，在多维空间是一个实体，由一个点，中心点 (Centroid)，植被状态向量来代表。我们在设计取样过程中的样本独立原则，不等于不保证我们所采的样本线性独立。样本—变量矩阵  $A$  中的所有样本向量都是植被状态向量的同簇向量，是植被状态向量的变异。如果考虑到取样误差，矩阵中所有样本向量全一样。样本—变量矩阵中实质上只有一个线性独立向量。所以方阵  $A$  不可能满秩，也即无逆。而且，当植被样本—变量方阵  $A$  不满秩时，任何数学处理都不能使它的秩增加。也即是说，如果  $A$  无逆，则  $AAT$ ， $AYAT$ ， $YAYT$ ，也都无逆。其中  $T$  是转置矩阵的符号， $AT$  是矩阵  $A$  的转置， $YT$  是矩阵  $Y$  的转置， $Y$  是任意的方阵。

### 2.3.1 结论

当  $A$ ， $B$  是样本—变量矩阵 (双下标变量) 时，植被状态转移矩阵无通解。

### 2.3.2 $A$ ， $B$ 是 $m$ -植被状态向量，单下标变量时

$AX = B$  方程组有解的必要条件是状态转移矩阵  $X$  的  $m \times m$  个元素中，

#### 2.3.2.1) 只有 $m$ 个待定分量/元素，其他 $m \times (m - 1)$

元素已知，或为零。

2.3.2.2) 状态转移矩阵的对角线元素的绝对值大于非对角线元素。这  $m$  个未知待定元素分布在对角线上。

讨论 2.3.2.1.  $X$  是状态转移矩阵，所有的  $m \times m$  个元素都是未知，待定的，则未知数多于方程个数，植被动态分析问题无确定解。为了简化问题，使问题有确定解，我们先求  $m$  个未知数，而把其他的  $m \times (m - 1)$  个元素的值设为 0。

讨论 2.3.2.2. 如果我们把  $m \times (m - 1)$  个元素的值预设为零，然后解出  $m$  个非零解，那么这  $m$  个非零解在矩阵中如何排列呢。为了充分考虑到所有变量，这  $m$  个元素应当排列在对角线上。也就是说，植被状态转移矩阵要采取对角矩阵的形式。这个结论也可以从“状态转移矩阵的对角线元素的绝对值大于非对角线元素的绝对值”的前提假定推出来。也就是说：变量自相关作用大于互相关作用。因为生物学上有生殖隔离 (Reproductive Isolation) 做保障。生物变量的现在值取决于生物的历史值和环境，而不取决于其他变量（虽然其他生物可能有影响，但所有其他生物的影响应当归纳进特定物种的环境）。这样我们便导出：状态转移矩阵对角线元素的值是植被状态向量对应分量的商的结论。用不太确切的说法是，植被状态转移矩阵的元素是先后植被状态在空间两点连线的斜率。

### 2.3 结论

植被状态转移矩阵 (State Transition Matrix) 的唯一确定解如果存在，应当是对角矩阵 (Diagonal Matrix) 的形式。而对角矩阵无异于多元向量。也即，多元向量是研究植被动态正确的数学工具。进而是多元指数增长系统动态分析的正确数学工具。

我们在这里简要地、模式性地讨论了在系统动态分析中应用向量的必要性和必然性。即使从矩阵出发，我们最后的解也



只能是对角矩阵的形式。在实际应用中，使用矩阵求解草原监测问题几乎是不可能的。我们所面对的草原植被，有很多植物种，一般群落也有一百多个植物种。而考虑到每个植物种都可能携带着植被演替的重要信息，我们在植被动态分析中没有考虑使用消元法。这样，面对如此高维空间，矩阵求逆的问题远远超过笔者的数学水平，而且即使有解的话，也超过普通读者的操作能力和经济负担能力。与此同时，简单易行的多元向量分析为我们提供了系统动态分析的有效数学工具。

(《超球面模型基本知识》第二章向量的四则运算)

## WHAT CAN WE DISCOVER FROM 1,2,3 TO 2,3,4? An Introduction to Instantaneous Trend Analysis.

Lets consider a situation where vegetation recovering from a fire event, such as Yellowstone National Park. The first year, say 1994, the data collected were 1, 2, 3, and the second year, 1995, were 2, 3, 4, respectively. These measurements represent the abundances of trees, shrubs, and grasses.

The data will look like these:

Year \ Plants	Trees	Shrubs	Grasses	Total
1994	1	2	3	6
1995	2	3	4	9

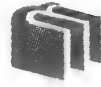
What can we discover from 1-2-3 to 2-3-4?

1) All of us can see that the vegetation changed from 1994 to 1995.

2) Most of us will agree that the vegetation have increased, but different plants may increase at different speeds. These may be considered as a temporal dynamics of vegetation.

趋势分析 及其在生态股市中的应用





3) Some of us would analyze the data and found out that:

Trees increased  $2/1 = 200\%$ ,

Shrubs increased  $3/2 = 150\%$ , but

Grasses increased  $4/3 = 133\%$ , and

The Total increased  $9/6 = 150\%$ , respectively.

The data will look like these:

Year \ Plants	Trees	Shrubs	Grasses	Total
1994	1	2	3	6
1995	2	3	4	9
Increasing	200%	150%	133%	150%

4)A

However, there may not be many of us who agree that the grasses are relatively decreasing, while all the figures appear to be increasing. To discover the instantaneous changing trends for each plant at given year, we use the total increasing ratio to adjust each single increasing ratio and name this as Trend Values, short for Multivariate Instantaneous Trends for 1995. The Total increasing ratio is  $9/6 = 150\%$ . After using this Total ratio to adjust the ratio of each plant, the Trend Values for each plant are:

$$T_{(t)} = 200\% / 150\% = 1.33,$$

$$T_{(s)} = 150\% / 150\% = 1.0, \text{ and}$$

$$T_{(g)} = 133\% / 150\% = 0.89, \text{ respectively.}$$

The above analysis are expressed in the following table 4)A:

4) A

Year \ Plants	Trees	Shrubs	Grasses	Total
1994	1	2	3	6
1995	2	3	4	9
Increasing	200%	150%	133%	150%
Trend Value	1.33	1.00	0.89	

The Trend Value is the increasing ratio of plants adjusted by that of the Total. The Trend Value for each plant can also be obtained by another means.

4) B:

The portions of the trees, shrubs, and grasses for the 1994 were plants divided by total:

$$Trees(\%) = 1/6 = 17\%,$$

$$Shrubs(\%) = 2/6 = 33\%, \text{ and}$$

$$Grasses(\%) = 3/6 = 50\%, \text{ respectively.}$$

The portions for 1995 were:

$$Trees(\%) = 2/9 = 22\%,$$

$$Shrubs(\%) = 3/9 = 33\%, \text{ and}$$

$$Grasses(\%) = 4/9 = 44\%, \text{ respectively.}$$

The trends for the three plants are portions (95) divided by portions(94):

$$T_{(t)} = 22/17 = 1.33,$$

$$T_{(s)} = 33/33 = 1.00, \text{ and}$$

$$T_{(g)} = 44/50 = 0.89, \text{ respectively.}$$

The above trend analysis are shown in the following table 4) B:

趋势分析 及其在生态股市中的应用



Year \ Plants	Trees		Shrubs		Grasses		Total
	#	%	#	%	#	%	
1994	1	17	2	33	3	50	6
1995	2	22	3	33	4	45	9
Trend Value		1.33		1		0.89	

The percentage is the ratio of the value of the plants over total. And the Trend Value is the ratio of percent of 1995 over 1994.

5) We can rank the plants by their Trend Values:

$$T_{(t)} = 1.33$$

$$T_{(s)} = 1.00$$

$$T_{(g)} = 0.88$$

The conclusion: The trend for trees is increasing, as the Trend Value of the trees is greater than one; while the trend of grasses is decreasing, as the Trend Value of grasses is less than one.

6) Without the concept of multivariate instantaneous trend, it is very difficult to convince people that this land will be dominated by trees and most of the grasses will be gone in 10 years, provided the trends remain the same in the next ten years:

$$1.33^{10} = 17 \text{ and}$$

$$0.88^{10} = 0.31.$$

The above trend analysis and prediction can be shown in the following table:

Year \ Plants	Trees	Shrubs	Grasses	Total
	1994	1	2	3
1995	2	3	4	9

Increasing	200%	150%	133%	150%
Trend Value	1.33	1	0.88	
$T^{10}$	17.31	1	0.31	

### 7) Conclusion:

We can definite the adjusted increasing ratio as instantaneous trend for plant  $i$  at time  $k$ , in a simplified form,

$$T_{(i,k)} = X'_{(i,k)} / [X'_{(i,K-1)}]$$

Where  $X'$  are relative abundance expressed in a percentage form:

$$X'_{(i)} = \frac{X_{(i)}}{\sum X_{(i)}}$$

With the trend analysis we discovered that Trees are increasing the most, the Shrubs are stable, but Grasses are decreasing. Furthermore, we may make a best prediction based on existing information that the vegetation will be dominated by trees in ten years, instead of the present grass community.

### 8) Discussion:

In temporal dynamic analysis, the MDSM assume that the changes of plants can only be discovered by comparison of the present over past of the plants, unless it is proved otherwise. In other words, it is the presumption of MDSM that all plants are function of environment expressed by community, and every plant is important and should be kept in temporal dynamics analysis.

The instantaneous trend value is a function of time. This makes it more flexible and different from the long term average trend.

### 9) Further Discussion:

We extend this discussion to two-dimensional space. Please recall



two ancient great Mathematicians: Shang Gao (?) from China and Pythagoras (BC 500 – 580) from Greece.

They discovered the relationship between the vector length and its components in two dimensional space:

$$3^2 + 4^2 = 5^2.$$

In other words, in two dimensional space,

$$3 + 4 = 5, \text{ instead of } 7.$$

While the sum is interpreted as vector length, the percentage is interpreted as cosine values:

$$3/5 = 0.6, \text{ and } 4/5 = 0.8.$$

The two cosine values determine the direction of the vector. When the cosine values of 0.6 and 0.8 change, the vector will change direction on the plane, and vice versa.

This can further be extended to multi-dimensional space:

vector length of  $x_{(i)}$  is the square root of the sum of the squares,

$$L_{[x]} = \sqrt{\sum x_{(i)}^2}$$

In our case, vector length of  $(1,2,3) = \sqrt{14} = 3.74$ , and vector length of  $(2,3,4) = \sqrt{29} = 5.39$ .

We can use these values replacing the sum and begin our discussion all over again. This time, our discussion is in 3-space and the trees, shrubs, and grasses are represented by three axes. We can imagine the state vector of vegetation is rotating in 3-space, from  $(1/3.74, 2/3.74, 3/3.74)$  in 1994 to  $(2/5.39, 3/5.39, 4/5.39)$ , 1995.

This supplies a new tool that we may use to investigate vegetation changes over time.

Summary of 9)

Using a vector sum instead of scalar sum in multi-space, the

changes in vegetation can be seen as the state vector rotating in multi-space. Thus, people can monitor the vegetation dynamics by tracing the movement of the state vector on the unit hyper-sphere. This new method of performing Multivariate-Instantaneous-Trend analysis and system monitoring was named as the Multi-Dimensional Sphere Model (MDSM).

In the MDSM, an observation is expressed as a point or an  $m$ -vector in  $m$ -space, and a community is the centroid of the observations. The state of the vegetation is then a standardized (normalized) centroid, i. e., the center of the projections of the observations on the unit hyper-sphere.

In MDSM, the distance expresses the quantity while the direction expresses the quality. In other words, the distance of a vector relates to the production of a vegetation, while direction of the vector contains the composition information of the vegetation. MDSM considers that the vegetation  $A = (1, 2, 3)$  equals to vegetation  $A' = (10, 20, 30)$ , as the two are in the same direction; but the two are different from vegetation  $B = (3, 2, 1)$ , although  $A$  and  $B$  have the same vector length. Furthermore, the trend may express the slope of the trace on the hyper-sphere, which indicates the reasons that have caused the changes.

P.S. What is expected after 1-2-3 and 2-3-4?

If people ever wanted to predict the state of the vegetation for 1996 based on the existing information, what they may do? People can predict the vegetation using the information collected from last two years. The MDSM supposes the trend remains the same in neighborhood, and makes a prediction for next time interval:

$$P_{(k+1)} = D_{(k)} \times T_{(k)}$$

and this presumption of "exponential growth in neighborhood"

趋势

势分析 及其在生态股市中的应用



brings in prediction errors. In the latest research, the prediction was adjusted by updated observation, i.e., collected data from the next time interval:

$$E = (P + D)/2, \text{ and}$$

$$R = \sin \langle E - P \rangle .$$

Where, the  $E$ ,  $P$ ,  $D$ , and  $R$  are  $m$ -vectors of expectation, prediction, observation, and error, respectively.

In above example, as the 1995 were (2, 3, 4), and the trend were (1.33, 1, 0.89), so the prediction of 1996 based on the existing information are  $P_{96} = (2.66, 3, 3.56)$ .

If the collected data on 1996 were  $D = (3, 4, 5)$ , then the expectation for the true value of 1996 would be  $E = (P + D)/2 = (2.83, 3.5, 4.28)$ , and the prediction error would be  $R = (0.17, 0.5, 0.72)$ . This prediction error would be smaller than the differences from the two observations:  $96 - 95 = (1, 1, 1)$ .

The entire procedure of temporal dynamic analysis is shown in the following table:

Year \ Plants	Trees	Shrubs	Grasses	Total
1994	1	2	3	6
1995	2	3	4	9
Percentage	200%	150%	133%	150%
$T$ -Value-95	1.33	1	0.89	
$P$ -1996	2.66	3	3.56	
$D$ -1996	3	4	5	
$E$ -1996	2.83	3.5	4.28	
$R$ -1996	0.17	0.5	0.72	

Interested readers please come to Aug. 14, session 155 of Ecological Society of America meeting.

(1995年12月,超球面模型通讯2-12,以一组火灾迹地植被恢复的虚拟数据,阐述超球面模型应用于植被演替分析的推导过程)

趋势分析 及其在生态股市中的应用





## About "Ocosaic Coefficient"

Message 58/99 From Tugjayzhab Bai

Jan 19, 97 10:02:40 pm - 0700

Subject: ? Ochiai coefficient ?

To: mdsml@gpsrv1.gpsr.colostate.edu

Date: Sun, 19 Jan 1997 22:02:40 - 0700 (MST)

Cc: jbai (Tugjayzhab Bai), sino - ecol@biomod.fw.vt.edu,  
STATSOC - L@MHC.MTHOLYOKE.EDU

Is there any body knows anything about "Ochiai Coefficient"?

Dear friends,

I used to think that cosine value as a similarity coefficient was mentioned in "Multivariate Analysis in Community Ecology" (Gauch, 1982, Cambridge), and discussed and applied in vegetation classification in "Exploration of Numerical Classification of Form *Leymus Chinensis* ..." (Bai, 1982, University of Inner Mongolia). However, recently, when I was reading "Data Analysis in Community and Landscape Ecology" (Jongman et. al., 1995, Cambridge), I found that the cosine was named as "Ochiai coefficient". However, I could not find related reference in the book. Is there anybody can tell me more detail about the Ochiai' work, such as, is he a mathematician or an ecologist, is he the same person as the editor of "Advanced Studies in Pure Mathematics 18", and when he used cosine value as similarity of coef-

ficient?

Thank you very much.

Jay Bai

(Jongman, et. al., 1995, page 178)——“More emphasis is given to qualitative aspects by not considering a site as point but as a vector. Understandably, the direction of this vector tells us something about the relative abundances of species. The similarity of two sites can be expressed as some function of the angle between the vector of these sites. Quite common is the use of the cosine (or Ochiai coefficient):

$$\begin{aligned} \cos \langle Y_{(i)}, Y_{(j)} \rangle &= OS_{(i,j)} \\ &= \frac{\sum (Y_{(k,i)} \times Y_{(k,j)})}{\sqrt{\sum Y_{(k,i)}^2} \times \sqrt{\sum Y_{(k,j)}^2}} \end{aligned}$$

Message 57/99 From Philip Dixon

Jan 20, 97 10:03:21 am - 0500

Date: Mon, 20 Jan 1997 10:03:21 - 0500

To: jbai@lamar.ColoState.EDU

Subject: ? Ochiai coefficient ? - Reply

Dr. Bai,

I may be able to shed a small amount of light on the Ochiai coefficient.

However, I can't answer your questions about Ochiai's background. I suspect he was a member of the active Japanese population biology of the 1950's. To get much more info, you might ask some Japanese ecologists or someone like David Goodall who was active at that time.

I usually see Ochiai's coefficient associated with a presence/ab-

趋势

势分析 及其在生态股市中的应用



sence coefficient:

$$\frac{a}{\sqrt{a+b} \times \sqrt{a+c}}$$

(e.g. Goodall chapter in Whittaker 1982, Ordination of plant communities, or Digby and Kempton 1987, Multivariate analysis of ecological communities). The reference is Ochiai, A. 1957. Zoogeographical studies on the solid fishes found in Japan and its neighbouring regions. Bull. Japanese Society of Scientific Fisheries 22: 526 - 530. Of course, this coefficient is the 0/1 analog of the cosine coefficient for continuous data. This might be the reason van Tongeren and others synonymize the two. As an aside, many coefficients have multiple names. The Bray - Curtis is an especially prolific one.

Hope this helps,

Philip Dixon

(多维球面模型通讯 4-3, 在美国生态学会统计生态学分会的网址上征求余弦值作为相似系数的出处。得到的答复是, “Ochiai Coefficient” 虽然有时被称为余弦值, 但只被用于 0/1 数据, 而非连续数据。发表于 1997 年 1 月)

## A Case of Projection

From: marco [sandeco@MBOX.VOL.IT]  
 Sent: Tuesday, June 09, 1998 3:23 AM  
 To: ECOLOG - L@UMDD.UMD.EDU  
 Subject: number of species - Mathem. predetermination -

Hello everyone,

Who could be useful for this problem.

I am a biologist and in my ecology researches. I try to forecast what the evolution in the number of animal species will be? I assumed there is an interrelation with cohabitant living organisms.

Example:

In a pond we find 4 species of living organisms: I, II, III, IV. In monthly surveys we observe if their number increased, decreased, or remained constant thus assessing them respectively 1, 2, X.

After 10 observations the result will be:

I	II	III	IV
1	2	1	2
1	1	2	1
1	1	1	2
2	1	X	1
1	2	1	1



2	1	2	2
1	2	2	1
X	1	2	2
X	2	1	2
2	1	1	2

What would the next sequence of number will be ?

There always is a law which is non-visible to the eye but identifiable through a system. Is there a software available able to work this out?

Thanks.

Marco/Italy

=====

From: T. Jay Bai [bai@gpsr.colostate.edu]

Sent: Tuesday, June 16, 1998 7:53 AM

To: sandeco@MBOX.VOL.IT

Subject: RE: number of species.

Marco/Italy,

If you could give me the raw data instead of converted data, I can offer a try. Or you can download a DOS executable computer software, SMM-52, from <http://lamar.colostate.edu/~jbai> to try out your self. When run the program, your control numbers are 4, 10, which means four dimensional space, ten vectors.

Best,

T. Jay BAI, Ph.D.

- End -

=====

From: Marco Canepa [sandeco@tin.it]  
 Sent: Wednesday, June 17, 1998 11:09 AM  
 To: Jay Bai  
 Subject: Here are the data

Hello Jay,

Thank you for your message and your offer to give a try.

I gladly send to you in annex the raw data from a recent work. I tried to connect with the address you wrote, but I am sorry to say I have no idea on how to type that wave symbol preceding jbai .....

Warm regards. Marco

Marco Canepa  
 sandeco@tin.it

I	II	III	IV	V
504	280	1453	1344	368
466	271	1460	1368	374
485	275	1520	1352	410
465	267	1459	1361	430
436	261	1523	1429	430
473	260	1440	1415	409
516	265	1600	1388	434
481	253	1536	1347	406
531	235	1442	1387	384
620	232	1270	1406	361
624	241	1248	1438	388
670	236	1071	1425	381
639	241	1149	1454	411



621 (noite-247) : 51112 : sdi 1417 : 404

- End -

=====

From: Jay Bai [bai@gpsr.colostate.edu]

Sent: Thursday, June 18, 1998 5:45 PM

To: sandeco@MBOX.VOL.IT

Cc: MDSM (E-mail); bai@gpsr.colostate.edu

Subject: A Sample of Multivariate Temporal Dynamic Analysis

Dear Marco,

Please find enclosed multivariate temporal dynamic analysis output that you requested.

Your variables are I, II, III, IV, and V. I am not sure what they represent. Just as an example, I assumed that these species were plant species, and the values were their productions, and I further assumed that your data were vegetation data, i.e., all five species were collected from the same community. Under these presumptions, I performed a multivariate temporal dynamic analysis. If these data were other ecological data, then you can convert them to responded species accordingly.

There are two tables attached. The first table, MarcoData, is your original data. The second table, Marcoout, is the output of the computer software of MultiDimensional Sphere Model (MDSM).

The MDSM treats your vegetation data as 5 component vectors, 5-vector, and every species as a component of the vector. For example, the five values, 504, 280, 1453, 1344, and 368 together made the 5-vector that represents the vegetation of the first year. In other words, the column Data is a 5-vector representing the observation of the vege-

tation at given year, and same as the column  $P$  (projection),  $E$  (expectation),  $R$  (prediction error), and  $T$  (multivariate instantaneous trend, see below).

There are 14 identical tables inside the Marcoout for each year of the 14 years. There are eight columns and they are explained briefly below:

The first column is the variable names, such as plant species, I, II, III, IV, and V, vector length(V Length), cosine, and system monitoring coefficient(SMC).

$P$ : projection values of the given year based on the previous information. (But for the first year, the projection values are the same as the original vegetation observation data.)

$D$ : Observation Data. These are the values that the program read from the input vegetation data that you supplied.

$E$ : Expectation. Expected true values of the vegetation at given year ( $E = P + D$ , it is weighted averages from projections and observations).

$IV$ : Importance Values. It is the importance values, or relative composition, of the plant species in the vegetation.

$R$ : Projection Error.  $R = E - P$ . Please notice that, the error definition is the difference between the projections and expectations, instead of between the projections and observations.)

$T$ : Multivariate Instantaneous Trend.  $T = IV_{(k)} / IV_{(k-1)}$ . This model analyzes the composition change of the vegetation. It considers the composition change is the essential change of the vegetation, or any community.

$P_{(k+1)}$ : Vegetation Projection for the next year ( $k$  is the index of the year).

趋势

势分析 及其在生态股市中的应用





The first five rows of the output describe the five species. The last three rows describe the general condition of the vegetation.

V Length: vector sum, or vector lengths, describe the general situation of the vegetation

cosine: cosine values between the two vectors describing the correlation between the two vectors. The cosine values under column  $D$  express the correlation between the observations:  $\text{cosine} \langle D_{(k)}, D_{(k-1)} \rangle$ , while the values under column  $E$  express the correlation between the observation and expectation:  $\text{cosine} \langle D, E \rangle$ .

SMC: System Monitoring Coefficient,  $\text{SMC} = \text{cos} \langle E, D \rangle / \text{cos} \langle D, D \rangle$ . When the SMC value is greater than or close to one, we consider that the model is working for system monitoring and its projections were over fitting or correct fitting the observations.

The lowest SMC value during the 14 years is the year of 11: 0.9981, and the highest SMC values occurred on year of 10: 1.0030. As both of them are not much less than one, we consider the projections are fitting with the observations.

In your original email, you asked that: "What would the next sequence of number will be?" As a temporal dynamic analysis model, the MDSM projected the values for the 15th year's vegetation. The values for the plant species I, II, III, IV, and V, are:

597.59, 255.01, 1133.47, 1408.26, and 410.29, respectively ( $k = 14, k + 1 = 15$ ).

How close is this projection? We have to wait till next year to find out. But from the history of the analysis, i.e., the years of 2 - 13, we can expect that it would be a very close projection.

For your second question: "Is there a software available able to work this out?" The answer is yes. A DOS executable computer pro-

gram, SMM52, based on MDSM performing temporal dynamic analysis, can be downloaded from my webpage:

<http://lamar.colostate.edu/~jbai>

You can just click the webpage address, instead of typing it, as you mentioned that there is no "wave" key on your keyboard. Furthermore, there are some other references about this model that you may be interested to check out from the web page.

If you have any more questions, please contact us or send your message to

MDSM@grpsr.colostate.edu (2006 Note: Now, it is TJBMDSM@yahoo.com or TJBMDSM@MSN.NET)

Best Wishes,

T. Jay Bai, Ph.D.

Table One MarcoData (this can be omitted)

I	II	III	IV	V
504	280	1453	1344	368
466	271	1460	1368	374
485	275	1520	1352	410
465	267	1459	1361	430
436	261	1523	1429	430
473	260	1440	1415	409
516	265	1600	1388	434
481	253	1536	1347	406
531	235	1442	1387	384
620	232	1270	1406	361



624	241	1248	1438	388
670	236	1071	1425	381
639	241	1149	1454	411
621	247	1112	1417	404

Table Two Marco Out (This was an computer output ASCII file.  
We put them into a table for readers convenience.)

K = 1

	$P$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	504	504	504	0.24067	0	1	504
II	280	280	280	0.13370	0	1	280
III	1453	1453	1453	0.69384	0	1	1453
IV	1344	1344	1344	0.64179	0	1	1344
V	368	368	368	0.17572	0	1	368
V Length		2094.13	2094.13	1	0	2.23606	2094.13
cosine							
SMC							

K = 2

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	504	466	473.6	0.22517	-30.4	0.93562	436.00
II	280	271	272.8	0.12970	-7.2	0.97007	262.89
III	1453	1460	1458.6	0.69350	5.6	0.99951	1459.29
IV	1344	1368	1363.2	0.64815	19.2	1.00990	1381.54
V	368	374	372.8	0.17725	4.8	1.00866	377.24
V Length		2105.58	2103.21	1		37.4037	2107.06
cosine		0.9998	0.9999				
SMC			1.0001				

K = 3

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$F$	$P_{(k+1)}$
I	436.00	485	475.2	0.22203	39.19	0.98604	478.23
II	262.89	275	272.57	0.12736	9.68	0.98192	270.02
III	1459.29	1520	1507.85	0.70454	48.56	1.01591	1544.18
IV	1381.54	1352	1357.91	0.63447	-23.63	0.97890	1323.48
V	377.24	410	403.44	0.18851	26.20	1.06351	436.04
V Length		2148.78	2140.19	1	72.34		2151.24
cosine		0.9996	0.9995				
SMC				0.9999			

K = 4

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	478.23	465	467.646	0.22077	-10.58	0.99431	462.35
II	270.02	267	267.605	0.12633	-2.42	0.99194	264.84
III	1544.18	1459	1476.03	0.69682	-68.14	0.98904	1443.02
IV	1323.48	1361	1353.49	0.63897	30.01	1.00708	1370.64
V	436.04	430	431.208	0.20357	-4.83	1.07989	464.35
V Length		2110.31	2118.22	1	75.40		2111.99
cosine		0.9996	0.9995				
SMC			0.9998				

K = 5

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	462.35	436	441.27	0.20281	-21.08	0.91864	400.52

趋势分析 及其在生态股市中的应用



II	264.84	261	261.76	0.12031	-3.07	0.95232	248.55
III	1443.02	1523	1507.00	0.69263	63.98	0.99397	1513.82
IV	1370.64	1429	1417.32	0.65141	46.68	1.01946	1456.81
V	464.35	430	436.87	0.20078	-27.48	0.98633	424.12
V Length		2191.96	2175.77	1	86.50		2194.55
cosine		0.9997	0.9996				
SMC			0.9999				

K = 6

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	400.52	473	458.50	0.21404	57.97	1.05537	499.19
II	248.55	260	257.71	0.12030	9.15	0.99995	259.98
III	1513.82	1440	1454.76	0.67912	-59.06	0.98049	1411.91
IV	1456.81	1415	1423.36	0.66446	-33.45	1.02003	1443.34
V	424.12	409	412.02	0.19234	-12.09	0.95794	391.79
V Length	2194.55	2129.42	2142.12	1	90.54		2132.37
cosine		0.9995	0.9994				
SMC			1				

K = 7

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	499.19	516	512.63	0.23130	13.44	1.08064	557.61
II	259.98	265	263.99	0.11911	4.00	0.99010	262.37
III	1411.91	1600	1562.38	0.70494	150.46	1.03802	1660.84
IV	1443.34	1388	1399.06	0.63126	-44.27	0.95003	1318.64

V	391.79	434	425.55	0.19201	33.76	0.99827	433.25
V Length	2132.37	2238.61	2216.31	1	161.05		2250.49
cosine		0.9982	0.998				
SMC			0.9998				

K = 8

	$P_{(k)}$	D	E	IV	R	T	$P_{(k+1)}$
I	557.61	481	496.32	0.22853	-61.28	0.98804	475.25
II	262.37	253	254.87	0.11736	-7.50	0.98526	249.27
III	1660.84	1536	1560.96	0.71877	-99.87	1.01960	1566.11
IV	1318.64	1347	1341.32	0.61763	22.68	0.97841	1317.92
V	433.25	406	411.45	0.18945	-21.80	0.98669	400.59
V Length	2250.49	2152.65	2171.72	1	121.56		2153.63
cosine		0.9999	0.9991				
SMC			0.9992				

K = 9

	$P_{(k)}$	D	E	IV	R	T	$P_{(k+1)}$
I	475.25	531	519.85	0.24467	44.59	1.07062	568.50
II	249.27	235	237.85	0.11195	-11.41	0.95390	224.16
III	1566.11	1442	1466.82	0.69039	-99.29	0.96051	1385.06
IV	1317.92	1387	1373.18	0.64631	55.26	1.04644	1451.41
V	400.59	384	387.31	0.1823	-13.27	0.96221	369.49
V Length	2153.63	2118.43	2124.62	1	123.32		2129.54
cosine		0.9986	0.9984				
SMC			0.9998				

趋势

趋势分析 及其在生态股市中的应用



K = 10

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	568.50	620	609.7	0.29642	41.19	1.21147	751.11
II	224.16	232	230.43	0.11203	6.26	1.00071	232.16
III	1385.06	1270	1293.01	0.62863	-92.05	0.91054	1156.39
IV	1451.41	1406	1415.08	0.68797	-36.33	1.06445	1496.62
V	369.49	361	362.69	0.17633	-6.79	0.96727	349.18
V Length	2129.54	2039.18	2056.87	1	107.59		2077.77
cosine		0.9963	0.9993				
SMC			1.003				

K = 11

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	751.11	624	649.42	0.31548	-101.69	1.06431	664.13
II	232.16	241	239.23	0.11621	7.06	1.03737	250.00
III	1156.39	1248	1229.67	0.59736	73.28	0.95027	1185.93
IV	1496.62	1438	1449.72	0.70426	-46.90	1.02367	1472.04
V	349.18	388	380.23	0.18471	31.05	1.04753	406.44
V Length	2077.77	2055.07	2058.48	1	137.57		2059.64
cosine		0.9998	0.9978				
SMC			0.9981				

K = 12

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	664.13	670	668.82	0.33833	4.69	1.07243	718.53

II	250.00	236	238.80	0.12080	-11.20	1.03944	245.30
III	1185.93	1071	1093.98	0.55341	-91.95	0.92641	992.19
IV	1472.04	1425	1434.40	0.72562	-37.63	1.03031	1468.20
V	406.44	381	386.08	0.19530	-20.35	1.05734	402.84
V Length	2059.64	1956.38	1976.80	1	102.14		1969.47
cosine		0.997	0.9996				
SMC			1.0025				

K = 13

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	718.53	639	654.90	0.32637	-63.62	0.96464	616.40
II	245.30	241	241.86	0.12053	-3.44	0.99777	240.46
III	992.19	1149	1117.63	0.55697	125.44	1.00	1156.40
IV	1468.20	1454	1456.84	0.72602	-11.36	1.00055	1454.80
V	402.84	411	409.36	0.20401	6.52	1.04455	429.31
V Length	1969.47	2017.33	2006.61	1	141.31		2018.86
cosine		0.9994	0.9976				
SMC			0.9983				

K = 14

	$P_{(k)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(k+1)}$
I	616.40	621	620.08	0.31407	3.67	0.96230	597.59
II	240.46	247	245.69	0.12444	5.22	1.03245	255.01
III	1156.40	1112	1120.88	0.56773	-35.52	1.01930	1133.46
IV	1454.80	1417	1424.56	0.72154	-30.24	0.99383	1408.26





V	429.31	404	409.06	0.20719	-20.24	1.01559	410.298
V Length	2018.86	1963.23	1974.32	1	51.25		1964.29
cosine		1	0.9999				
SMC			0.9999				

- End -

=====

From: Marco Canepa [sandeco@tin.it]  
 Sent: Friday, June 19, 1998 1:51 PM  
 To: Jay Bai  
 Subject: Re: A Sample of Temporal Dynamic Analysis

Dear Jay,

First of all I heartly want to thank you for your comprehensive clear reply. I had no difficulty at all, so I believe, in understanding your explanations to your analysis though without entering much in the mathematical process of your model yet.

The data I supplied to your courtesy were an experiment, still in progress, aiming at studying the existence of interrelations between micro-organisms of different species in a controlled environment (an aquarium). Variables such as light, PH, temperature, pollutants etc. are kept constant thus not representing, de-facto, a variable but a constant value, at least at this phase. Variables are the number of individuals of given species only, thus stressing the relations existing between these species. We are not sure at all a relation exists between all of the 5 species introduced in the controlled environment nor the "weight" of relations. Maybe the population increase or decrease in "B" has noth-

ing to do with the values of the other species population. Therefore we conduct experiment with several aquariums, each with a number of species between 5 and 10 but varying the composition of the species. For example in aquarium "KK" we introduced species A C E G I M, in another species: A B C I M; and so on.

The good news is we do not have to wait till next year. Surveys are monthly and the 15th observation data are:

I	II	III	IV	V
520	234	1120	1404	370

I immediately noticed a correct prediction for III and IV, I is correct as far as trend, II and V are not very correct.

I think it is a good result with only 14 observations and we cannot expect to get the right prediction on any observation. Also it is not granted that I II III IV V are interrelated so much that a prediction can be made from their historical series of observations. Also the fall in the population of II and V could be very temporary, maybe lasted only a couple of days so that a medium term average would not have even moved.

That is what I wanted to report to you immediately and want to point out again how I appreciated your interest for my question. Before saying anything more I am looking forward to check out your webpage and look at your model.

Have a joyful day.


Best,

Marco Canepa

sandeco@tin.it

- End -

=====

 趋势分析 及其在生态股市中的应用



Jay Bai

From: Jay Bai [bai@gpsr.colostate.edu]

Sent: Monday, June 22, 1998 7:38 AM

To: Marco Canepa (E-mail)

Cc: ESA (E-mail); MDSM (E-mail)

Subject: Projection for 16th observation

Dear Marco,

This is the second response to your reply last Friday, 6-19-98.

You said in the mail:

"Surveys are monthly and the 15th observation data are:

I	II	III	IV	V
520	234	1120	1404	370

Based on your new information, I put the 15th observation to the MDSM model. The following table is the output.

The projection for 16th observation is:

I	II	III	IV	V
460	232	1148	1417	350

Let's wait and see how close is the projection this time.

Also, you may have down loaded the SMM52 program, by now.

If the output is different from this list, please let me know.

Thanks,

T. Jay BAI, Ph.D.

Table: Projection of observation 16

K = 15

	$P_{(15)}$	$D$	$E$	$IV$	$R$	$T$	$P_{(16)}$
I	597.59	520	535.51	0.277635	-62.07	0.883982	459.67
II	255.01	234	238.20	0.123494	-16.81	0.992368	232.21
III	1133.46	1120	1122.69	0.5825	-10.77	1.025224	1148.25
IV	1408.26	1404	1404.85	0.728333	-3.41	1.009408	1417.20
V	410.29	370	378.05	0.196002	-32.23	0.945994	350.01
Vector Length	1964.29	1920.33	1928.86	1	72.82		1927.35
cosine		0.9987	0.9995				
SMC							1.0008

- End -

=====

From: Marco Canepa [sandeco@tin.it]

Sent: Wednesday, July 01, 1998 11:26 AM

To: Jay Bai

Subject: R: Projection for 16th observation

Dear Jay,

Here are June last quarter observation (16th monthly) data compared to your projection:

I	II	III	IV	V
535	237	952	1422	352
460	232	1148	1417	350

Well, 3 ( II , IV , V ) out of 5 are very good, if not perfect. Is the system sharpening up ?

趋势

趋势分析 及其在生态股市中的应用



Please acknowledge that I passed you a wrong data concerning June 3rd quarter (June 22nd in my previous e-mail) concerning the value of  $\lambda$ . The right value was 895.

Warm regards.

Marco Canepa

sandeco@tin.it

- End -

(多维球面模型通讯 5-7, 多维球面模型的一个成功例子。

1998年6月发表于美国生态学会统计分会网站。[http://atlantic.evsc.virginia.edu/Mailarch/stat-ecol-mh/msg\\_00067.html](http://atlantic.evsc.virginia.edu/Mailarch/stat-ecol-mh/msg_00067.html))

一名意大利网友, Marco, 在美国生态学会网站征集生态系统实验数据预测方法。我们用多维球面模型根据他提供的 14 次观察数据, 预测了第 15 次, 再加入提供的第 15 次观察数据, 预测了第 16 次观察数据。按 Marco 的报告, 预测结果相当接近于他的实际观测数据。

	实测数据 16	预测数据 16
I	535	460
II	237	232
III	952	1148
IV	1422	1417
V	352	350

通讯由 7 封电子邮件组成。

日期	发信人	内容概括
6/9	Marco	向美国生态学会提出问题
6/16	Jay	索要原始观察数据
6/17	Marco	提供 14 个时间段的数据
6/18	Jay	提供数据分析结果, 预报第 15 次实验数据。
6/19	Marco	认为预报效果较好。提供第 15 次观察数据。
6/22	Jay	预报 16 次实验数据。
7/1	Marco	比较预测与实测。



## Three month summary

An investment using MultiDimensional Sphere Model (MDSM)  
Three Month Summary (6/19/98 - 9/18/98)

Dear Colleagues,

Both vegetation and the stock market are communities: vegetation is a community of plants, and the stock market is a community of stocks. All the components of a community share the same resources and compete for the resources; thus, they are mutually independent to each other. Every single component in the community exhibits an exponential growth.

Based on above presumptions, we have conducted a mutual fund investment trial since June 19, 1998 (for background information, please read MDSM Communication 506, *Let's Have Some Fun!*) This trial was to test and fine tune the Multi-Dimensional Sphere Model (MDSM), a model originally designed for vegetation science. Following is a summary of the test. Table One is the three month summary based on the weekly report of September 18, 1998. The weekly report (Table Two) is sorted by values of Average Random Investment (ARI, see bellow), i.e. A-value.

In Table One, the values of the Best 4 funds (they were: CF-SAX, CUSGX, COLIX, and CUTLX, see Table Two) in the last three months totaled \$ 1610.56. If I could have known and selected these

funds three months ago, the value of my investment of \$ 8,000 would be \$ 8052.80.

On the other hand, the Worst 4 funds in last three months were: CTCAX, CNTAX, NGCAX, and CSMIX. The sum of the worst four was \$ 1321.04. If I had unfortunately selected these four funds, then the value of my investment would be \$ 6605.22.

After setting the Best 4 as 100% and the Worst 4 as 0%, the MDSM was scoring 78.59%, as its value was \$ 7742.90. At the same time, Average Random Investment (ARI) was scoring 51.55%, as its value was \$7351.51. The temporary conclusion is that the MDSM performed better than ARI. This test will be continued, and follow up reports will be posted continuously on the MDSM mailing list (MDSM@gpsr.colostate.edu) and my home page.

If you are interested in knowing more about the game, please visit the web site. Copies of all the official transaction confirmations from Liberty Funds Service, Inc. (Colonial Investors Service) are available upon request.

Thank you for your attention.

T. Jay BAI, Ph.D.

Quantitative ecologist

Email: jbai@lamar.colostate.edu

Attachment:

Table One: comparison of the Best, MDSM, ARI, and the Worst.





	4 Funds	5 x 4 Funds	Difference from Worst	Percentage
Best 4	1610.56	\$8052.80	1447.58	100.00%
MDSM 20		\$7742.90	1137.68	78.59%
ARI 20		\$7351.51	746.29	51.55%
Worst 4	1321.04	\$6605.22	0.00	0.00%

The formulae to calculate the percentages are:

$(\text{MDSM} - \text{Worst}) / (\text{Best} - \text{Worst})$ , and

$(\text{ARI} - \text{Worst}) / (\text{Best} - \text{Worst})$ , respectively.

Table Two. Three months report of twenty mutual funds sorted by their returns.

	Pr619	Avalue	Ashare	Mvalue	Pr918	A-value	M-share	M-value
SGI	66.375				58.131			
TOTAL (\$)		800		790.79		\$7351.51		\$7742.90
(CIFAX)	9.25	400	43.24		9.81	\$424.22		\$0.00
CFSAX	10.88	400	36.76	1983.22	11.22	\$412.50		\$0.00
CUSGX	6.64	400	60.24	1983.22	6.80	\$409.64		\$0.00
COLIX	6.52	400	61.35	1997.18	6.55	\$401.84		\$0.00
CUTLX	19.97	400	20.03		19.30	\$386.58	103.219	\$1992.13
COSIX	7.31	400	54.72		6.98	\$381.94		\$0.00
CSBAX	15.53	400	25.76		14.59	\$375.79		\$0.00
COLFX	10.77	400	37.14		10.07	\$374.00		\$0.00
CGUAX	15.17	400	26.37		14.17	\$373.63		\$0.00
COLHX	7.25	400	55.17		6.62	\$365.24		\$0.00
SIMAX	13.81	400	28.96		12.51	\$362.35		\$0.00

CFGAX	19.54	400	20.47	1997.17	17.45	\$357.22		\$0.00
COLGX	21.75	400	18.39		19.36	\$356.05	97.95	\$1896.31
SRSAX	13.38	400	29.90		11.83	\$353.66		\$0.00
COEAX	15.11	400	26.47		13.28	\$351.56		\$0.00
CONAX	13.42	400	29.81		11.55	\$344.26		\$0.00
CTCAX	4.94	400	80.97		4.15	\$336.03		\$0.00
CNTAX	6.44	400	62.11		5.38	\$334.16	329.982	\$1775.30
NGCAX	9.39	400	42.60		7.83	\$333.55	265.538	\$2079.16
CSMIX	33.23	400	12.04		26.36	\$317.30		\$0.00

Where Prc619 are the prices on June 19, 1998;

Avalue are the invested values of an Average Random Investment;

Ashare are the shares of the ARI;

Mvalue are my investment on June 19 using MDSM.

Prc918 are the prices of September 18, 1998;

A-value are the present values of the ARI;

M-value are the present values of my IRA account using MDSM.

TOTAL is the total value at a given time with a given method.

\* CIFAX was merged into other funds after we started the game, so it is invalid now. It was included in the table only to keep the analysis balanced.

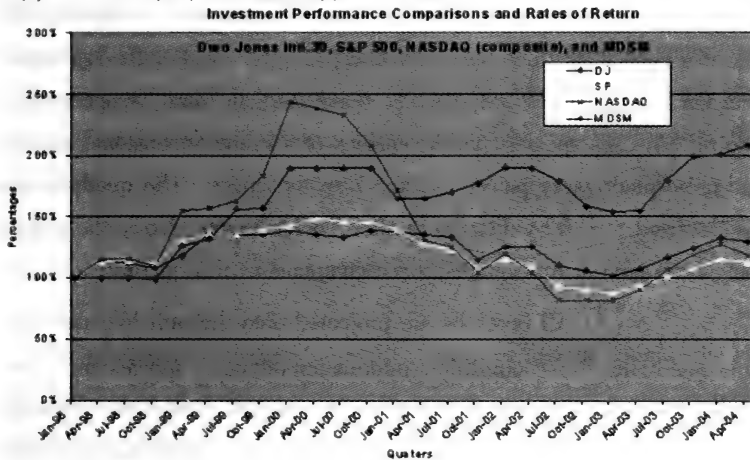
三个月总结 写于 1998 年 9 月份。转刊载在美国生态学会, 统计分会网站上 [http://atlantic.evsc.virginia.edu/Mailarch/stat-ecol-mh/msg\\_00076.html](http://atlantic.evsc.virginia.edu/Mailarch/stat-ecol-mh/msg_00076.html)。收入本书时, 经作者做了删减, 以方便传统读者。从赚钱的角度说, 三个月的投资实验没有赚钱。原因是股市下跌。商高指数从 65.375 下跌到 58.131; 市场平



势分析 及其在生态股市中的应用



均投资从 \$ 8,000 下跌到 \$ 7,351.51。MDSM 管理的投资也从 \$ 7,960.79 下跌到 \$ 7,742.90，但下跌的要比平均投资少。从科学研究的角度，用模型指导投资的实验成功了。以此为开端，投资实验持续做到 2004 年四月份。投资比较结果刊登在《The New Frontier of Investing》，引起美国证券委员会中部地区办公室的注意和约谈。以下图表得到许可，可以刊登在数据分析方法相关的科学研究的交流材料上，但在取得股票交易执照前，不可用来作为招揽投资的广告。



Source: Columbia/Liberty Mutual Funds Account Statements.

Note: MDSM Investment Performance shown is based on a personal IRA account managed by MDSM from June 19, 1998 to April 2004.

## Multi-Dimensional Sphere Model and Vegetation Instantaneous Trend Analysis

**Abstract** The Multi-Dimensional Sphere Model (MDSM), a new method for multivariate instantaneous trend analysis, is introduced. The model handles three subscript data,  $Z_{(i,j,k)}$ , e.g., for vegetation analysis,  $i$ ,  $j$ , and  $k$  are species, quadrats, and time, respectively. The MDSM uses species, as dimensions of a multi-dimensional space, and uses quadrats as points (vectors) in the space. The quadrats are standardized to 1.0 by division by their vector length, i.e., the square root of the sum of the squares of the components of a quadrat,  $q'_{(i)} = q_{(i)} / \sqrt{\sum q_{(i)}^2}$ . All quadrats are projected onto the unit hypersphere. This maintains the composition information of each species for every quadrat in the data set, and makes all quadrats comparable because their vector lengths equal 1.0. When performing trend analysis, the MDSM defines the quotient of components of previous ( $k-1$ ) and present ( $k$ ) state vectors as an instantaneous trend at a given time. This is referred to as a trend vector, and describes vegetation composition change over time,  $T_{(k)} = Z'_{(k)} / Z'_{(k-1)}$ . The components of a trend vector (here called the  $t$ -value of the species) carry information from both previous and present states for species and community. This trend can then be extended to project the future states of the vegetation,  $P_{(k+1)} = Z_{(k)} \times T_{(k)}$ . The MDSM combines correlation analysis, cluster analysis, trend analysis, and prediction of future vegetation

趋势分析 及其在生态股市中的应用



states, making it a powerful and promising multivariate analysis method. The model was tested with data from the Land Condition Trend Analysis program at Fort Carson in southeastern Colorado. The model shows promising results for vegetation trend analysis; however, geometric meaning of the vector quotient is not yet clear. To improve our understanding, comparison with an additive model and a validation analysis are needed.

**Key Words:**  $m$ -space, vector analysis, hypersphere, multivariate time series, centralization, centroid vector, vector inverse, standardization, state vector, Importance Value, multivariate instantaneous trend, Trend analysis, trend vector, trend value, similarity coefficient.

## 1. Introduction

With ecosystems under increasing stress worldwide, numerous efforts have been undertaken to monitor changes in the biota. An example of such an effort is the U.S. Army Land Condition Trend Analysis (LCTA) program. From a network of thousands of permanent plots on military installations across the United States and Germany, data are collected periodically to characterize the natural resources and to track changes over time. Fort Carson, a 55,600-hectare training installation in the eastern foothills of the Rocky Mountains in central Colorado, has used LCTA to monitor land condition since 1986. Data from Fort Carson were used to test a new multivariate analysis model, the Multi-Dimensional Sphere Model (MDSM). The model analyzes multivariate instantaneous trend in a vector form to express the magnitude, direction, and rate of instantaneous change in vegetation composition at a given

time. In addition to trend analysis, the MDSM is used for correlation analysis, vegetation classification, and system monitoring.

The MDSM was designed for multivariate time series, three subscript data  $Z_{(i,j,k)}$ , or three way data. When used for vegetation analysis,  $i$ ,  $j$ , and  $k$  represent species, quadrats, and time, respectively. MDSM fixes  $i$  as the dimension of multivariate space,  $m$ -space; groups and combines  $j$  to eliminate it from the analysis; and performs trend analysis over time.

The principal advantage of the MDSM over other methods of multivariate analysis is its use of direction of a multivariate vector,  $m$ -vectors, to express the relation between the vegetation. Using  $m$ -vectors, represented here by bolded letters, MDSM can simultaneously analyze multiple species independently to reflect composition change. Other methods of multivariate analysis, such as least squares regression expressed by the formula

$$Y = a_{(1)}x_{(1)} + a_{(2)}x_{(2)} + \dots + a_{(m)}x_{(m)} + r,$$

are strongly influenced by a few dominant species. They only reflect changes in these dominant species, and are not capable of showing all vegetation composition changes. A second advantage of MDSM is that it works with species auto-correlation, instead of interrelation. MDSM emphasizes the study of vegetation state changes over time. When performing trend analysis, MDSM defines trend ( $t$ ) as present over past:

$$t_{(k)} = z'_{(k)} / z'_{(k-1)}$$

and extends this trend to the next time interval to make a projection ( $p$ ) based on existing information:

$$p_{(k+1)} = Z_{(k)} \times T_{(k)}.$$

A third advantage of the MDSM is that it uses instantaneous

**趋势分析** 及其在生态股市中的应用



trend, instead of long-term average trend. The time interval is more flexible, and the results more precisely reflect the vegetation changes. Simplicity of calculation is another advantage of the MDSM. All calculations can be done on a handheld calculator. The MDSM can be described simply as an extension of division and percentage calculations from a scalar to a vector.

The MDSM is demonstrated in this paper with a sample data set of two species and three quadrats (Table 1). The multi-dimensional space,  $m$ -space, built from these sample data is a 2-dimensional surface defined by the orthogonal axes of  $X$  and  $Y$  with three points  $A$ ,  $B$ , and  $D$ . The quadrats  $A$ ,  $B$ , and  $D$  are represented as points and defined by the 2-vectors  $\mathbf{a} = (40, 80)$ ,  $\mathbf{b} = (64, 77)$ , and  $\mathbf{d} = (117, 53)$  in Figure 1.

*T. Jay Bai et al. / Ecological Modelling 97 (1997) 75-86*

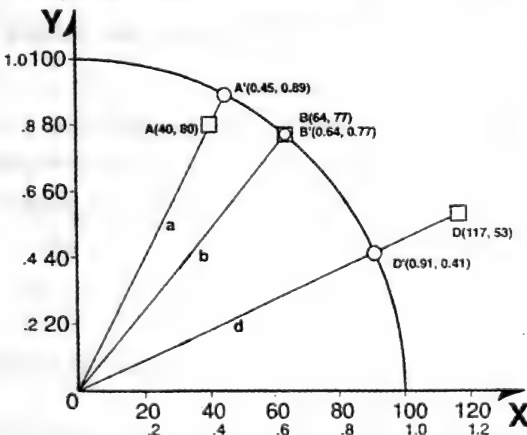


Fig. 1. Data standardization procedure to project data onto the hypersphere.

Table 1

Simplified quadrat-species matrix on which the MDSM is based.

Species data are abundance values (e.g., percent cover, density, or frequency)

	Quadrat A	Quadrat B	Quadrat D
Species x	40	64	117
Species y	80	77	53

## 2. Methods

Application of the Multi-Dimensional Sphere Model to vegetation analysis and synthesis involves four phases: quadrat data are divided by their vector length (standardization); determination of quadrat similarities and clustering of quadrats based on similarities (centralization); trend analysis and projection. These four steps are outlined below.

### 2.1 Step 1: Standardization of data by quadrat

Standardization of quadrat data projects the quadrat points from an  $m$ -space to the unit hypersphere by dividing each element of an  $m$ -vector by the vector's length.

$$Q' = \frac{Q}{L_{(q)}}$$

The length of a vector is the square root of the sum of the squares of the elements of a quadrat

$$L_{(q)} = \sqrt{\sum Q_{(i)}^2} = |Q|$$

therefore,

$$Q'_{(i)} = \frac{Q_{(i)}}{|Q|}$$





where  $Q$  is the quadrat,  $Q_{(i)}$  is the  $i$ th species of the quadrat,  $L_{(q)}$  is the vector length of the quadrat represented as  $|Q|$ ,  $Q'$  is the standardized quadrat, or the projection of the quadrat on the hypersphere, and  $Q'_{(i)}$  is the  $i$ th species of the standardized quadrat. In this paper  $Q'_{(i)}$  is referred to as the Importance Value (IV) of  $i$ th species. It can also be interpreted as the cosine value of  $i$ th component in the standardized vector. This process standardizes the different vector lengths to unity, while retaining their composition ratio.

The standardization procedure is illustrated in Figure 1. The two dimensional space defined by the orthogonal axes of  $X$  and  $Y$  contains three points,  $A$ ,  $B$ , and  $D$ , represented as squares. The lengths of these vectors are:

$$|A| = OA = \sqrt{40^2 + 80^2} = \sqrt{8000} = 89$$

$$|B| = OB = \sqrt{64^2 + 77^2} = \sqrt{10000} = 100$$

$$|D| = OD = \sqrt{117^2 + 53^2} = \sqrt{16498} = 128$$

Therefore, the standardized vectors  $\mathbf{a}'$ ,  $\mathbf{b}'$ , and  $\mathbf{d}'$ , shown as circles, are:

$$A' = \frac{A_{(i)}}{|A|} = \left(\frac{40, 80}{89}\right) = (0.45, 0.90)$$

$$B' = \frac{B_{(i)}}{|B|} = \frac{64, 77}{100} = (0.64, 0.77)$$

$$D' = \frac{D_{(i)}}{|D|} = \frac{117, 53}{128} = (0.91, 0.41)$$

After standardization, the lengths of standardized vectors are 1, and the end points of these vectors fall on the unit hypersphere:

$$OA' = \sqrt{0.45^2 + 0.90^2} = 1$$

$$OB' = \sqrt{0.64^2 + 0.77^2} = 1$$

$$OD' = \sqrt{0.91^2 + 0.41^2} = 1$$

The standardized  $m$ -vectors retain the same ratio of composition:

$$A'(X:Y) = 0.45:0.89 = (40/89):(80/89) = 40:80 = A(X:$$

$Y)$

for  $m$ -vector  $A$ , and

$$D'(X:Y) = 0.91:0.41 = (117/128):(53/128) = 117:53 = D$$

$(X:Y)$

for  $m$ -vector  $D$ .

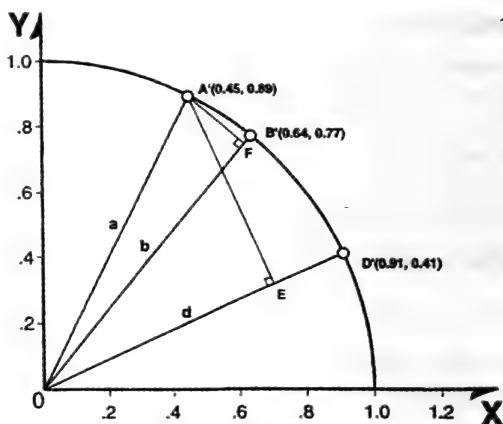


Fig. 2. Cosine values as Similarity Coefficients.

## 2.2 Step 2: Similarity Coefficient

The standardized  $m$ -vectors  $\mathbf{a}' = (0.45, 0.89)$ ,  $\mathbf{b}' = (0.64, 0.77)$  and  $\mathbf{d}' = (0.91, 0.41)$  are plotted again in Figure 2. Based on their components, quadrat  $B' = (0.64, 0.77)$  is located between  $A' = (0.45, 0.89)$  and  $D' = (0.91, 0.41)$ , and closer to  $A'$  than to  $D'$ .

When dimensions are more than two, or the quadrats are numerous, it is impossible to visualize the entire data structure. Consequently, a similarity coefficient is needed to determine the relative position



of the quadrats in  $m$ -space. The similarity coefficient between two  $m$ -vectors (quadrats) is defined by MDSM as the cosine of the angle between the two  $m$ -vectors (Gauch 1982, Ludwig and Reynolds, 1988; Orloci, 1967).

$$SC_{(a,b)} = \cos \angle AOB = \cos \angle BOA = \frac{\sum A_{(i)} \times B_{(i)}}{|A| \times |B|}$$
$$= \sum A'_{(i)} \times B'_{(i)}$$

Where SC is the Similarity Coefficient, and AOB is the angle between vectors **a** and **b**. Because the lengths of the standardized vectors are one, vector lengths are omitted in the following calculation:

$$SC_{(a,b)} = \cos \angle AOB = \cos \angle BOA = OF$$
$$= (0.45 \times 0.64 + 0.90 \times 0.77)$$
$$= 0.97$$

$$\arccos 0.97 = 13^\circ = \angle AOB$$

The angle between **a** and **b** is therefore 13 degrees.

$$SC_{(a,d)} = \cos \angle AOD = \cos \angle DOA = OE$$
$$= (0.45 \times 0.91 + 0.89 \times 0.41)$$
$$= 0.77$$

$$\arccos 0.78 = 39^\circ = \angle BOD$$

The angle between **a** and **d** is therefore 39 degrees. The cosine value calculated by MDSM shows that point  $B'$  is closer to point  $A'$  than to point  $D'$  in the 2-space. Therefore, quadrat  $B$  is more similar to quadrat  $A$  than to quadrat  $D$  in species composition.

### 2.3 Step 3: Clustering Quadrats

In trend analysis, MDSM analyzes vegetation state over time, as represented by state vectors. To determine a state vector of vegetation, the model groups quadrats based on their coefficients of similarity to generate a centroid  $m$ -vector representing the combined quadrats. We

assume similar vegetation types will react similarly to environmental stressors and will have the same trend. A result of quadrat standardization, or projection onto the hypersphere, is the formation of clusters of quadrats. As an example, Figure 3 shows the projected locations of three standardized quadrats  $A' = (0.45, 0.89)$ ,  $B' = (0.64, 0.77)$ , and  $D' = (0.91, 0.41)$ .  $B'$  is closer to  $A'$  than to  $D'$ . If  $A'$  and  $B'$  are combined to form a cluster  $G$ , the coordinates of  $G$  would be the average of  $A'$  and  $B'$ :  $A' = (0.45, 0.89)$  and  $B' = (0.64, 0.77)$

$$G = \frac{A' + B'}{2} = \left( \frac{0.45 + 0.64}{2}, \frac{0.89 + 0.77}{2} \right) \\ = (0.54, 0.83)$$

$G$  must be standardized in the same manner as described previously. Figure 3 shows that  $G$  is the center of chord  $A'B'$ , and  $G'$  is the center of arc  $A'B'$ .

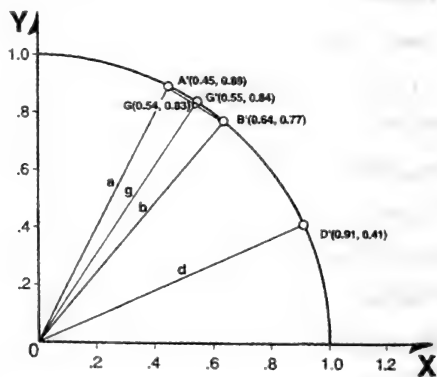


Fig. 3. Clustering of the standardized samples on the hypersphere.



#### 2.4 Step 4: Trend Vector

Discussion to this point has been based on quadrats sampled in the same time period but from different locations. However, for trend analysis, the MDSM is applied to quadrats sampled in different time periods but from the same location. We can use the quadrats  $A$  and  $B$  to define them as a time series, 1989 to 1993 for example, where  $1989 = 89' = A'(0.45, 0.89)$  and  $1993 = 93' = B'(0.64, 0.77)$ . The 2-vectors  $89'$  and  $93'$ , are defined as state vectors and represent vegetation conditions in 1989 and 1993 from the same location (Figure 4). If the coordinates of the two states are not the same, then some changes in composition has occurred during the sampling period. For natural resource monitoring and trend analysis, the following questions become important:

- 1) What are the magnitude and direction of tre changes in species composition during the sampling period?
- 2) If the trend remains the same, what would be the next state of the vegetation?
- 3) If the trend remains the same, then how long could this system remain in a steady stata, and when it does shift, what composition whould it change to?

Even when reasons for the vegetation change are not fully known, the MDSM assumes that people can make a projection based on present conditions. This situation can be expressed as:

$$89' \times t = 93'$$

where  $89'$  and  $93'$  are state 2-vectors representing the condition of vegetation in 1989 and 1993, respectively, and  $t$  is the trend vector representing the combination of the unknown factors that caused the

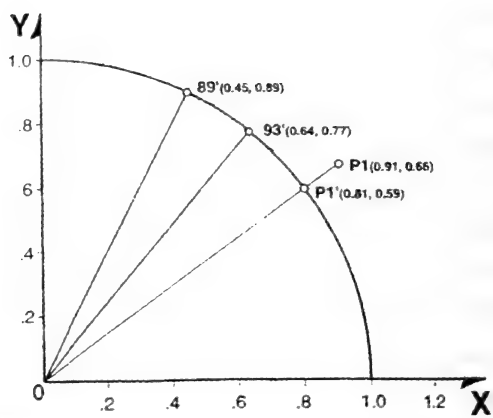


Fig. 4. Vegetation difference and trend from 1989 to 1993, and predicted state for 1997.

vegetation shifted from 1989 to 1993.

A basic premise of the MDSM is an alternative definition of multiplication and division of vectors, which has been used in spreadsheet software:

Current definition:

	Vector	Matrix
ADDITION:	$c_{(i)} = a_{(i)} + b_{(i)}$	$C_{(i,j)} = A_{(i,j)} + B_{(i,j)}$
SUBTRACTION:	$c_{(i)} = a_{(i)} - b_{(i)}$	$C_{(i,j)} = A_{(i,j)} - B_{(i,j)}$
MULTIPLICATION:	$c = \sum a_{(i)} \times b_{(i)}$	$C_{(i,j)} = \sum A_{(i,k)} \times B_{(k,j)}$
or	$C_{(i,j)} = a_{(i)} \times b_{(j)}$	
DIVISION:	No definition.	Exists only for some square matrices

Alternative definition:

MULTIPLICATION:  $c_{(i)} = a_{(i)} \times b_{(i)}$

The product of two *m*-vectors is an *m*-vector in the same *m*-space. The elements of the resultant *m*-vector are the products of the



corresponding elements of the two  $m$ -vectors.

SQUARE:  $\mathbf{a}^2 = \mathbf{a} \times \mathbf{a} = \mathbf{a}_{(i)}^2$

DIVISION:  $\mathbf{c}_{(i)} = \mathbf{a}_{(i)}/\mathbf{b}_{(i)}$

The quotient of two  $m$ -vectors is an  $m$ -vector in the same  $m$ -space. The elements of the quotient are the quotients of the corresponding elements of the two  $m$ -vectors.

INVERSION:  $\mathbf{a}^{-1} = 1/\mathbf{a}_{(i)}$

The inverse of an  $m$ -vector is an  $m$ -vector in the same  $m$ -space, whose elements are the inverse of the elements of the  $m$ -vector.

With this alternative definition of multiplication and division of vectors, the trend vector,  $\mathbf{t}$ , can be expressed as the quotient of 93' and 89':

if  $89' \times \mathbf{t} = 93'$

then  $\mathbf{t} = 93'/89'$

$\mathbf{t}_{(i)} = 93'_{(i)}/89'_{(i)} = 93_{(i)}/|93| \times |89|/89_{(i)}$

The  $t$ -value also permits a projection of the next and future values of the state vectors:

$B' \times \mathbf{t} + r_{(1)} = P_{(1)}$

and,

$B' \times t^2 + r_{(2)} = P_{(2)}$

.....

$B' \times t^k + r_{(k)} = P_{(k)}$ ,

where  $P_{(k)}$  is the projected future value of a state vector for  $k$  time intervals later, and  $r$  is the estimated error. For example, the time interval for the LCTA data at Fort Carson is four years. The given sample year was 1993. The prediction formula for the LCTA data is expressed as:

$93' \times \mathbf{t} + r = 1997$

$$93' \times t^2 + r_{(2)} = 2001,$$

.....

The time interval in the example is four years. The line connecting 89' and 93' in Figure 4 is indicative of the difference between the two states. The slope of the line indicates the direction and rate of change. The trend vector then operates on the state vector to project the state at the next time interval and beyond.

The trend vector  $t$  in Figure 4 is calculated as follows:

$$t = 93'/89'$$

$$t_{(i)} = 93'_{(i)}/89'_{(i)} = (0.64/0.45, 0.77/0.89) = (1.42, 0.86)$$

This trend vector shows that between the sampling dates, species  $X$  increased from 0.45 to 0.64, at a rate of 1.42; and species  $Y$  decreased from 0.89 to 0.77, at a rate of 0.86. Figure 4 also shows the dotted line of 89' 93' extended to  $P_{(1)}$ :

$$P_{(1)} = 1997 = B' \times t,$$

the values of  $P_{(1)}$  are:

$$P_{(1)} = B' \times t = (0.64 \times 1.42, 0.77 \times 0.86) = (0.91, 0.67)$$

This vector is then standardized giving  $P_{(1)}'$ :

$$P_{(1)}' = 97' = P_{(1)}/|97|$$

$$\begin{aligned} P_{(i,1)}' &= P_{(i,1)}/\sqrt{\sum P_{(i,1)}^2} \\ &= (0.91/\sqrt{0.91^2 + 0.67^2}, 0.67/\sqrt{0.91^2 + 0.67^2}) \\ &= (0.81, 0.59) \end{aligned}$$

The MDSM treats each species separately, and all species are included in the calculations, even those with comparatively small importance values. When performing instantaneous trend analysis, the model compares the present importance value of each species with its previous value to generate  $t$ -values. By comparing all  $t$ -values, a complete pic-





ture of vegetation community trend at given time can be obtained.

### 3. A trial trend analysis

The MDSM was applied to a trend analysis using Land Condition Trend Analysis data from Fort Carson. Results are presented in Table 2. For this analysis, 35 plant species were selected by correlation analysis and field experience. The correlation analysis was used to ensure that the selected species were not highly correlated. Species having a correlation of 0.97 or greater are combined to form a composite species. Species whose value fluctuates dramatically (from field observations) were not included in this trial analysis to avoid introducing too much noise. We did, however, keep a few widely fluctuating species, such as *Melilotus officinalis*, to show their activity in MDSM. Data were obtained from 199 permanent quadrats sampled in 1989 and 1993. To emphasize the trend analysis, all quadrats were combined and analyzed as one community representing Fort Carson as a whole.

Table 2

Results of a trial MDSM multivariate instantaneous trend analysis

Species	1989'	t-values	1993'	1997	1997'	2001	2001'
<i>Melilotus officinalis</i>	0.0042	10.5715	0.0440				
<i>Bromus japonicus</i>	0.0101	6.1026	0.0615				
<i>Suañon hystrix</i>	0.0203	3.5678	0.0724				
<i>Oryzopsis hymenoides</i>	0.0323	2.9095	0.0939	0.2731	0.2233	0.7946	0.1406
<i>Bromus inermis</i>	0.0160	2.0444	0.0327	0.0669	0.0547	0.1368	0.0242
<i>Juncus sp.</i>	0.0300	1.8238	0.0547	0.0998	0.0816	0.1820	0.0322
<i>Salix sp.</i>	0.0145	1.4994	0.0217	0.0326	0.0266	0.0489	0.0086
<i>Atriplex confertifolia</i>	0.0082	1.4497	0.0119	0.0172	0.0141	0.0250	0.0044

<i>Sarcobatus vermiculatus</i>	0.0032	1.4436	0.0045	0.0066	0.0054	0.0095	0.0017
<i>Atriplex canescens</i>	0.0097	1.3800	0.0134	0.0185	0.0151	0.0255	0.0045
<i>Frankenia jamesii</i>	0.0107	1.3681	0.0147	0.0200	0.0164	0.0274	0.0049
<i>Chenopodium incanum</i>	0.0112	1.3404	0.0150	0.0202	0.0165	0.0270	0.0048
<i>Cercocarpus montanus</i>	0.0568	1.1870	0.0675	0.0801	0.0655	0.0950	0.0168
<i>Quercus gambelii</i>	0.0672	1.1680	0.0785	0.0916	0.0749	0.1070	0.0189
<i>Pascopyrum smithii</i>	0.2018	1.1615	0.2344	0.2722	0.2226	0.3162	0.0559
<i>Oryzopsis micrantha</i>	0.0159	1.1457	0.0182	0.0208	0.0170	0.0239	0.0042
<i>Yucca glauca</i>	0.0406	1.1083	0.0450	0.0498	0.0408	0.0552	0.0098
<i>Pinus edulis</i>	0.2992	1.0781	0.3225	0.3477	0.2483	0.3749	0.0663
<i>Krascheninnikovia lanata</i>	0.0140	1.0657	0.0149	0.0159	0.0130	0.0169	0.0030
<i>Aristida purpurea</i>	0.1167	1.0588	0.1236	0.1308	0.1070	0.1385	0.0245
<i>Juniperus monosperma</i>	0.3439	1.0458	0.3597	0.3762	0.3075	0.3934	0.0696
<i>Rhus thilobata</i>	0.0276	1.0345	0.0286	0.0295	0.0241	0.0306	0.0054
<i>Pinus ponderosa</i>	0.0478	1.0236	0.0489	0.0500	0.0409	0.0512	0.0091
<i>Kochia scoparia</i>	0.0365	0.9956	0.0364	0.0362	0.0296	0.0361	0.0064
<i>Stipa sp.</i>	0.0964	0.9802	0.0945	0.0926	0.0757	0.0908	0.0161
<i>Bouteloua gracilis</i>	0.8238	0.9652	0.7952	0.7675	0.6275	0.7408	0.1311
<i>Andropogon gerardii</i>	0.0276	0.9613	0.0265	0.0255	0.0209	0.0245	0.0043
<i>Bouteloua curtipendula</i>	0.0689	0.9438	0.0651	0.0614	0.0502	0.0580	0.0103
<i>Opuntia imbricata</i>	0.0067	0.8890	0.0059	0.0053	0.0043	0.0047	0.0008
<i>Opuntia polyacantha</i>	0.0246	0.8482	0.0208	0.0177	0.0145	0.0150	0.0027
<i>Helianthus petiolaris</i>	0.0053	0.7638	0.0040	0.0031	0.0025	0.0024	0.0004
<i>Hilaria jamesii</i>	0.0509	0.6625	0.0337	0.0223	0.0183	0.0148	0.0026
<i>Gutierrezia sarothrae</i>	0.0338	0.5760	0.0195				
<i>Sporobolus sp.</i>	0.1337	0.5282	0.0706				
<i>Salsola kali</i>	0.0672	0.2802	0.0188				
Length	1	1	1	1.2231	1	5.6511	1
Trend Index		1.2188					



The second and fourth columns in Table 2 are the state 35-vectors for 1989 and 1993, shown as 1989' and 1993', respectively. These values are the standardized averages for each species over all 199 quadrats. The third column is the trend 35-vector. The elements in the trend vector,  $t_{(i)}$ , are the quotients of 1993'  $_{(i)}$  and 1989'  $_{(i)}$  and indicate the ratio of change for a given species during this time period. The expectation for the  $t$ -value is one. If the  $t$ -value for a species is equal to or greater than 1.0, the species' relative importance remained constant or increased over time. Conversely, if the  $t$ -value is less than 1.0, the species' importance value decreased over time.

The MDSM assumes most species usually behave linearly in a neighborhood, i.e., the expectation of  $t_{(i)}$  is one. (Jameson 1986). Data from Table 2 were sorted by the  $t$ -value in descending order. The top and bottom 10 percent of the  $t$ -values that are extremely away from one were excluded from projection to avoid skewing the projection. Examples of calculated species trends are: *Oryzopsis hymenoides* ( $t_{(4)} = 2.9095$ ), which increased by a factor of nearly three from 1989 to 1993; and *Bouteloua gracilis* ( $t_{(26)} = 0.9652$ ) which decreased slightly during the same time period. The fifth column in Table 2 is the projected state 35-vectors for 1997, while the sixth column is the standardized state 35-vector for 1997. The rows in Table 2 show the changes for each species over time. For example, the  $t$ -value for *Pascopyrum smithii*,  $t_{(15)} = 1.1615$ , indicates a slight increase over four years. Its importance value increased from 0.2018 in 1989 to 0.2344 by 1993, while the projected value for 2001 after standardization decreased to 0.0559. The increase in the projected value for *Pascopyrum smithii* and its concomitant decrease in importance value is a unique feature of the MDSM and can be explained by reference to other species

in the vector. Trend analysis by the MDSM is based on standardized vectors that include information from both the original vector lengths and their elements. The MDSM utilizes not only the ratios of species' values, but also the ratio of the lengths of the  $m$ -vectors, an indication of the overall vegetation condition. Although the instantaneous trend indicate the  $t_{(15)}$  is grater than one, compare with all other species, it' IV is relatively decreasing. (Average trend value, pseudo trend index, equals to 1.2. See below.)

The increase of *Pascopyrum smithii*. and decrease of *Bouteloua gracilis*, two important species in the shortgrass prairie at Fort Carson, may indicate a slight vegetation shift from *Bouteloua gracilis* to *Pascopyrum smithii*. This shift may be caused by management practices, such as removal of grazing pressure or soil surface damaged by military vehicles.

Further analysis of the trend values of *Bouteloua gracilis* can supply more information. As  $t_{(26)} = 0.9652$  and  $0.9652^{20} = 0.4924$ , MDSM shows that if the decreasing trend of *Bouteloua gracilis* continues for another  $4 \times 20 = 80$  years, then this important species may no longer be dominant in this ecosystem.

The average  $t$ -value of 1.2188 for 4th to 32nd species, the pseudo trend index, can be used as a relative indicator of overall condition to compare different sites. We used this pseudo trend index to label a map of Fort Carson with colors corresponding to levels of the index. The resultant map matched well with both the experience of the range manager and a usage map of Fort Carson. In another study utilizing the MDSM on an eleven-year data set, the predicted values were highly correlated with interpolated values and actual values ( $r > 0.94$ ).



## Conclusions and Discussion

The Multi-Dimensional Sphere Model constructs a unit hypersphere by using species as dimensions, and quadrats as standardized  $m$ -vectors whose endpoints lie on the hypersphere. The MDSM uses the cosine of the angles between  $m$ -vectors as similarity coefficients, indicating the relative position of the quadrats on the hypersphere. It also uses cosine values to establish the direction of the  $m$ -vector, and illustrates vegetation changes by comparing the previous and present cosine values. In other words, It is interpreted as instantaneous trend that the ratio of the  $m$ -state vector over time. The MDSM extends the trend to project the next and future states.

In the Fort Carson data, the number of quadrats ( $j$ ) sampled = 199, so there are 199 values for each plant species. MDSM used the centroid vector, made of averages, instead of a matrix to characterize the vegetation. This is important, as trend analysis assumes a homogeneous community, so there would be only one trend for a community at a given time. The values of the centroid vector and state vector fluctuate less as the number of quadrats increases.

MDSM considers the vector direction carries the composition information instead of distance. This is reflected in data standardization. Standardization also shows that MDSM excludes any specific interaction between special pairs. Each plant species interacts with every other species.  $Q_{(i)}$  reacts with the sum of  $Q_{(j)}$ ,  $j = 1, 2, \dots, i, \dots, n$ , including  $i$ . This interrelation of one to all is reflected in the standardization:

$$q'_{(i)} = IV_{(i)} = \frac{q_{(i)}}{|q|} = \frac{q_{(i)}}{\sqrt{\sum q_i^2}}$$

This describes a basic interrelationship of species within a community.

If we convert Brockwell's time series to vegetation, then the three parts of a time series can be expressed as:

$$Z_{(k)} = t_{(k)} + u_{(k)} + v_{(k)}$$

where  $t_{(k)}$  is the "trend component",  $u_{(k)}$  is a function with known period referred to as an "seasonal component", and  $v_{(k)}$  is a "random noise component" (Brockwell and Davis, 1991). The trend component remains in the data and can be analyzed after MDSM filters out the "random noise component (sampling error)" and "seasonal component (annual component)" by centralization and standardization, respectively. MDSM expresses the trend as Instantaneous Trend: present over past as an extension of "average changing trend". Instantaneous Trend can be calculated for any time interval supported by the data. This makes the trend analysis more flexible and more accurate.

In the LCTA example, the time interval for sampling was four years. The model projected vegetation condition at subsequent four-year intervals. However, the time interval,  $k$ , can be set at any fraction.

For the LCTA data, when  $k$  takes the value of 0.25,  $t$ -values are

$$t = \sqrt[4]{\frac{b}{a}}$$

With these  $t$ -values, the MDSM may be used to project vegetation condition in consecutive years instead of four-year intervals. It can also be used to interpolate values for the years 1990, 1991, and 1992.

The MDSM allows the investigator to use several scales in data collection (e.g. using cover and dominance for different species in the same study) without affecting the trend results. Such scale changes are a transformation. Transformations of scale can be useful in classifica-

趋势

势分析 及其在生态股市中的应用



tion interpretation because dimensions of an ordination can be expanded to increase resolution.

There are limitations to the application of the MDSM, because  $t$ -values are derived from the division of state vectors, and the elements of the state vector cannot be zero. It may be acceptable in some cases, however, to omit species from the analysis that have zero values. Another limitation is that the prediction error grows with increasing time intervals (Figure 4). All state vectors have a sampling error which could be passed to the  $t$ -values. When  $t$  is raised to the power of  $k$  for a projection, the error is also raised to the power of  $k$ . Thus prediction accuracy decreases with the size of  $k$ . To increase the accuracy of prediction, data must be collected in subsequent years to adjust the projection, in other word the  $k$  value should be kept small (Jameson 1986).

The MDSM projection of future states does not fall on a straight line because the  $t$ -values are raised to the power of  $k$ . The projection line will approximate a straight line only if all  $t$ -values and  $k$  are very close to 1.0. This suggests that the MDSM as trend analysis described here should not be used to monitor ecosystems undergoing dramatic changes. When the MDSM is used on a rapidly changing ecosystem, and greater accuracy of prediction is required, an additive model, linear model, may be more appropriate. MDSM, a multiplicative model, was compared with interpolation, an additive model, using an eleven-year data set. Results was presented to 49th annual meeting of society for Range Management, Wichita Kansas.

MDSM can be used to analyze vegetation trend, or changes, over time. Preliminary tests of the MDSM using LCTA vegetation data from Fort Carson in Colorado show the method to be very promising. Results

from the MDSM can be interpreted as a measure of successional change in vegetation. Linkage of the MDSM to succession theory will be useful in understanding the model. Tilman (1988) developed a theory of vegetation change that suggests secondary succession may be due to the transient dynamics of competitive displacement. A major feature of Tilman's theory is that plant life histories are a trade-off between growth rate and the ability to acquire resources. In early succession, rapid growth rates are advantageous for acquisition of territory. In later successional species, interactions become more important, and only strong competitors can increase in size. This theory suggests that in different series, species' growth rates are different on average, and that growth rate can vary for a single species through time, in response to competition. These changes in growth rates are of value when interpreting succession, and may be approximated (relative to the community) by the species transition vectors generated by the MDSM. In another words, the concept of a fixed intrinsic growth rate might be replaced by an empirical growth rate. Greater significance can be attached to the results from the MDSM when they are viewed from the standpoint of this theory of succession.

Back face

This research was conducted 1994, but published in 1997. There have some new developments that did not included in the published paper. Thus, the author posted a revised version in web page. Important concepts are  $m$ -space,  $m$ -vectors,  $m$ -exponential equation. There are differences between the slop and the ratio, projection and prediction. But these will be discussed in another paper.

趋势分析 及其在生态股市中的应用





## Acknowledgments

Initial guidance from Academician Bo Li at the Chinese Academy of Sciences (CAS) and Professor Jie Chen at Inner Mongolian University, P.R. China, were essential to the early formation of the model. Though not always in agreement, the first author's discussions with D. Jameson, Department of Rangeland Ecosystem Science, Colorado State University (C.S.U.), were extremely helpful. D. Anderson, Fort Carson LCTA coordinator, supplied the data. G. Gertner, Department of Forestry, University of Illinois, F. Smith, Department of Earth Resources and J. She, Department of Physics, C.S.U., reviewed earlier versions of this manuscript. CEMML staff at C.S.U. assisted with technical aspects of the manuscript preparation. Discussion of mathematics with P.D. Chen, Institute of Applied Mathematics, CAS, Y.L. Shi, Department of Mathematics, C.S.U, and D. Levinson, Department of Mathematics, Colorado College, were very helpful.

(Published in *Eldological Modelling* 97(1997) 75 – 86) by T. Jay Bai, Tom Cottrell, Dun-yuan Hao, Tala Te, and Robert J. Brozka)

## A Critique of Matrix Solutions for Ecology

The discussion about "A Critique of Ecology" on Ecol-1 is interesting. I think one of the reasons that ecologists have not been able to make predictions is that we did not have a suitable tool. When we handle every single variable, we can't find the rules behind the phenomena. When we handle the community as a whole using matrices, the variables can only be treated as linear. The formal name of matrix algebra is linear algebra. It has addition and subtraction, but generally does not have division or inverse, at least for ecological data. This fatal limitation has restricted its application on temporal or spatial dynamic analysis on ecology.

There is no definite matrix solution for the equation:  $AX = B$ , where  $A$  and  $B$  are two known sequential states of an ecosystem/community, and  $X$  is  $m \times m$  transition matrix that needs to be determined. We can consider two cases:

1)  $A$  and  $B$  are sample-variable matrices. There is no inverse for matrix  $A$ , thus there is no solution for  $X = B/A$ . Furthermore, if there was no inverse for  $A$ , then there wouldn't be an inverse in the forms of  $AA'$ ,  $AYA'$ , or  $YAY'$ , either. Although these solutions may be mathematically sound, they won't work for ecology.

The main reason that there is no inverse of a variable-sample ma-



trix for ecology is that the rank of the matrix is not full. As long as the samples were collected from a homogeneous vegetation/community, then even if they were sampled randomly, there would not be enough independent vectors in the matrix. All the samples would be varieties of the centroid vector. In other words, if the sampling variances were considered, then there would be only one independent vector in the matrix. Furthermore, if the matrix was not full rank, then there would not be a transformation that could increase the rank.

2)  $A$  and  $B$  are multi-component vectors,  $m$ -vectors, which were made of averages of the sample-variable matrix over samples, but  $X$  is an  $m \times m$  matrix. Since the number of unknown components,  $m \times m$ , is greater than the number of equations,  $m$ , there are many answers, i.e. no definite answer. The transition matrix  $X$  has definite solution, only if the number of components that need to be determined is  $m$ , but the other  $m \times (m-1)$  components are either known or zeroes. As we know nothing about the transition matrix, to determine these values of the  $m$  components, those other  $m \times (m-1)$  components should be set to zeroes first. Then the next question would be how to locate these  $m$  non-zero components and those  $m \times (m-1)$  zeroes in the transition matrix  $X$ . As there is "reproductive isolation", we can assume that the autocorrelation of a species would be greater than inter correlation with other species, i.e.,  $|X_{(i,i)}| > |X_{(i,j)}|$ ,  $i = 1, 2, \dots, m$ ,  $j = 1, 2, \dots, m$ ,  $i < j$ . Therefore, these  $m$  non-zero components should be placed in the diagonal positions of the transition matrix,  $X_{(i,i)} < > 0$ , but those  $m \times (m-1)$  zeroes should be in the off diagonal positions,  $X_{(i,j)} = 0$ ,  $i < j$ . It means that the transition matrix in ecology should be in diagonal form, and  $X_{(i,i)} = B_{(i)}/A_{(i)}$ ,  $X_{(i,j)} = 0$ .

However, a diagonal matrix is nothing else than a vector. This suggests that the  $m$ -vector with division and multiplication may be the correct tool for temporal dynamic analysis for community ecology.

If we treat our object as a community/system using  $m$ -vectors, and describe and analyze them by applying a vector algorithm, then we might discover some new rules for community/system. Then we could extend these rules to make projections for the future. This new data synthesis and analysis method based on  $m$ -vectors, named Multi-Dimensional Sphere Model(MDSM), was published recently in Ecological Modelling, 97/1-2, pp.75-86.

In other words, as the exponential growth is the essence of biology (Vandermeer 1981), we may adjust the exponential growth equation, by replacing the intrinsic rate with empirical rate and extending the mono-species model to a multi-species situation, then we may be able to use  $m$ -exponential equation to make predictions.

(Bulletin of the Ecological Society of America Vol.79 No 2, April 1998)



## Testing an Ecological Model using Mutual Fund Data

An ecological model, Multi-Dimensional Sphere Model (MDSM) originally developed for Land Condition Trend Analysis (LCTA), was further tested with mutual fund data. MDSM is a new multivariate data analysis method based on extended vector analysis. Vectors are the quantities stating both magnitude and direction. MDSM uses multi-component vectors, so it can handle more than three variables simultaneously. MDSM has defined vector division, so it can manipulate exponential growth. Thus, it is suitable for variables in sciences of ecology, financing, information, and economics. When applying to mutual fund data, MDSM built a multi-dimensional variable space corresponding to the funds. MDSM expressed the market with a point in the space, and defined it as market vector. It expressed the magnitude of the market with the vector length, but expressed the state of market with vector direction. MDSM used the quotient of cosine values, present over previous, to express the changing trends of the funds, and traded funds based on their trend values. The test was conducted from July 1998 to January 2000. The returns of an Individual Retirement Account directed by MDSM were 25.82% and 31.99% for the first seven months and the second year, while the average returns of the market were 2.64% and 9.82%, respectively. The conclusion is that this new data analysis method can be used for vegetation classification, rangeland monitoring, or temporal dynamic analysis for other systems.

(此文发表在《美国生态学会 2001 年年会》)

## Vegetation Science and Vector Analysis

**Abstract** Vegetation is a community of plants. Vegetation with multi-plant species can be described by multi-component vectors,  $m$ -vectors. An  $m$ -vector is a quantity that has magnitude and direction. The magnitude of the vector expresses the mass of the vegetation, while the direction of the vector expresses the composition of the vegetation. The directions of  $m$ -vectors are essential to quantify plant communities. Directions can be used for vegetation classification and vegetation dynamic analysis.

**Key words:** MDSM,  $m$ -space,  $m$ -vector, vegetation composition, community.

### Introduction

As part of the discussion on Ecol-L about "A Critique for Ecology", a commentary entitled "A Critique of Matrix Solutions for Ecology" was published in Bulletin of the Ecological Society of America, April, 1998. Since then, the author has received several responses about the commentary. Some of these responses suggested that a vector is a kind of matrix and that the author seemed to be inflating the importance of the vector.

This essay is to point out the distinction between the usage of vectors and matrices in vegetation analysis. Although both  $m \times 1$  matrices

趋势分析 及其在生态股市中的应用



and multi-component vectors,  $m$ -vectors, are multicomponent arrays and an  $m$ -vector can be treated as  $m \times 1$  matrix sometimes, the  $m$ -vector is not an  $m \times 1$  matrix. In fact, an  $m$ -vector is more than an  $m \times 1$  matrix. Vectors not only have the magnitude, the same as a matrix, but also have a direction, and are thus different from a matrix. By definition, a vector is a quantity that has magnitude and direction (Grolier, 1995). In Chinese, a vector is called *xiangliang*, which means direction (i.e., *xiang*), and magnitude (i.e., *liang*). It is the direction that makes vectors different from other quantities. It also makes a difference in applying an  $m$ -array to vegetation analysis, with or without direction.

## An Example for Vegetation

The following is an example of what the direction of an  $m$ -vector means to vegetation science. In vegetation science, people consider the composition of the vegetation to mean more than the total biomass. Imagine, a three dimensional space (3-space), where the three dimensions are Trees, Shrubs, and Grasses. In this 3-space, point  $A = (3, 1, 0)$  is different from point  $B = (0, 1, 3)$ , although the two have the same magnitude equal to the square root of 10. How does one distinguish two vectors when their magnitudes are the same? The answer is to use the position of the points in 3-space, in other words, the direction of the 3-vectors in the 3-space. The position of point  $A$ , or direction of the vector  $OA$  (where  $O$  is the origin, and  $A$  is the point in space), is closer to the first axis of Trees. Consequently,  $A$  can be classified as a woodland. On the other hand,  $B$  is closer to the third axis of Grasses, and may be classified as grassland. The two 3-vectors,  $A = (3, 1, 0)$

and  $B = (0, 1, 3)$ , represent different composition and are different types of vegetation. For the same reason, point  $2A = (6, 2, 0)$  in the same 3-space may be classified as woodland the same as  $A = (3, 1, 0)$ , since  $A$  and  $2A$  have the same composition ratio, yet  $A$  and  $2A$  have a difference in magnitude. This difference, however, is not essential to separate the two in vegetation science. This difference may be caused by different sampling areas or different sampling times. For example,  $2A$  might have twice the sampling area as  $A$  did, or  $2A$  might have a wetter sampling year than  $A$  did. However, those are minor factors, and would not change the vegetation classification. What is essential to vegetation analysis is whether  $A$  and  $2A$  have the same composition; that is, whether  $A$  and  $2A$  have the same direction in multi-species space ( $m$ -space) (Gauch, 1982). Generally speaking, all the points representing the same vegetation are located on the same ray in the  $m$ -space.

In other words, all  $m$ -vectors representing the same vegetation have the same direction in the  $m$ -space. Just as direction is essential to a vector, composition is essential to vegetation analysis. Different directions in  $m$ -space equate to different vegetation, and the same direction means the same vegetation. Furthermore, any compositional change in vegetation can be expressed by changes in vector direction. Therefore, vegetation dynamics, changes in composition, can be expressed by vector rotation in the  $m$ -space. People can monitor vegetation change by monitoring the rotation of the state vector of the vegetation in the  $m$ -space.

The direction of an  $m$ -vector in the  $m$ -space can be expressed by the vector's  $m$ -cosine values (Gauch, 1982, Jongman et. al., 1995):





$$\begin{aligned} \text{Direction } A &= \text{cosine } A_{(i)} = \frac{A_{(i)}}{|A_{(i)}|} \\ &= \text{Direction } 2A = \text{cosine } 2A_{(i)} \\ &= \frac{2A_{(i)}}{|2A_{(i)}|}, \quad i = 1, 2, \dots, m, \end{aligned}$$

where  $|A_{(i)}|$  is the magnitude of the vector, the square root of the sum of the squares (in another field, it is also called Shang Gao Index). For example, the direction of vector  $OA$  and  $O2A$  above can be expressed as

$$\begin{aligned} \text{Direction } A &= \text{cosine } A_{(i)} = \frac{A_{(i)}}{|A_{(i)}|} \\ &= \text{Direction } 2A = \text{cosine } 2A_{(i)} = \frac{2A_{(i)}}{|2A_{(i)}|} \\ &= \frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}, \frac{0}{\sqrt{10}}, \\ &= \frac{6}{\sqrt{40}}, \frac{2}{\sqrt{40}}, \frac{4}{\sqrt{40}} = A' \end{aligned}$$

We notice that the cosine  $\langle A \rangle = \text{cosine } \langle 2A \rangle = \text{cosine } \langle A' \rangle$  are also a 3-vectors, when  $A$ ,  $2A$ , and  $A'$  are 3-vectors. These are their projections on the three axes. Furthermore, the relation between any two vectors is expressed by a scalar which is the cosine value between the two vectors, or their projection to each other. For example, the

$$\begin{aligned} \text{cosine } \langle OA, O2A \rangle &= \text{cosine } \angle A-O-2A = 1, \text{ but the} \\ \text{cosine } \langle OA, OB \rangle &= \text{cosine } \angle A-O-B = 1/10 = 0.1. \end{aligned}$$

Using  $m$ -vector and  $m$ -space, people can build a standard vegetation classification system. This system can accept as many species as required, and can handle as many samples as are available. The differences in sampling size, such as  $10 \text{ m}^2$  versus  $100 \text{ m}^2$ , sampling shape,

such as line versus square, and measurement, such as, weight versus cover, would be filtered out by standardization. Thus, the only remaining information is the composition.

As different vectors can have the same magnitude, magnitude by itself tells us little about the composition of a vegetation. On the other hand, after knowing direction, people can determine any individual component or/and the vector magnitude given only a single component. For example, if we know the type of vegetation was woodland, and the components of the vegetation were 3, 1, and 0, or direction of the state vector in the 3-space is  $(\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}, \frac{0}{\sqrt{10}})$ , as discussed above, then given a shrub value of 2, we can project the vegetation as (6, 2, 0), and the magnitude of the vegetation would be calculated as the square root of 40. This is the principle that the Multi-Dimensional Sphere Model (MDSM) uses for system dynamic analysis.

## Conclusion

Vegetation occurs as communities, and communities can be accurately expressed as  $m$ -vectors. An  $m$ -vector is fully described by its magnitude and direction. Vegetation with  $m$  species is fully described by the magnitude and direction of its  $m$ -state vector. The magnitude of the vector expresses the amount of material, energy, and information that the vegetation contains, while the direction expresses the distribution of those material, energy, and information among the  $m$  components. The direction of a vector can be expressed by cosine values. The cosine value between two vectors expressing their relation is a scalar. However, the cosine values associated with each of the  $m$  axes express-

趋势

势分析 及其在生态股市中的应用



ing the vector's direction in the  $m$ -space is itself an  $m$ -vector. More specifically, it is the projection of the vegetation on the unit multidimensional sphere.

(Acta Ecologica Sinica 2002, 22(6):950-953)

## Multidimensional Sphere Model and its Application to Vegetation Classification

**Abstract** We introduce application of the Multi-Dimensional Sphere Model (MDSM) to vegetation classification. Vegetation samples can be projected as identities in a multi-plant species space. Samples can be classified by clustering their one to one images in the  $m$ -space. MDSM uses the cosine of the angle formed by two sample vectors in multi-dimensional space as a similarity coefficient. Cosine value expresses both species composition and relative abundance, reflecting the true structure of vegetation data in the  $m$ -space. The use of cosine values for similarity coefficients is widely accepted by mathematicians and computer scientists, however, the method has received little evaluation in the ecological literature. In addition to cosine similarity, the MDSM procedure uses fuzzy matrix composition to conduct the similarity from a single member to all members of the group. These two features make the MDSM classification unique, relative to distance based vegetation classification methods. In this paper we apply MDSM to vegetation data and discuss the results. We provide access to an Excel spreadsheet program which performs MDSM, and to the data used in this paper.

**Key Words:** multivariate space, MDSM, vector analysis,  $Q$ -mode, normalization, unit hypersphere, chord distance, vegetation cluster analysis



## Introduction

Multi-Dimensional Sphere Model (MDSM) was designed for the Land Condition Trend Analysis program at Fort Carson Colorado. MDSM was first introduced in 1995 as a method for multivariate temporal dynamic analysis (TDA), or trend analysis, and further explored to cluster quadrats into groups using cosine values, formed a centroid based on these groups, and then performed trend analysis on vegetation data. TDA reveals progressive changes in a natural resource such as vegetation. When used for temporal dynamic analysis, MDSM calculates each component's changing trend and predicts the future state of vegetation.

Classification is useful in landscape management for grouping or clustering entities based on a measure of similarity (Gauch 1982), yielding distinct groups which may be useful units for natural resource management or temporal dynamic analysis. While determination of groups of samples (clusters) that comprise a coherent unit is often the purpose of a classification, vegetation classification is also a necessary precursor to TDA. TDA requires a set of relatively homogeneous samples to provide a single moving trace in  $m$ -space. MDSM is a complete data analysis tool for vegetation resources: it allows the determination of management units (clusters), their changes over time, and can be used to assess data homogeneity.

In the current paper we discuss in detail the use of cosine value as a similarity coefficient and report on the use of MDSM for numerical classification. Further we describe the application of MDSM to a small vegetation data set. We provide access to an Excel workbook and data to perform the analyses. The workbook will be used as an outline for

our discussion. Interested readers are encouraged to download the MDSM workbook at <http://www.cwu.edu/biology/faculty/currentFaculty/cottrell/index.html> and study the use of MDSM with the provided data or their own data.

## Background

### Vegetation and its expression in multi-dimensional space

As with most multivariate vegetation data analysis, MDSM considers vegetation as a multi-dimensional hypervolume, where the  $m$  dimensions correspond to plant species that make up the vegetation. Every vegetation sample is an  $m$ -component vector mapped to a point on a unit hypersphere, and samples and points in  $m$ -space are related one to one.

Our vegetation data includes 87 plant species for which we have cover data, corresponding to 87 dimensions in  $m$ -space, here  $m$  equals 87. Six letter abbreviations of all plant species are listed in column A of the Excel workbook page DATA. Our 16 vegetation plots (samples, or observations) are represented by 16 points in this 87-space. For example, in the B2:B88 on the DATA page, the 87 numbers are ordinates of our sample A, and express position of the sample in 87-space. Because our data contain 16 plots (each with numerous species) we have 16 points (16 position vectors) in the 87-space.

Multidimensional space is visualized as either a sample space ( $R$  mode) which is based on resemblances between objects, or a variable space ( $Q$  mode) based on resemblances between descriptors. MDSM was designed for  $Q$  mode analyses in which the first step is to compute a matrix of association among the samples. This is how we apply MDSM

趋势分析 及其在生态股市中的应用



in classification and trend analysis, however, MDSM allows either approach in data analysis simply by inverting the data table.

The distribution of the sample points or vectors in multi-dimensional space expresses the important aspects of data structure. Samples are combined into groups when they are close in the  $m$ -space, i.e. similar. Generally, the number of shared species is divided by the total number of species in the two sites and the similarity coefficient is expressed as percentage. A threshold similarity value can be specified, resulting in the fusion of similar samples into coherent groups, which are visualized by a dendrogram. Usually vegetation studies require explicit stand selection criteria to ensure adequate homogeneity among samples. After data collection, cluster analysis can be used to assess homogeneity to ensure that samples are similar enough to allow synthesis into meaningful groups. Outliers will be indicated by their failure to coalesce into the other samples at either *a priori* or *a posteriori* specified threshold levels of similarity. Cluster analysis is thus an explicit filter that allows the removal of non-homogeneous samples.

#### **Cosine Similarity Index:**

The MDSM workbook includes three separate but integrated units. The first unit calculates a similarity coefficient (workbook page S for similarity) for all samples. The second unit conducts fuzzy matrix composition on the calculated similarity coefficients in order to meet the conductivity requirement. The third unit of the workbook clusters samples based on their similarity, and is run iteratively until all clusters coalesce into a single group.

Similarity coefficients are widely utilized in vegetation science. Euclidean distance is one common measure of similarity between samples (Jongman et al. 1995):

$$ED_{(A,B)} = \sqrt{\sum_{j=i}^p (Y_{Aj} - Y_{Bj})^2}$$

where  $Y_{Ai}$  is the abundance value for species  $i$  in sample  $A$ . MDSM utilizes an alternative similarity measure called cosine similarity where the similarity of samples (or clusters) is defined as the cosine of the angle between their respective vectors. If the total number of species is  $m$ , then we can use a dot product to quantitatively express the abundances of the shared species of any two samples as a measurement of closeness. Here, the closeness of two identities in  $m$ -space is the same as similarity in community ecology (Affifi and Clark, 1984).

The dot product is:

$$\sum (Y_{ia} \times Y_{ib})$$

where  $i = 1, 2, \dots, m$

Note that in this equation  $Y$  indicates the abundance value of component  $i$  in samples  $a$  and  $b$ . The combined total number of species in samples  $a$  and  $b$  is indicated by  $m$ . The product of  $Y_{ia}$  and  $Y_{ib}$  equals zero when either  $Y_{ia}$  or  $Y_{ib}$  is zero, and is only positive when both sites are non zero for component  $i$ . The maximum value possible for a given dot product occurs when the component values are identical in the two samples.

This closeness measure is relativized by division by the maximum possible dot product. Because sample  $a$  is the most similar to itself, its maximum closeness value is  $\sum (Y_{ia} \times Y_{ia}) = \sum Y_{(ia)}^2, i = 1, 2, \dots, m$ , similarly, the maximum closeness value for  $b$  is  $\sum (Y_{ib} \times Y_{ib}), i = 1, 2, \dots, m$ . The geometric average of these two closeness values is the square root of the product of the two:

$$\text{Geometric Average} = \sqrt{\sum Y_{ai}^2 \times \sum Y_{bi}^2}$$

where  $i = 1, 2, \dots, m$ .







This provides a very good calibration for a similarity measurement.

Utilizing the dot product of  $a$  and  $b$  and dividing by the geometric average, we define our cosine similarity coefficient:

$$S. C_{ab} = \sum \left( \frac{Y_{ia}}{\sqrt{\sum Y_{ia}^2}} \times \frac{Y_{ib}}{\sqrt{\sum Y_{ib}^2}} \right) = \cosin_{(a,b)}$$

where  $i$  is the index of species,  $m$  is the total number of plant species and  $Y$  indicates species abundance or other sampled value. This is exactly the cosine value between the two sample vectors in  $m$ -space.

Interested readers might want to compare this similarity coefficient value in  $Q$  analysis with the correlation coefficient  $r^2$  in  $R$  analysis. The similarity coefficient in MDSM is indexed by the total number of species ( $i = 1, 2, \dots, m$ ) while  $R$  analysis is indexed by the total number of samples ( $j = 1, 2, \dots, n$ ). Further, the correlation coefficient  $r^2$  is the cosine value between two variables measured from the centroid of the sample space. The similarity coefficient in  $Q$  analysis, however, is the measurement from the original data, not the data centroid. To further illustrate, below is the correlation coefficient  $r^2$ :

$$r^2 = \frac{\sum_j ((Y_{aj} - \bar{Y}_a) \times (Y_{bj} - \bar{Y}_b))}{\sqrt{\sum_j (Y_{aj} - \bar{Y}_a)^2 \times \sum_j (Y_{bj} - \bar{Y}_b)^2}}$$

where  $n$  is the total number of samples,  $Y$  indicates the variable values, and  $j$  is the index of samples.  $R$  analysis can be seen here to be (centered) by subtraction of the mean values of  $Y_a$  and  $Y_b$ .

Cosine similarity coefficient can also be expressed as product of two standardized sample vectors, the page entitled COS in the MDSM Excel worksheet shows similarity coefficients calculated from two standardized samples:

$$S. C_{ab} = \sum \left( \frac{Y_{ai}}{\sqrt{\sum_i Y_{ai}^2}} \times \frac{Y_{bi}}{\sqrt{\sum_i Y_{bi}^2}} \right) = \cosine(a, b)$$

where  $i = 1, 2, \dots, m$ .

The divisors in this equation are the vector lengths, or norms. The workbook calculates vector lengths (which are the square-root of the sum of the square, or SSS) and stores them in row 89 of page DATA. The procedure of dividing a vector by its norm converts all vector lengths to one, causing the projection of all sample points in  $m$ -space to fall on the unit hypersphere. From this, the name Multi-Dimensional Sphere Model or MDSM is derived.

The cosine value between any two vectors is calculated in the workbook by the macro SIMILARITY and stored on the workbook page S. For example, cell B3 in page S indicates the cosine similarity coefficient (SC) between sample  $A$  and  $B$ . Cell B17 in page S indicates the SC between samples  $A$  and  $P$ , and so on.

Since  $SC_{(AB)} = 0.6127$ , and  $SC_{(AP)} = 0.5031$ , sample  $A$  is closer (more similar) to  $B$  than to  $P$  in the 87-space. By checking the original data and noting species composition we see this is correct.

### Clustering method using fuzzy matrix composition

The second and third units of the MDSM workbook cluster samples based on their similarity in multivariable space. Eventually all the samples are joined together to form one cluster. This large cluster results from the progressive joining of individual samples and later, primary and secondary clusters. The clustering method used by MDSM involves fuzzy matrix composition (FMC) and conducts similarity coefficients between samples. As an example, this occurs when  $A$  is similar to  $B$ , and  $B$  is similar to  $C$ , then these similarities can be used to relate  $A$  to  $C$ . In cluster analysis after  $A$ ,  $B$ , and  $C$  cluster into a group, the decision of whether  $D$  should join  $ABC$ , or  $EFG$  is based on the similarity of the entire group with  $D$ . The method for joining primary groups (or sample units) into secondary groups defines the different



clustering methods.

The basic formula used in fuzzy matrix composition is given by

$$F_{(a,b)}^2 = \text{MAX}(\text{MIN}(F_{(a,j)}, F_{(j,b)}))$$

in which the elements of the composed fuzzy matrix  $F^2$  are the maximum values of the two corresponding minimum values for each species in a previous fuzzy matrix  $F$ . FMC conducts fuzzy matrix composition on the primary similarity matrix  $S$ , to form  $F$ , then  $F^2$  then  $F^4$ , . . . until  $F_n$  is formed.  $F_n$  represents the final fuzzy matrix of similarity coefficients. In the  $F_n$  matrix each cluster will have a single similarity with any other cluster, regardless of the number of samples comprising the cluster, and at this stage the matrix meets the requirement of conductivity and can be used for classification of the data set. Application of FMC a number of times equal to or less than half the matrix dimension will make the fuzzy similarity matrix meet the conductivity requirement for classification (Wang, 1980). If the conductivity requirement was not met then sample  $D$  would have six similarity coefficients with groups  $ABC$  and  $EFG$ .

Methods other than fuzzy matrix composition handle multiple similarity coefficients in a variety of ways. Furthest neighbor clustering defines the distance between any two clusters as the greatest distance (or least similarity) between any two members in the different clusters. A second common method UPGMA (unweighted pair group average) uses the average distance between all member pairs in the two different clusters (Sneath and Sokal, 1973). Choice of distance coefficients is well explained in Legendre and Legendre (1998). While UPGMA, furthest neighbor, and other methods are more common, we have found several published records involving use of fuzzy matrix composition in natural resource management (e.g. Schaefer et al., 2001). To date, however,

fuzzy matrix composition has not been widely used for ecological questions.

An instructive comparison involves our data matrix prior to application of FMC (in the S matrix of the MDSM workbook) to the data matrix after application of FMC (F<sub>-</sub> matrix of the workbook). Refer to Figure 1, and the Excel workbook, for the following discussion.

Any sample in S matrix has a number of similarity coefficients equal to the number of samples in the data. But in the F<sub>-</sub> matrix, sites may have multiple identical SC. For example, in our vegetation data, sample A in S matrix has 15 different SC's, one with each other sample. However, in matrix F<sub>-</sub> sample A has only one similarity coefficient (0.6355) for all 15 other samples and therefore will join all other samples at one time (Figure 1). The late fusion of sample A with all other samples indicates that it is least similar to the other vegetation samples. By contrast sample B from matrix F<sub>-</sub> has its highest SC (0.9122) with sample I. This results in the joining of samples B and I. Once B-I forms, this cluster has an SC of 0.8387 with a large cluster: N-O-M-G-C-D-F-K and will fuse at this similarity level.

After application of FMC any sample in the F<sub>-</sub> matrix will have an identical SC with all component sites of any other extant cluster. This illustrates the conductivity of similarity among samples, and it also explains why the samples must have a certain level of homogeneity before clustering can begin in temporal dynamic analysis of vegetation data.

## Example Analysis and Results

### The vegetation data

The Algific talus slope data are from a small data set collected in



1997 and 1998. Data consist of cover values for all vascular plants in a rare plant community in Iowa, USA. Algific talus slope vegetation occurs on steep unstable talus slopes. These slopes allow cold air drainage from the permanent ice caves behind the talus, and are notable for their prevalence of cold tolerant northern species. Vegetation cover data were recorded by cover classes, then standardized to the cover class midpoint. Means calculated from  $25 \times 1/4\text{m}^2$  quadrats comprise each sample. Quadrat placement on transects utilized random starts with following systematic placement. The data set consists of 16 samples, and a total of 87 species.

#### Algific talus slope clustering results:

MDSM joined the 16 samples into a few primary clusters, while two samples, *A* and *J*, join at lower similarity coefficients of 0.6355 and 0.7000, respectively (Figure 1). All other samples join at similarity coefficients greater than 0.80. Several tightly linked clusters are evident in the dendrogram, including samples *N*, *O*, and *M*; samples *C* and *D*; and samples *E* and *L*. Two species, *Cystopteris bulbifera* and *Impatiens pallida*, account for most of the data variation. The former comprises 25% of the total variance, while the latter comprises 51% of total variance. We will briefly discuss the biology of these clusters here.

Cluster *NOM* consists of 3 vegetation samples (25 quadrats each) adjacent to the same cliff system, within 2 km of one another. The samples were taken on steep and unstable talus sites. The dominant species in these samples is *Cystopteris bulbifera*. These three samples are markedly different from other samples by their small but important cover values for *Solidago flexicaulis*, a common species in moist woods of northeast Iowa. However, it is not clear whether *S. flexicaulis* indi-

icates any specific abiotic or community attribute about these three samples. All vegetation data contain a certain amount of stochasticity: some species will be present simply because their seeds fall in an environment that is not hostile. On the other hand it is likely that the very cold soil temperature at all three of these sites results in their similar vegetation. The close proximity and similar geology of these samples yield similar plant community attributes such as species composition and abundance.

A second cluster, *CD*, in addition to having large amounts of the fern *Cystopteris bulbifera*, both *C* and *D* are marked by an abundance of *Rubus ideaus*. Both samples are notable for their heavily shaded and unstable talus. The trailing growth form of *R. ideaus* allows it to cover areas not otherwise easily inhabited by other plants. This growth form might be advantageous in the constantly moving rocky substrate.

With the addition of sample *G* the *NOM* and *CD* clusters join to form a major cluster with a similarity coefficient of 0.9125. This group of 6 samples characterizes the most pristine algific vegetation in the area. Human induced disturbances at all 6 sites were minimal, however, the unstable nature of the steep talus produces frequent natural disturbance events. It is interesting to note that while we interpret the clusters based on abiotic factors such as stability and soil temperatures, measures of these factors were not included in the data base. Thus, this provides a form of indirect gradient analysis involving clustered samples.

At the other end of the *C. bulbifera-I. pallida* gradient (exclusive of the 2 disjunct samples *A* and *J*), is the cluster *ELP*. The distinguishing vegetation feature linking these 3 sites is the dominance of *Impatiens pallida*. This species is common on moist wooded slopes, of-



ten on colluvial substrates. In our samples we found that *I. pallida* is perhaps more common in areas that are less shady than those dominated by *C. bulbifera*, though *I. pallida* is similarly well adapted to the loose rocky talus slopes. *Impatiens pallida* is a highly competitive herb that can reach impressive heights (1.5 m or more), overtopping and shading most other forest herbs.

The only other tight cluster formed is that of samples *B* and *I*. These samples are marked by moderate amounts of both *C. bulbifera* and *I. pallida*. Their similarity in composition and abundances, particularly their shared high cover values for *Polymnia canadensis* resulted in their early fusion with one another. We have noticed *P. canadensis* is common at cliff bases, perhaps profiting from increased shading and runoff. Both sites *B* and *I* are marked by steep cliffs which rise abruptly at the upper end of the sampled slopes, and by an associated swath of *Polymnia canadensis*.

The tight clustering of sites *NOMGCD* and perhaps also sites *F* and *K* is suggestive that these samples depict heavily shaded Algific slopes with very cold soils. The remaining samples (again excluding *A* and *J*) are found on less steep talus with perhaps greater solar incidence.

Samples *A* and *J* are significantly different from the other 14 samples. Reference to the species abundance values show that each of these samples is dominated by a different species which is nowhere common in the other samples. Sample *A* is unique in being dominated by *Cornus alternifolia*. This species is an associate of *Acer saccharum* and occurs in moist woodlands, forest margins, and is shade tolerant. Sample *J* is dominated by *Eupatorium rugosum*, an herb found in rich woods, often at the base of wooded bluffs or on rock outcrops. While

this herb occurs in other samples, it is nowhere as dominant as in sample *J*. Presence of *E. rugosum* may indicate a level of disturbance not reached by the other samples. We note too that both samples *J* and *A* had low covers of *C. bulbifera* and *I. pallida*, further isolating them from the former samples.

The exposition of these outliers illustrates an important result of the application of MDSM to multivariate vegetation data. Discontinuities in vegetation data usually indicate dissimilar habitats or land use histories. This discontinuity was not immediately evident without use of a data reduction method. MDSM results suggest that these two outlying samples should not be treated in the same manner as the other samples during the development of a management plan for Algific slopes in northeast Iowa. Management decisions based on an average of all the sampled algific sites would likely not address the vegetation of the first 14 samples if sites *A* and *J* were included. The remaining 14 sites are more similar and can potentially be addressed by the same management options. If a finer grain management is desired the clusters with higher similarity can be used. The use of MDSM as a quality control method to discriminate outliers is a valuable tool for natural resource land managers.

## Discussion and Conclusions

We have provided a brief background to the cosine similarity coefficient utilized by MDSM. We describe the use of fuzzy matrix composition in clustering samples which are ordered by this similarity coefficient. The downloadable Excel workbook and provided data set will aid readers who might be interested in adopting MDSM in their vegetation





resource management programs.

Vegetation data are multi-dimensional and are easily handled as vectors. Vector direction describes data composition, while vector length defines element abundances. Vector length can also be viewed as a measure of magnitude, or energy, when considering, for example, range production. While vectors can be of any length (i.e. abundances can take any value), standardization to unit length in the MDSM procedure removes this variability from the data. After this standardization, vector direction measured by cosine becomes useful as a similarity coefficient.

In general, as the relative cover of a plant species increases, that species is more descriptive of the vegetation. We should emphasize, however, that it is the relative cover which is important, not the absolute cover. For example, by our cosine similarity measure the three element vector  $A = (1, 2, 3)$  and a second with the same relative cover  $10A = (10, 20, 30)$  represent the same vegetation. The different lengths of the vectors of sites  $A$  and  $10A$  may simply result from different sample areas. For example, sample  $10A$  may contain 10 times the area of  $A$ , or perhaps one was sampled in a dry year and one in a wet year. It is the vector direction (angle), rather than magnitude, which carries the essential information of the vegetation. MDSM projects the points from  $m$ -space to a unit hypersphere eliminating the differences in vector lengths and thereby filters out the noise from sample area and sample date while emphasizing the vegetation composition. We believe that the angle between vectors is more germane than distance between projected sites for vegetation science. Comparison of vegetation (vectors) can be accomplished by reference solely to the cosine of the angle between vectors.

Using vectors to express the vegetation is the basis of MDSM. To classify the vegetation samples we cluster their images in  $m$ -space. These images are generally distributed unevenly in the  $m$ -space. Certain areas of the space will have low densities of samples while others will have high densities. The goal of a classification is to use the low-density areas as belts to separate higher density areas. Most classification software utilizes some measure of distance, or converted distance to determine cluster membership. Euclidean distance is commonly used but when a Euclidean distance measure is based on species abundance it can suffer from the paradox that two communities sharing all their species may be at a greater Euclidean distance than two communities sharing no species. This paradox is solved by using the Euclidean distance between the two points that have been standardized to the unit hypersphere. This distance is called the chord distance. The essential part of Orloci's chord distance, however, is the cosine of the angle between the two site vectors and the information in Orloci's chord distance is completely captured in the cosine. Use of the cosine of the angle between two sites is the MDSM approach.

Once the similarity coefficient between two samples is determined, the problem of identifying clusters of samples based on this similarity arises. Our method of fuzzy matrix composition produces results similar to nearest neighbor clustering. Nearest neighbor clustering has been criticized for its tendency to produce clusters which link ever more dissimilar units. This is referred to as the problem of chaining. We will address this issue in depth in a later paper, but at this point we suggest that MDSM produces the least variance projection of data onto the hypersphere. This can be visualized by an example of a bivariate normal sample. Such a sample has a concentration ellipse with its

趋

势分析 及其在生态股市中的应用



long axis on a radius, and which can be standardized and projected on to the unit circle. The resulting distribution will be the least variance projection. In this view, chaining does not play a role in the classification. However, it is possible that if a classification continues until all samples are included, and some samples are much different than others, unlike samples will eventually be linked. Careful screening of samples prior to classification will eliminate these linkages of highly dissimilar samples.

We have shown that the fuzzy clustering method applied to the MDSM cosine similarity coefficient yields interpretable groups when applied to a data set having both similar samples and samples which are discontinuous. This method is both theoretically sound and mathematically simple. We believe that the MDSM procedures are an important addition to the natural resource management toolbox and could have wider application to studies of biodiversity.

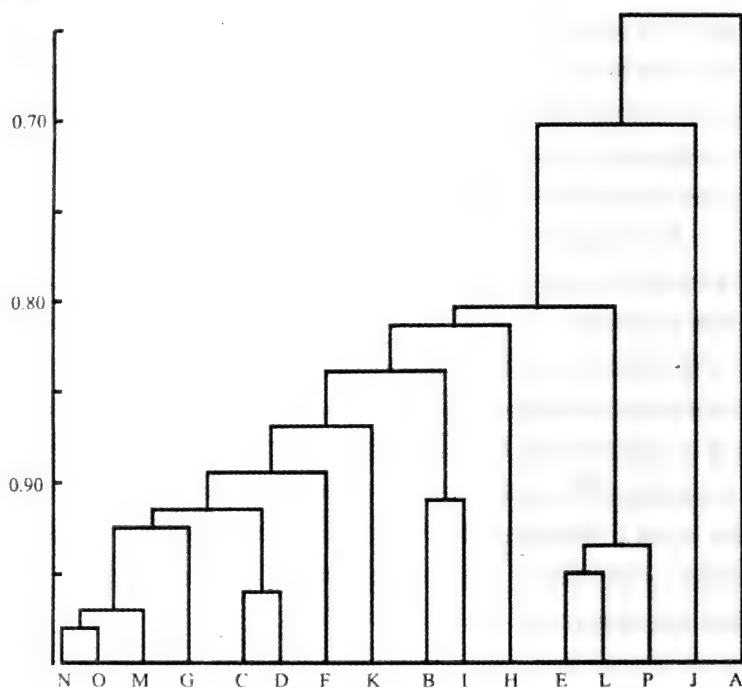


Figure 1. Dendrogram showing clusters of vegetation samples generated by fuzzy matrix clustering of cosine similarity values. Maximum similarity occurs between samples N and O. Samples J and A are outliers and do not coalesce into the larger group until similarities of 0.7000 and 0.6355.

By Tom R. Cottrell, T. Jay Bai and Dale Shanner.

Dept. of Biological Sciences, Central Washington University, Ellensburg, WA 98926, USA.



## 词汇、符号

- 多维空间** 用多个坐标轴撑起的空间，空间的点由一组有序的数，或由位置向量，连接原点和空中点的向量表示
- 变量空间** 以变量为坐标轴建立的空间；便于进行  $Q$  分析。
- 样本空间** 以样本为坐标轴建立的空间；便于进行  $R$  分析。
- $m$ -空间** 由  $m$  个变量为坐标轴构成的  $m$  维变量空间。
- $m$ -数组** 有一定序列的  $m$  个数，与  $m$  维空间的点（位置向量）对应。
- $m$ -空间的点** 空间点的数值表示为  $m$  维数组，用点的坐标表示。数组中的数称为分量。
- 向量** 既有方向又有量值的量。
- 位置向量** 连接原点和空间中的点的向量，用该点的坐标（数组）表示。与  $m$ -空间的点或  $m$ -数组相同。
- $P$**  大写字母表示空间的点  $P$
- $PQ$**  以点  $P$  和点  $Q$  为端点的一条线段
- $PQ$**  斜体大写表示有方向的，起点为  $P$  终点为  $Q$  的向量（又称矢量）

向量  $PQ$   
 $OP, P$

**P**

$|p| = |OP| = OP$

$\langle p, q \rangle$

$\cos \langle p, q \rangle$

和向量

差向量

积向量

商向量

逆向量

中心向量, 形心向量

同簇向量

中心化

向量标准化

余弦值向量

状态向量

趋势向量,  $T$

预报向量,  $P$

由  $Q$  点坐标减去  $P$  点坐标所得的向量。  
 起点为原点, 终点为  $P$  的向量, 原点  
 $O$  可以省略不写;

黑体小写也表示向量。

向量的长度, 模, 等于各分量的平方和  
 的算术根。

向量  $p, q$  的夹角,  $\angle POQ$

两向量夹角的余弦值

以分量的和做分量的向量。

以分量的差做分量的向量。

以分量的积做分量的向量。

以分量的商做分量的向量。

积向量的所有分量等于 1 的两向量互为  
 逆向量; 分量互为倒数的两向量; 以分  
 量的倒数做分量的向量。

若干个向量的分量的代数平均所组成的  
 向量; 是多边形的形心。

组成形心向量的各向量互称同簇向量。

求形心向量的过程。

向量除以向量长度, 把空间的点投影到  
 单位超球面的过程。

标准化向量; 分量等于余弦值。

经过标准化, 中心化, 可以代表系统状  
 态的向量。

表示系统运动趋势的向量, 是后前两状  
 态向量的商。可以被解释为状态转移向  
 量。

把趋势值作用于系统得到的向量, 用来

趋

势分析 及其在生态股市中的应用



预报。

观察值和预报值的加权平均。

第  $i$  变量时间  $k$  的趋势值。

第  $i$  变量时间  $k$  的预报值。

第  $i$  变量时间  $k$  的期望值。

第  $i$  变量时间  $k$  的误差值。

自变量和因变量的关系： $Y = a + bx$ 。

自变量和因变量的关系： $Y = a \cdot e^{(bx)}$ 。

双下标变量，可以列表表示。

非对角线元素为零的矩阵。

记录  $N$  个样本  $M$  个变量的， $M * N$  个数排成的矩阵。

三下标变量，三向数据，晶阵。

第一下标，标记变量，取值从 1 到  $m$ 。

第二下标，标记样本，取值从 1 到  $n$ 。

第三下标，标记时间，取值从 0 到  $o$ 。

三下标变量，“立体矩阵”。

把多维空间的点投影到单位超球面上，用平面几何的工具分析投影，推断多维空间点的分布，动态的数学工具。

提出“勾三股四弦五”的中国古代数学家。

股市上所有股票价格平方和的算术根。无偏估计。

多个相同或相类的事物按一定的秩序和关系所组合而成的总体。

组成系统的各元素由于竞争共同资源而

期望向量， $E$

$T_{(i,k)}$

$P_{(i,k)}$

$E_{(i,k)}$

$R_{(i,k)}$

线性增长，算术增长

指数增长，几何增长

矩阵

对角矩阵

样本 - 变量矩阵

$D_{(i,j,k)}$

$I$

$J$

$K$

晶阵

超球面模型

商高

商高指数

系统

竞争系统

系统惯性	互相制约的系统。 本书对系统的假定：系统的趋势值，二阶趋势值的期望值等 1；趋势值作用于系统产生预报值。
植被	由植物组成的地表覆盖物的总称。
同比植被	两块组成比例相同的植被；两者互为缩影，在多维空间共线。
草原	一种分布在开阔地，以旱生禾草为主的原生植被。
草场	作为放牧对象的草本植被。
草地	作为耕作对象的草本植被。
十一律	生态系统里每一个营养级的生物量大约是其下面营养级生物量的 10% 的自然规律。旧译作“金字塔”。
R-分析	在样本空间分析变量之间的关系。
Q-分析	在变量空间分析样本之间的关系。
技术分析	认为历史价格经过适当分析可以反映未来价格的市场理论。





## 参考文献

### 英文部分

1. Affifi, A. A. and V. Clark, 1984: Computer - aided Multivariate Analysis. Lifetime Learning Publications, Belmont California.
2. Allen, H. Franklin and R. Karjalainen, 1999: Using genetic algorithms to find technical trading rules, *Journal of Financial Economics* 51, 245 - 271.
3. Anonymous. 1985: Rangeland Monitoring: Trend Studies United States Department of the Interior, Bureau of Land Management technical reference 4400 - 4.
4. Babbs, Simon H., and K. Ben Nowman, 1999: Kalman filtering of generalized Vasicek term structure models, *Journal of Financial and Quantitative Analysis* 34, 115 - 130
5. Bai, T. J., and Linn, J., 1994: Range Condition Trend Analysis for Fort Carson by Multidimensional Space Model. Presentation at Land Rehabilitation and Maintenance 3rd Conference, Aberdeen, MD. USA.
6. Bonham, C. D. 1989: Measurements for Terrestrial Vegetation John Wiley and Sons N. Y.
7. Borland. Appendix D, 1987: Array features. in *Quattro Pro User's Guide*.

8. Brockwell, P. J. and Davis, R. A., 1991: *Time Series: Theory and Methods*. Second edition. Springer - Verlag.

9. Brown, J. Stephen, William Goetzmann, and Alok Kumar, 1998: *The Dow Theory: William Peter Hamilton's track record reconsidered*, New York University. Working paper.

10. Cowles, Alfred, 1934: Can stock market forecasters forecast? *Econometrica* 1, 309 - 324.

11. Crowe, Michael J., 1967: *A History of Vector Analysis*. University of Notre Dame Press.

12. Davis, Harry F., Snider, A. D., 1979: *Introduction to Vector Analysis*. Allyn and Bacon, Inc., 4<sup>th</sup> Ed.

13. Eberhardt, L. L. & Simmons M. A. 1992: Assessing rates of increase from trend data. *J. Wildl. manage.*, 56(3):603 - 610

14. Gauch, H. G., 1982: *Multivariate Analysis in Community Ecology*. Cambridge University Press.

15. Gerrodette, T. 1987: A power analysis for detecting trends. *Ecology*, 1987, 68(5):1364 - 1372

17. Howden, W. 1997: Systems Testing and Statistical Test Data Coverage, *Procs. COMPSAC '97*, Washington, DC, p. 500 - 504

17. Jameson, D. A. 1985: A Priori analysis of allowable interval between measurements as a test of model validity. *Applied Mathematics and Computation*, 17:93 - 105

18. Jameson, D. A. 1986: Sampling intensity for monitoring of environmental systems. *Applied Mathematics and Computation*, 18: 71 - 76

19. Jameson, D. A. 1989: *Environmental Monitoring and Adaptive Management*. Colorado State University.

20. Jassby, A. D. & Powell, T. M. 1990: Detecting changes in



ecological time series. *Ecology*, 71(6):2044 – 2052

21. Jolliffe, I. T., 1986: *Principal Component Analysis*. Springer – Verlag.

22. Jongman, R. H. G., C. J. F. Terbraak & O. F. R. Van-Tongeren, 1995: *Data Analysis in Community and Landscape Ecology*. Cambridge University Press.

23. Kapur, R., S. Patil, T. Snethen, T. W. Williams, 1996: A Weighted Random Pattern Test Generation System, *IEEE Transactions on Computer – Aided Design of Integrated Circuits and Systems*, vol. 15, no. 8, August 1996, p. 1020 – 1025.

24. Kent, M., and P. Coker. 1992: *Vegetation Description and Analysis*. John Wiley & Sons, New York, New York.

25. Legendre, P. and L. Legendre. 1998: *Numerical Ecology*: second English edition. Elsevier, New York, New York.

26. Leon, Steven J., 1994: *Linear Algebra with Applications*. 4<sup>th</sup> ed. Prentice Hall.

27. Ludwig, J. A. and J. R. Reynolds. 1988: *Statistical Ecology: A primer on methods and computing*. John Wiley & Sons, New York.

28. Malaiya, Y. K., 1995: Antirandom Testing: Getting the Most out of Black – Box Testing, *Procs. ISSRE '95*, Toulouse, Oct. 1995, p. 86 – 95.

29. Mardia, K. V., 1972: *Statistics of Directional Data*. Academic Press.

30. Mclendon, T. and Redente, E. F., 1991: Nitrogen and phosphorus effects on secondary succession dynamics on a semi – arid sagebrush site. *Ecology*, 72(6):2016 – 2024

31. Neely, Christopher, Paul Weller, and Rob Dittmar, 1997: Is

technical analysis in the foreign exchange market profitable? A genetic programming approach, *Journal of Financial and Quantitative Analysis* 32, 405 - 426.

32. Nichols, J. D., Sauer, J. R., Pollock, K. H., & Hestbeck, J. B., 1992: Estimating transition probabilities for stage - bases, population projection matrices using capture recapture data. *Ecology*, 73(1):306 - 312

33. Orloci, L. 1967: An agglomerative method for classification of plant communities. *Journal of Ecology* 55:193 - 205.

34. Pennacchi, G, George, 1991: Identifying the dynamics of real interest rates and inflation: Evidence using survey data, *Review of Financial Studies* 4, 53 - 86.

35. Schaefer J. A., A. M., Veitch, F. H. Harrington, W. K. Brown, J. B. Theberge, and S. N. Luttich, 2001: Fuzzy structure and spatial dynamics of a declining woodland caribou population. *Oecologia* 126:507 - 514.

36. SHINICHI TAMURA, 1971: Pattern classification based on fuzzy relations. Vo.1 SMC - 1, 1.

37. Sneath, P. H. and R. R. Sokal. 1973: *Numerical Taxonomy: The principles and practice of numerical classification*. W. H. Freeman and Company, San Francisco, California.

38. Spellerberg, I. F., 1994: *Monitoring ecological change*. Cambridge University Press.

39. Tazik, D. J., Warren, S. D., Diersing, V. E., Shaw, R. B., Brozka, R. J., Bagley, C.F., and Whitworth, W. R., 1992: U.S. Army Land Condition - Trend Analysis (LCTA) Plot Inventory Field Methods. USACERL Tech. Rep. N - 92/03. 62 p.

40. Tilman, D., 1988: *Plant Strategies and the Dynamics and*



Structure of Plant Communities. Princeton Univ. Press, Princeton, New Jersey.

41. Vandermeer, John, 1981: Elementary Mathematical Ecology. John Wiley & Sons, Inc.

42. West, N. E., 1983: Choice of vegetation variables to monitor range condition and trend. Pages 636 - 639 IN: Bell, J. F. and Atterbury, T. (eds.) Proceedings of International Conference on Renewable Resource Inventories for Monitoring changes. August 15 - 19, Oregon State university, Corvallis.

43. Whittaker, R. H., 1978: Classification of plant communities.

44. Williams, W. T. and Dale, M. B. 1965: Fundamental problems in numerical taxonomy. Adv. Bot. Res. 2:35 - 68.

45. Young, Eutiquio C. 1993: Vector and Tensor Analysis. Marcel Dekker, Inc. 1993.

## 中文部分

1. 巴世杰, 于中流等, 1984: 对内蒙古的饲草资源及其养畜的初步估价。

2. 北京工业大学计算站, 1976: 怎样使用计算机。科学出版社。

3. 查普曼等, S.B., 1976: 植物生态学的方法。阳含熙等译, 1980, 科学出版社。

4. 何森等, 1995: 板齿鼠种群中长期预测的时间序列模型。走向 21 世纪的中国生态学。中国生态学会。

5. 黄金嗣, 1980: 海晏县草原退化调查及防治措施的探讨(内部发行)。

6. 吉林师范大学草原研究室, 1960: 东北西部及内蒙东部的草原。东北西部及内蒙东部第一次草原科学报告会论文集, 5-17。

7. 考克斯, G.W., 1972: 普通生态学实验手册。蒋有绪译, 1979, 科学出版社。

8. 李博, 1962: 内蒙古地带性植被的基本类型及其生态地理规律。内蒙古大学学报(自然科学), (2) 47-74。

9. 刘钟令, 1960: 内蒙古草原区植被概貌。内蒙古大学学报(自然科学), (2) 41-47。

10. 刘书润, 1981: 内蒙古锡林河流域植物名录(油印稿)。

11. 美利坚百科全书 1999 年网络版, 向量分析词条, Gibbs 词条。

12. 内蒙古草原综合实验研究中心地区综合考察队, 1965: 内蒙古锡林郭勒种畜场地区的植被与草场资源。内蒙古草原研究所 1975 年铅印。

13. 皮洛, E.CY., 1969: 数学生态学引论。卢泽愚译, 1978, 科学出版社。

14. 谭浩强等, 1980: BASIC 语言。科学普及出版社。

15. 许鹏等, 紫泥泉种羊场冬、春、秋草场产量和营养物质动态的初步研究。

16. 汪培庄, 1980: 模糊集理论及其简单应用(北京大学数量生态学讲习班油印讲义)。

17. 吴征隘等, 1980: 中国植被。科学出版社。

18. 阳含熙, 1979: 植物群落数量分类的研究, 林业科学, 15 (4) 244-255。

19. 阳含熙, 1980: 相似系数的探讨, 自然资源, 1980 (1)。



20. 阳含熙, 卢泽愚, 1981: 植物生态学的数量分类方法。科学出版社。

21. 赵松岭等, 1982: 植物群落演替的线性与非线性及数字预测, 生态学报, 1982 (1、3)。

22. 中国科学院蒙宁综合考察队, 1981: 内蒙植被。

И. Х. Блюмянталь и Л. А. Личмелова, 1960.

Геоботанический Очерк Формаций острши ложнопы реинного  
Аггруппум Pseudoаггруппум Геоботаника. Вып. 48. 42 -  
124. Б. А. Выков, 1962: Доминанты раститдльного покрова  
советского союза. Том II изд. Анказ Алма - ата.

## 谁为他插上飞翔的翅膀

一位留美博士希望在西部大开发中发挥一技之长，但回国整整一年过去，却找不到一家接收单位——

在人才缺乏的内蒙古，留美生态学博士白·图格吉扎布（又名白捷）回国已整整一年了，却找不到一家接收单位，让他感到“非常失望”。

去年2月，白捷辞去了美国农业部在科罗拉多州的一个科研机构的工作，决心回国。行前，中国驻芝加哥总领事馆为其出具了证明，并向他介绍：国内正在进行的西部大开发十分重视生态建设，回去会大有用武之地。

对于国内的生态环境问题，白捷并不陌生。早在出国前的1984年，白捷就主持了“草原退化趋势预测及对策研究”的课题。在研究中，他大胆地提出了中国草原正在退化的观点，而且预测到北京大风天数有增加的趋势——到2000年，八级以上大风天数将达66天。课题完成后，他与同事提出了建立草原监测网的初步设想。“图格吉扎布是国内较早搞草原退化研究的人，他当时的一些观点和预测，后来都得到了验证。”中国农科院草原研究所党委书记苏和介绍说。

1985年，白捷离开中国农科院草原研究所（所址在呼和浩特市）到了美国。4年后，他在美国科罗拉多州立大学获得

势分析 及其在生态股市中的应用





数量生态学博士学位。在美国的15年中，他一直没有放弃建立草原监测网的设想，试图找到一个趋势分析、动态监测草原生态系统变化的方法。1995年，他成功研究出“超球面模型”方法。

“建立草原监测网，是我一生最大的愿望。自从研究出‘超球面模型’，我就期待着找一个合适的时间回国，将这个模型用在内蒙古草原生态建设上。国家实施西部大开发战略，我认为是我回国的最好时机。”白捷这样介绍他回国的初衷。

1998年，白捷与中国农科院及草原研究所的领导联系回国事宜，得到很好的答复。“农科院领导还表示，不仅欢迎我回来，而且我太太的工作也可以安排在草原研究所。”

然而，回国前一个月，草原研究所经过多次党政联席会议研究，向白捷发去了一份言辞委婉的电子邮件，答复说：“经所领导研究，意见如下：一、鉴于您目前所从事的关于股票方面的工作以及您的兴趣，目前我们不开展此项业务，这也不是草原研究所的研究重点。二、依据1999年中国农业科学院院工作会议精神，我所实行全员聘任制，对新进我所人员包括长期在国外的留学人员一律以合同制聘任，经双向选择，择优录用……三、就目前我所的情况和您电子邮件中的意见，我所建议您到国家留学生人才交流中心或院属有关单位人事部门咨询有关情况，以得到对您回国的具体落实。”而此时，白捷已辞去了工作，并且拿上了回国的机票。

生态学博士怎么搞起了股票研究？事情原来是这样的。

白捷“超球面模型”的研究成果，实际上是一个解决系统问题的方法和工具，它需要大量的数据作为研究对象和进行实际验证。而在草原生态的研究过程中，得到大量的监测数据相当困难，而且周期很长。一次，中国科学院院士、著名生态学专家阳含熙先生到美国访问，认为白捷的这项研究“有新意”、

“值得继续研究和推广”，但鉴于草原研究的实际情况，建议白捷用股票的数据进行试验，以使模型得到验证。

股市与草原毫无关系，但它们都是动态的多元指数增长系统，经过对数据的处理，可以预测它的发展趋势。

受到启发，白捷用报纸上的股票数据进行试验，发现用模型分析的股票走势，竟然同股票实际走势相一致。为了得到进一步的检验，不懂股票的白捷在1998年6月到2000年1月，分两个阶段依靠模型在股市投资，结果回报率分别达到25.8%和31.99%，而相应的股市平均回报率却是2.64%和9.82%。模型在股市的应用取得了一定的效果。可白捷没有想到，他在股票市场上的试验，恰恰成了拒绝接收他的主要理由。

白捷如期回国。在回国后很长的一段时间内，他没有放弃与草原研究所的接触，答复还是上述三条。直到七八月份，中国农科院的领导明确告诉他：协调失败，草原所拒绝接收他。

草原所回不去，到高校总该可以吧，白捷信心十足地与母校内蒙古大学进行联系。白捷的本科和硕士研究生的学位，都是在内蒙古大学获得的。他还是内蒙古培养的第一批12名理学硕士之一。然而，他的申请仍然遭到了拒绝，理由依然是“专业问题”。

许多专家认为，白捷的研究有一定的“创新”和“实用价值”，值得进一步立项研究。白捷回国时，以草原研究所名义申请了“留学归国人员科研启动基金”。但没有了接收单位，研究经费申请不到，也就无法开展进一步的研究。

在内蒙古的几个月中，白捷还遇到了许多未曾预料到的问题。回国后，白捷的户口问题迟迟得不到解决。“我去派出所，派出所说我的户口底子找不到了；去公安局，公安局说要上缴护照才能落户；去公安厅，公安厅说我有外国居留权，按规定

趋

## 趋势分析 及其在生态股市中的应用



不能落户，必须和上级部门请示才能办理；去外国专家局，专家局说他们只给外国专家发居留证，留学归国人员他们管不着……”白捷说：“单位找不到，我想申请注册一个公司，搞自己的研究，可工商局又提出了种种条件。”万般无奈，白捷给自治区两位领导写信反映了自己的遭遇，以求得帮助。

在自治区领导的过问下，白捷的户口和公司注册问题很快得到了落实，但工作问题一直得不到解决。在西部大开发中，内蒙古虽然制定了一些吸引人才的政策，但对留学归国人员，没有任何针对性的政策。不久前，国家人事部颁布了《关于鼓励海外高层次留学人才回国工作的意见》，其中的规定，也是在有了工作单位后才有可能执行的。

2000年8月，白捷的护照到期了，他不得不返回美国。期间，他的事情受到中共内蒙古自治区党委组织部进一步的关注。有人将这个�息告诉他，建议他再次回国解决工作问题。抱着一线希望，白捷在2000年底又回到内蒙古。这次，一直没有申请成为美国公民的白捷，向当地政府提出了加入美国国籍的申请，并得到批准。“我想，有一个外国专家的身份，回去找工作可能会好一些。”白捷说。

又过去了几个月，白捷的工作单位仍然得不到落实。依靠自己申请的生态数据分析公司，白捷继续做模型应用试验的研究工作。内蒙古自治区政协副主席、内蒙古大学经济与社会发展研究中心主任、原内蒙古大学常务副校长许柏年为他提供了简单的办公场所和一部电话。许柏年认为：“白捷的遭遇虽然是一个个例，但也说明我们在西部大开发的人才引进工作中，在思想上、组织上和条件上存在准备不足的问题。”

几天前，在有关部门的协调下，白捷与内蒙古农业大学进行了接触，但还没有最后的结果。“这将是接触的最后一个单位，我真不愿意这样。”白捷叹口气说。

## Fast Antirandom (FAR) Test Generation To Improve Code Coverage

Andre Bai  
Tom Chen  
Amjad Hajjar

Anneliese von Mayrhauser  
Charles Anderson

Dept. of Electrical Engineering  
Colorado State University  
Fort Collins, CO 80523

Dept. of Computer Science  
Colorado State University  
Fort Collins, CO 80523

970-491-6574

970-491-7016

FAX: 970-491-2249

970-491-7491

andre@engr.colostate.edu

avm@cs.colostate.edu

chen@longs.lance.colostate.edu

anderson@cs.colostate.edu

hajjar@engr.colostate.edu

**Key words:** test data generation, antirandom testing

### 1 Introduction

Testing techniques employ a variety of mechanisms, automated, tool assisted, and manual, for test generation. One of the techniques that has gained support and has shown to be useful in a series of empirical evaluations [5, 6] is anti-random testing. The basic premise of anti-random testing is that in order to achieve higher coverage (of whatever type) one should, after having exercised a set of tests, now choose tests that are as different as possible from the tests previously used. Such a strategy will ensure that tests are not repeated. The distance measure is Hamming and Cartesian distance. New test patterns are chosen that maximize this distance.

In previous analyses, this approach has improved code coverage for boundary conditions, and has proved more efficient than random testing [5, 4].

The basic method has the following two disadvantages.

1. The method essentially requires enumeration of the input space and computation of distance for each potential input vector. This prevents scale-up to large test sets and/or long input vectors. Computations become too expensive.
2. The input vectors on which the anti-random vectors are computed have to be binary. The current way around this problem is to use "checkpoint encoding" [6]. Non-binary inputs are grouped into partitions which are then given a binary encoding. This binary encoding is used for anti-random test generation. The anti-random vectors computed are mapped back into the actual input space by selecting from each of the partitions identified by the binary encoding. Unless the partitions are very small, this approach can lead to the problems identified by [2]. On the other hand, when we have many small partitions, the size of the input vector grows and computation becomes expensive, and quickly impossible.

Our objective was to find a more efficient method to generate anti-random test patterns that would be computationally feasible for large input vectors and long sequences of tests. This would enable a promising technique to be applied to larger problems.



Section 2 illustrates the approach to developing FAR. We start with an informal example to explain the basic approach, i. e., the concept of balancing points using the Multidimensional Sphere Model [1]. This is followed by the explanation of the algorithm and an example of its application. Section 3 analyzes the complexity of FAR compared to the enumerative algorithm and evaluates how close the FAR generated inputs are compared with those by the enumerative method. Section 4 illustrates a use of FAR made possible through its speed of computation. One can group a number of inputs as one packed input and use it to include the effects of dependencies of inputs in the test generation. Section 5 reports on applying FAR to seven different programs, including packing.

We conclude by summarizing our results and pointing out open questions with regards to anti-random testing in general and the FAR method in particular.

## 2 Approach

### 2.1 Method

The anti-random test vector is defined as a vector that has the maximum distance from all previous vectors which were applied during a test [5, 6]. For binary vectors, there are two ways to calculate the distance, Hamming Distance (HD) and the Cartesian Distance (CD). Hamming Distance is defined as the number of corresponding bits that are different between two binary vectors. The Cartesian Distance between two vectors  $A_M$  and  $B_M$  is defined as:

$$CD(A_M, B_M) = \sqrt{(A_M - B_M)^2 + (A_{M-1} - B_{M-1})^2 + \dots + (A_1 - B_1)^2}.$$

For example, for two binary vectors  $A=[1,1,0]$  and  $B=[0,1,1]$ , the Hamming Distance between the vectors is 2 (the first and third bits differ). Their Cartesian Distance is  $\sqrt{(0-1)^2 + (1-1)^2 + (1-0)^2} = \sqrt{2}$ .

Exhaustive anti-random test generation calculates the Hamming Distance (HD) and Cartesian Distance (CD) for all potential input test vectors. The Fast Anti-Random (FAR) approach generates new test sequences by centralizing all existing input test vectors into one test vector. A centroid vector of a set of vectors is their average. Next, FAR finds all potential input test vectors that are orthogonal to the centralized vector. Vectors are orthogonal if their dot product is zero. Finally, FAR finds an anti-random vector with maximum distance from the centroid vector.

Let  $M$  be the number of bits in each input vector.  $M$  is the dimension of the input space. Let  $N$  be the number of such vectors (i. e., the length of the test sequence). The set of all  $M$ -bit binary vectors defines a Euclidean  $M$ -space, and can be denoted by  $2^M$ . The elements of  $2^M$ , points in the  $M$ -space, are called  $M$ -vectors. The FAR algorithm finds a new point in  $M$ -space, provided  $N$  points, to make the  $N+1$  points distributed as evenly as possible in the  $M$ -space. It is interpreted as finding a point with maximum distance from the existing  $N$  points in  $M$ -space[5].

A test sequence of  $N$  vectors is a sampling of the  $M$ -dimensional space with  $2^M$  points in this space [1]. FAR constructs an Anti-random test sequence that balances the points in this space, given the known test sequence of  $N$  vectors.

To illustrate the concept of balanced space, Figure 1 shows a three dimensional binary space with a total of eight possible points in the space. Vectors A, B and D represent three given points in the space. Vector C is the centroid vector of A, B and D. If the centroid vector is rounded to binary integers, then vector C is the same as vector A. The property of the centroid vector is that it has the minimum distance to all the existing vectors. Therefore, one of the orthogonal vectors to the centroid vector will have the maximum distance to all existing vectors. In Figure 1, vector F is the orthogonal vector of the centroid vector and has the maximum distance to vectors A, B and D. Vector F is the anti-random vector. The vectors A, F are symmetrical to each other. So are B and D. This makes the three dimensional space balanced. Table 1 shows the sum of Cartesian distance between all vectors in the space to vectors A, B and D. Vector F has indeed the largest distance to the existing vectors.

More formally, the FAR method consists of the following three steps:

1. Centralization: This step determines the centroid vector of the random test sequence by using the *Multidimensional Sphere Model* [1]. This involves calculating the average of the  $N$  vectors, then rounding the resulting values to 0 or 1. Given  $N$  binary test vectors, first sum the  $N$  vectors, then

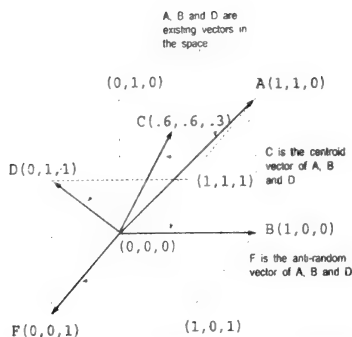


Figure 1: Three dimensional binary vector space

Vectors	Vector A	Vector B	Vector D	CD
(0,0,0)	$\sqrt{2}$	1	$\sqrt{2}$	$1+\sqrt{2}+\sqrt{2}$
(0,1,0)	1	$\sqrt{2}$	1	$1+1+\sqrt{2}$
(1,0,1)	$\sqrt{2}$	1	$\sqrt{2}$	$1+\sqrt{2}+\sqrt{2}$
(1,1,1)	1	$\sqrt{2}$	1	$1+\sqrt{2}+\sqrt{2}$
F(0,0,1)	$\sqrt{3}$	$\sqrt{2}$	1	$1+\sqrt{2}+\sqrt{3}$

Table 1: The Cartesian distance between vectors

divide the vector sum by  $N$ . The result is the centroid vector, a vector whose elements are between 0 and 1. Because we need a binary vector, we project these floating-point numbers into binary numbers by rounding either down to 0 or up to 1. This projection between a floating point number and a binary one will introduce truncation error. This is shown in Figure 1 as the change of from the centroid vector C to the binary centroid vector A.

FAR rounds vector elements to 0 for values less than 0.5. if they are greater it rounds them to 1. If the value equals 0.5, FAR randomly selects either 0 or 1. This creates a binary centroid vector.

2. Orthogonal Vector Calculation: This step finds all orthogonal vectors to the centroid vector. The vector farthest from the centroid vector will be an orthogonal vector, making orthogonality a necessary condition to being farthest away. Two vectors are orthogonal, if and only if their dot product is zero[7]. Orthogonal vectors include those with maximum distance provided all vectors are nonnegative. Not all orthogonal vectors will have maximum distance to the centroid vector, but the vector with the maximum distance is orthogonal to the centroid.
3. Maximum Distance Calculation: Since all new test vectors generated are orthogonal to the given centroid vector, the orthogonal vector with the maximum vector length will have the maximum distance. For binary vectors this is the vector containing the most 1's.

For implementation, steps 2 and 3 can be combined, as inverting the binary centroid will result in an orthogonal vector with maximum distance.

During the process of centralization, a centroid vector bit might be 0.5. This means that the space is balanced with respect to that particular vector element. When a vector element in the centroid vector becomes 0.5, FAR randomly selects it to be 0 or 1. When generating an anti-random test sequence based



on the centroid, FAR iterates until all vector elements in the centroid vector become 0.5, implying the  $M$ -space is balanced. While the FAR algorithm is an approximation, it is fairly effective until the centroid vector has too many entries of 0.5 (the space is balanced.) Thus we would expect to use the FAR algorithm in cases where enumeration is not possible and the input space is large. The concept of a balanced space provides a natural stopping criterion for this technique.

## 2.2 Example

Let us now illustrate with an example how to generate a six bit anti-random vector when given five input test vectors. Assume a six dimensional input space ( $M = 6$ ) with five input test vectors ( $N = 5$ ). The existing test vectors are given in Table 2. We now use FAR to generate anti-random vectors to these test vectors.

The Test Vectors	
A=	[1,1,1,0,1,0]
B=	[0,1,0,1,1,0]
C=	[1,0,0,0,1,1]
D=	[0,1,1,0,0,0]
E=	[1,0,0,0,1]

Table 2: The existing test sequence

**Step 1: Centralization:** Find the centroid vector  $X$  of the five given test vectors. Vector  $X$  is the average of vectors A through E. Vector  $X'$  is the rounded binary centroid vector. The computation of vectors  $X$  and  $X'$  is shown in Table 3.

Centroid Vector X Calculation	
$X = (A+B+C+D+E)/5$	
$= [3, 3, 2, 1, 3, 2]/5$	
$= [0.6, 0.6, 0.4, 0.2, 0.6, 0.4]$	
Corresponding Binary Centroid $X'$	
$X' = [1, 1, 0, 0, 1, 0]$	

Table 3: The centralization procedure and the binary centroid  $X'$

**Step 2: Orthogonal Vector Calculation:** Find all vectors in  $M$ -space that are orthogonal to the centroid vector  $X'$ . The list of all the orthogonal vectors is shown in Table 4.

**Step 3: Maximum Distance Calculation:** Next, FAR finds the orthogonal vector that has maximum vector length, i. e., the binary vector containing the most 1's. For this example,  $X_1 = [0, 0, 1, 1, 0, 1]$  contains the most 1's, therefore, it is the anti-random vector for  $X'$ .

## 3 Evaluation of FAR

### 3.1 Computational Complexity of FAR

We compare the computational complexity between the exhaustive search algorithm and the search algorithm proposed here in FAR. The exhaustive anti-random test generation is completed by first obtaining a binary test sequence. Next, find a new test vector for which the sum of Hamming distance and Cartesian distance from all previous vectors is greatest. However, calculating the total Hamming distance and total Cartesian distance generally requires enumeration of the input space and distance computation for each

Orthogonal Vectors
X1=[0,0,1,1,0,1]
X2=[0,0,0,1,0,1]
X3=[0,0,1,0,0,1]
X4=[0,0,1,1,0,0]
X5=[0,0,0,0,0,1]
X6=[0,0,1,0,0,0]
X7=[0,0,0,1,0,0]
X8=[0,0,0,0,0,0]

Table 4: All vectors that are orthogonal to X'

potential input vector. Since there are  $2^M-N$  potential input vectors, ( $M$  being the number of bits in a binary vector, and  $N$  being the length of the test sequence given), the complexity of exhaustive calculation then becomes of the order  $(2^M-N)*N$ .

Conversely, the FAR algorithm is based on the calculation of the centroid vector. Given  $N$  test inputs with  $M$  bits, the FAR algorithm treats the given test sequence as a two dimensional array of size  $N*M$ . It first sums each column of the two dimensional array and then divides it by the number of rows to obtain a centroid vector. It then rounds each bit of the centroid vector to 0 or 1. A new anti-random vector is constructed by inverting each bit of the centroid vector. Therefore, the FAR algorithm only requires on the order of  $N*M$  calculations. Thirty sets of anti-random generations were run to illustrate the complexity of the FAR algorithm, as compared to an exhaustive search algorithm. In each run we generated 200 binary vectors, each with a bit length of 15. We used the FAR algorithm to find one anti-random vector based on 200 test vectors and then compared the result with an anti-random vector generated by the exhaustive search algorithm. In the experiment,  $M$  is 15, and  $N$  is 200, then the complexity of the exhaustive search algorithm becomes on the order of  $(2^{15}-200)*200$ , or  $6.5*10^6$ . By comparison, the complexity of FAR is only on the order of 3000. Therefore, we would expect that in this study, the cost for generating one anti-random vector using the exhaustive search algorithm will be 3 degrees of magnitude higher than the anti-random vector generated by FAR. Figure 2 shows the cost of finding one anti-random test vector using the FAR algorithm versus the exhaustive search algorithm, as measured in CPU time. The X axis is the number of experiments and the Y axis is the CPU time.

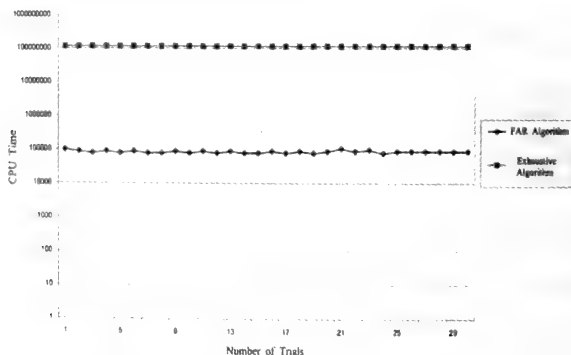


Figure 2: The CPU time of FAR algorithm vs. exhaustive algorithm





As expected, the results show that the cost for generating one anti-random vector using the exhaustive search algorithm is indeed 3 orders of magnitude higher than the FAR algorithm.

### 3.2 Quality of FAR

Because FAR uses an approximation of the centroid vector by rounding bits to 0 or 1, a rounding error will occur when given an even number of input test vectors. On the other hand, the cost for generating the anti-random test sequence using the FAR algorithm is low compared to the exhaustive search algorithm. Thus we are interested in examining the quality of the anti-random vectors generated by the FAR algorithm. The quality of the anti-random vector generated by FAR is measured by the distance between the vector generated by FAR and that generated by the exhaustive search algorithm from the same set of existing test inputs. The same data described in section 3.1. was used to measure the quality of the anti-random vectors generated by FAR. Figure 3 shows the differences in the anti-random vectors generated by FAR versus using the exhaustive search algorithm. Distance is measured in terms of the Hamming distance of the two anti-random vectors. The X axis identifies each run in the sequence, while the Y axis shows the Hamming distance between the anti-random vector generated by the FAR algorithm versus the exhaustive search algorithm for a given run.

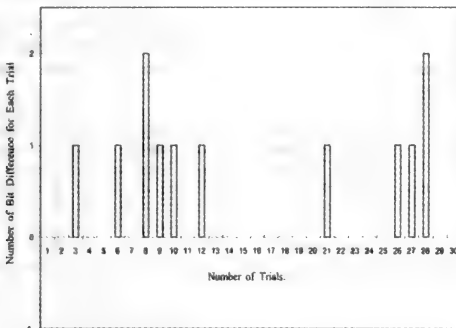


Figure 3: Quality of FAR algorithm vs. exhaustive algorithm

The results show that for the majority of experiments, the Hamming distance between the anti-random vector generated by the FAR algorithm and the exhaustive search algorithm is zero. There are only two cases where Hamming distance between the anti-random vector generated by FAR versus the exhaustive search algorithm is more than 1. Given that the FAR algorithm is three degrees of magnitude faster than exhaustive search, we conclude that the overall performance of the FAR algorithm by far exceeds that of the exhaustive search algorithm.

## 4 Input Packing

In testing we often have to deal with the issue of retained state [3], that is, the same test input pattern can generate different execution behavior depending on its order in the sequence. When testing for coverage, this can make a huge difference in what parts of the code get executed. In this case we are interested in generating sequences of tests, rather than single, independent input patterns. Therefore, the key to successful testing is generating appropriate sequences. Usually, even a finite window of sequences is better than a single, independent generation. Because of the speed of the FAR generation method, we can generate test sequences of desired length through packing. *Packing of size k* is defined as concatenating k input patterns of size M into one new input of size kM. When packing, we are modeling sequences of test inputs,

of length  $k$ . A packed input with packing size  $k$  is called a *scenario segment*. For packed input, FAR generates new scenario segments that are as far away from the current usage pattern as possible.

Packing can be expected to be beneficial in situations where code execution depends heavily on the order in which a series of inputs is applied in addition to the value of the individual input. Optimal packing size is problem dependent and currently must be determined by the tester.

Most programs will have state dependent aspects. In situations with small input space (for example, a Tic-Tac-Toe game), this can lead to exhaustive testing of all inputs without full code coverage. Applying packing increases the input space of low dimensional input vectors and alleviates this situation.

## 5 Applications

We have applied the FAR technique to test code written in VHDL, a hardware design language. Each of the VHDL programs represents a model of a hardware design. We selected this language for our empirical investigation, because its input variables (*signals*) are binary vectors. Using the FAR technique we generated anti-random test sequences with vectors of varying size ( $M$  varied from 2 to 62). For the larger dimensions, enumerative computation is clearly out of the question. The FAR algorithm quickly handled even the larger dimensions.

Using seven different VHDL models, we performed an empirical study which generated a series of random test cases first, then created anti-random test sequences for these tests. Table 5 lists for each model its name, the number of lines of VHDL code, the number of branches, the number of control bits and the number of data bits.

Name	LOC	Branches	Control	Data
b01	115	28	1	2
b03	141	27	1	4
b04	80	17	3	9
b08	130	33	2	8
b10	167	44	3	8
b13	296	79	3	8
sys7	3785	591	7	62

Table 5: Benchmark characteristics

For each model, we compared anti-random testing versus random testing. First, we applied random tests to each model followed by an anti-random test. We compared the branch coverage increase of the anti-random test against cases where a random test was followed by another random test with the same number of inputs as for the anti-random strategy, so that the total number of patterns for both setups are the same. Table 6 shows the branch coverage after the first phase (random test), the branch coverage after the second phase (random test), and the percentage increase after the second phase. Table 7 shows the branch coverage after the first phase (random test), the branch coverage after the second phase (anti-random test), and the percentage increase after the second phase.

Figure 4a shows the graph for a branch coverage increase from applying a set of random test sequences followed by another set of random test sequences. Figure 4b shows the branch coverage increase when applying the same set of random test sequences followed by a set of anti-random test sequences. The X axis of Figure 4 is branch coverage percent for seven different models at the end of phase 1. The Y axis of Figure 4 is the percentage increase after the second phase of the test.

Figure 5 is the curve fitted graph of the comparison of Figure 4a and Figure 4b. The solid line is the case where a random test sequence was followed by more random test inputs. The dashed line represents the case where a random test sequence was followed by anti-random test inputs. The X axis of Figure 5 is the branch coverage in percent for the seven models after executing the random patterns of phase 1. The Y axis is the percent increase in branch coverage after the second phase of the test. We used this format to



Name	b01	b03	b04	b08	b10	b13	sys7
First Phase (Random Test)	89.29	77.78	70.59	72.73	70.45	69.62	57.53
Second Phase (Random Test)	89.29	77.78	70.59	72.73	72.73	69.62	70.05
Branch Coverage Increase	0	0	0	0	2.28	0	12.52

Table 6: Random test followed by random test

Name	b01	b03	b04	b08	b10	b13	sys7
First Phase (Random Test)	89.29	77.78	70.59	72.73	70.45	69.62	57.53
Second Phase (Anti-random Test)	92.86	77.78	76.47	72.73	70.45	69.62	69.71
Branch Coverage Increase	3.57	0	5.88	0	0	0	12.18

Table 7: Random test followed by anti-random test

visualize the effect of the coverage level at the beginning of phase 2 on possible further coverage. Usually, it is harder to achieve additional coverage once a certain coverage level is reached.

The results show that using anti-random tests after a set of random test sequences tended to improve branch coverage, as opposed to using further random test sequences. It is not, however, a strong result in favor of anti-random testing. The amount of coverage increase for anti-random versus random testing in the second phase was more for two models, less for two models, and was the same for three models. This could be due to the issue of retained state, or ordering of inputs, particularly since the VHDL models that show the least improvement have state-dependent code.

To investigate this further, we used the sys7 model to evaluate the effectiveness of packing. The sys7 model is a highly state sensitive model where the order of the test patterns executed will have a significant impact on branch coverage increase.

The FAR algorithm generates an anti-random test sequence very quickly even with a big  $M$ -space. This allowed us to use packing to improve the branch coverage of the sys7 model. In the first phase of the experiment we treated the model with 200 random test sequences, and then generated 25 anti-random test sequences using packing sizes  $k$  of 1, 5, 10, 15, 20, 25 and 30 of the original input vectors.

In the second phase of the study, we treated the sys7 model with 25 anti-random test sequences that were generated and measured cumulative branch coverage for each packing size. We repeated this 19 times with different seeds for generating the random test vectors. In the same set up, we also tested the case where 225 random tests were applied to the model, and measured the branch coverage for each test. In this way we wanted to average the effect of the random seed used for the random test generation. The average branch coverage of the 19 runs for each packing size compared to the 225 random tests is shown in Table 8. Anti-random testing improved branch coverage in all cases over pure random testing. The amount of improvement varied with the packing size.

Figure 6 graphs the average branch coverage in percent for each packing size for the sys7 model. The X axis is the packing size. The Y axis is the branch coverage in percent. From the figure, we can clearly tell that the order of the test inputs in the sequence of execution had an impact on branch coverage. The two best packing sizes for the sys7 program are either relatively small ( $k=5$ ) or large ( $k=30$ ). We believe packing size is problem and code structure dependent.

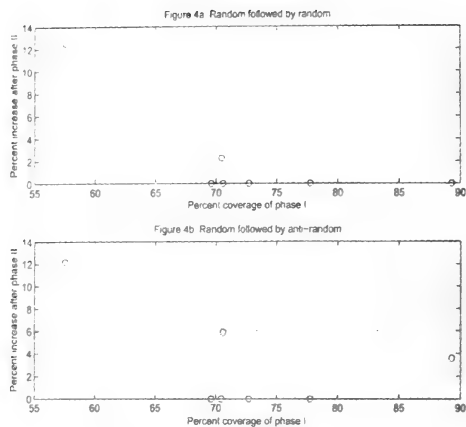


Figure 4: The graph of branch coverage increase

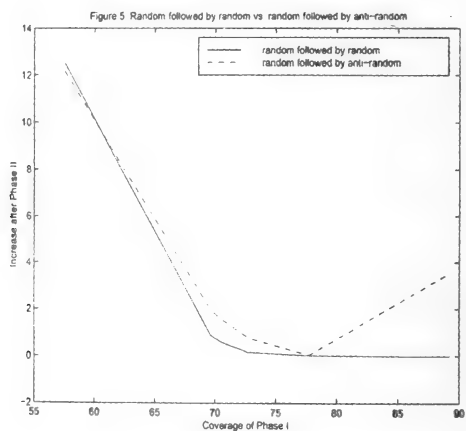


Figure 5: The curve fitted graph of Figure 4

## 6 Conclusion

We demonstrated a Fast Anti-Random Test generation algorithm that is able to improve test coverage of branches. It can be used to generate single, independent input patterns, or sequences of related inputs via packing. This paper shows its use on VHDL as the application language. When applying it to programs with more sophisticated data types and inputs, a mapping or encoding can be used to map the binary vectors into the representation required for the actual input to the program. This mapping could potentially be





Packing Size	225 random	1	5	10	15	20	25	30
Average Branch Cov.	67.78	67.97	68.52	67.85	67.86	67.86	68.06	68.19

Table 8: Average branch coverage for each packing size

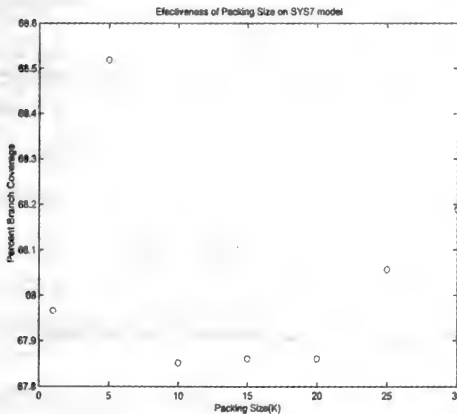


Figure 6: Effectiveness of packing size on sys7 model

checkpoint encoding [6].

While our small study was not able to achieve large improvements in branch coverage, as for example reported in [5, 6], several things must be considered.

- Based on work by Hamlet, et al. [2], and Tsoukalas, et al. [8] we cannot expect any single testing technique to be consistently superior. Thus anti-random testing will not always be the "better" technique. When it is advisable, it is of course important to have a fast implementation like FAR.
- Packing improved FAR's performance in uncovering new branches. The best packing size depends on the code under test (its state transition behavior).
- Even if only moderate increases in branch coverage are possible, this should be compared to how many more random tests one would have to execute to achieve the same coverage as via anti-random testing. Such a comparison would indicate the amount savings in test execution and validation.

Ultimately, one would want to combine a testing technique like FAR with a stopping rule (e. g., [3]), so that it will be used only as long as it yields results. Secondly, each testing technique should come with rules of thumb when their use is likely to be advantageous. For anti-random testing we are working on both of these issues.

## Acknowledgements

This work was supported by the National Science Foundation through grant MIP-9628770. Yashwant Malaiya of Colorado State University contributed to our understanding of anti-random test-generation

**Colorado  
State  
University**

Center for Ecological Management  
of Military Lands

Department of Forest Sciences  
Fort Collins, Colorado 80523  
USA

Telephone (303) 481-2748  
FAX (303) 481-2713

April 5, 1994

MEMORANDUM

TO: Jay Bai, Dave Anderson, Bob Brozka, Dave Kowalski

FROM: Mary Huwa

SUBJECT: Ownership of Model Developed by Jay Bai

Per discussions at our meeting on March 18, 1994, concerning ownership of the computer model developed by Jay Bai, I contacted Legal Counsel at Colorado State University for an opinion. I was referred to Kathleen Byington, President, Colorado State University Research Foundation (CSURF). CSURF provides copyright and patent policy expertise for CSU. I spoke with Ms. Byington about the issue and was asked to submit a one-page summary of the situation. She then conferred with Jim Brown, Assistant Vice President for Research/Patent Officer, Colorado State University. Their opinion is that ownership of the model belongs with Inner Mongolia University.

Any communications referencing the model should credit Inner Mongolia University. Any distribution would be subject to their approval.

cc: Bob Shaw  
Gary Belew

趋势分析 及其在生态股市中的应用

**CERTIFICATE OF REGISTRATION**

**FORM TX**

For a Literary Work



This Certificate issued under the seal of the Copyright Office in accordance with title 17, United States Code, attests that registration has been made for the work identified below. The information on this certificate has been made a part of the Copyright Office records.

TXU 714-554



*Margareth Peters*

EFFECTIVE DATE OF REGISTRATION  
MAY 24 1995

REGISTER OF COPYRIGHTS  
United States of America

DO NOT WRITE ABOVE THIS LINE. IF YOU NEED MORE SPACE, USE A SEPARATE CONTINUATION SHEET.

附件 4

**1** TITLE OF THIS WORK ▼  
STOCK MARKET MONITOR 52 (SMM 52)  
PREVIOUS OR ALTERNATIVE TITLES ▼ SMM 52

PUBLICATION AS A CONTRIBUTION ▼ If this work was published as a contribution to a periodical, serial, or collection, give information about the collective work in which the contribution appeared. Title of Collective Work ▼

If published in a periodical or serial give: Volume ▼ Number ▼ Issue Date ▼ On Pages ▼

**2 a** NAME OF AUTHOR ▼ TUGJAYZHB JAY BAI DATES OF BIRTH AND DEATH  
Year Born ▼ 1945 Year Died ▼

Was this contribution to the work a "work made for hire"?  Yes  No  
AUTHOR'S NATIONALITY OR DOMICILE  
Name of Country: CHINA  
OR  
Citizen of: CHINA  
OR  
Domiciled in: CHINA

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK  
Anonymous?  Yes  No  
If the answer to either of these questions is "Yes" see detailed instructions

NATURE OF AUTHORSHIP ▼ Briefly describe nature of material created by this author in which copyright is claimed.  
COMPUTER PROGRAM

**2 b** NAME OF AUTHOR ▼ ANGGIN BAI DATES OF BIRTH AND DEATH  
Year Born ▼ 1975 Year Died ▼

Was this contribution to the work a "work made for hire"?  Yes  No  
AUTHOR'S NATIONALITY OR DOMICILE  
Name of Country: CHINA  
OR  
Citizen of: CHINA  
OR  
Domiciled in: CHINA

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK  
Anonymous?  Yes  No  
If the answer to either of these questions is "Yes" see detailed instructions

NATURE OF AUTHORSHIP ▼ Briefly describe nature of material created by this author in which copyright is claimed.  
PROGRAM CODE

**3 a** YEAR IN WHICH CREATION OF THIS WORK WAS COMPLETED 1995 **b** DATE AND NATION OF FIRST PUBLICATION OF THIS PARTICULAR WORK  
Complete this information only if a new work has been published. Month ▼ Day ▼ Year ▼

Was this contribution to the work a "work made for hire"?  Yes  No  
AUTHOR'S NATIONALITY OR DOMICILE  
Name of Country: CHINA  
OR  
Citizen of: CHINA  
OR  
Domiciled in: CHINA

WAS THIS AUTHOR'S CONTRIBUTION TO THE WORK  
Anonymous?  Yes  No  
If the answer to either of these questions is "Yes" see detailed instructions

NATURE OF AUTHORSHIP ▼ Briefly describe nature of material created by this author in which copyright is claimed.

**4** COPYRIGHT CLAIMANT(S) (SI) Name and address must be given even if the claimant is the same as the author given in space 2. TUGJAYZHB JAY and ANGIN BAI  
615 JOANNE ST.  
FORT COLLINS, CO 80524

APPLICATION RECEIVED  
**MAY 24 1995**  
ONE DEPOSIT RECEIVED  
  
TWO DEPOSITS RECEIVED  
**MAY 24 1995**  
REMITTANCE NUMBER AND DATE

TRANSFER If the claimant(s) named here in space 4 is/are different from the author(s) named in space 2, give a brief statement of how the claimant(s) obtained ownership of the copyright.

EXAMINED BY \_\_\_\_\_ FORM TX

CHECKED BY *ll*

CORRESPONDENCE  
 Yes

FOR  
COPYRIGHT  
OFF  
UL  
ONE

DO NOT WRITE ABOVE THIS LINE. IF YOU NEED MORE SPACE, USE A SEPARATE CONTINUATION SHEET.

PREVIOUS REGISTRATION Has registration for this work, or for an earlier version of this work, already been made in the Copyright Office?

- Yes  No If your answer is "Yes," why is another registration being sought? (Check appropriate box) ▼
- a.  This is the first published edition of a work previously registered in unpublished form.
- b.  This is the first application submitted by the author or copyright claimant.
- c.  This is a changed version of the work, as shown by space 6 on this application.

5

If your answer is "Yes," give Previous Registration Number ▼ Year of Registration ▼

DERIVATIVE WORK OR COMPILATION Complete both space 6a and 6b for a derivative work; complete only 6b for a compilation.

a. Preexisting Material Identify any preexisting work or works that this work is based on or incorporates. ▼

b. Material Added to This Work Give a brief, general statement of the material that has been added to this work and in which copyright is claimed. ▼

6

See instructions before completing this space

—space deleted—

7

REPRODUCTION FOR USE OF BLIND OR PHYSICALLY HANDICAPPED INDIVIDUALS A signature on this form at space 10 and a check in one of the boxes here in space 8 constitutes a non-exclusive grant of permission to the Library of Congress to reproduce and distribute solely for the blind and physically handicapped and under the conditions and limitations prescribed by the regulations of the Copyright Office: (1) copies of the work identified in space 1 of this application in Braille for similar tactile symbols; or (2) photocopies embodying a fraction of a reading of that work; or (3) both.

Copies and Photocopies

Copies Only

Photocopies Only

See instructions

8

DEPOSIT ACCOUNT If the registration fee is to be charged to a Deposit Account established in the Copyright Office, give name and number of Account.

Name ▼ Account Number ▼

9

CORRESPONDENCE Give name and address to which correspondence about this application should be sent. Name/Address/Apt./City/State/ZIP ▼

T. JAY BAI  
615 JOANNE ST.  
FORT COLLINS, CO 80524

Area Code and Telephone Number ▶ H-970-495-9716 W-719-526-1693 (800-478-8171)

Be sure to give your telephone please

CERTIFICATION I, the undersigned, hereby certify that I am the

Check only one →

- author
- other copyright claimant
- owner of exclusive rights
- authorized agent of

of the work identified in this application and that the statements made by me in this application are correct to the best of my knowledge.

Name of author or other copyright claimant, or owner of exclusive rights ▼

10

Typed or printed name and date ▼ If this application gives a date of publication in space 3, do not sign and submit it before that date.

TUGJAYZAB JAY BAI, ANGLIN BAI

date ▶ 5-20-95

Handwritten signature (Do not write in this space) *Tugjayzab Bai Anglin Bai*

MAIL CERTIFICATE TO

Certificate will be mailed in window envelope

Name ▼  
T. JAY BAI  
Number/Street/Apartment Number ▼  
615 JOANNE ST.  
City/State/ZIP ▼  
FORT COLLINS, CO 80524

Complete all necessary spaces Sign your application in space 10

1. Application form  
2. Nonreturnable \$20 filing fee in check or money order payable to Registrar of Copyrights U.S. Deposit Receipt

Registrar of Copyrights Library of Congress Washington, D.C. 20540

11

The Copyright fee has been paid and the fee has been credited to the Copyright Fee Index. The next edition must be due to the Copyright Office after July 1 (to determine the applicable rate).

17 U.S.C. § 101(d). Any person who knowingly makes a false representation of a material fact in the application for copyright registration provided for by Section 409 or in any written statement filed in connection with the application, shall be liable not more than \$2,500.





UNITED STATES  
**SECURITIES AND EXCHANGE COMMISSION**  
CENTRAL REGIONAL OFFICE  
1801 CALIFORNIA STREET  
SUITE 1500  
DENVER, COLORADO 80202-2656

August 30, 2005

VIA CERTIFIED MAIL

Dr. Tugayzhab J. Bai  
615 Joanne St.  
Ft. Collins, CO 80524-3684

Re: MDSM Excelerator Hedge Group and MDSM Excelerator Benchmark Group

Dear Dr. Bai:

It has come to our attention that you were involved in the above-mentioned private offerings. We would like to meet with you to discuss these offerings. You should be aware that you are not required to participate in any such meeting or provide any information to the staff of the Commission. However, if you choose to voluntarily meet with the staff, please contact Paula Weisz at (303) 844-1088 to arrange a convenient time to meet in our office at 1801 California Street, Suite 1500, Denver. We also ask that you bring the following records to the meeting.

1. Copies of all offering documents and any other advertisements or promotional, pamphlets, or materials regarding MDSM Excelerator Hedge Group or the MDSM Excelerator Benchmark Group distributed to more than one person.
2. Records supporting MDSM's performance as referenced in the brochure entitled, "The New Frontier of Investing."
3. A list of current and past investors in the two MDSM offerings together with addresses, telephone numbers, and amounts invested.

If you have any questions, please call Paula Weisz or Bill Johnson at (303) 844-1000

Sincerely,

Dale E. Coffin, Assistant Regional Director  
Investment Company/Investment Adviser Regulation

Enclosure: SEC 1662

中科院植物所图书馆



S0048463

2006.10.24

赠送

17.80元















# 内 容 提 要

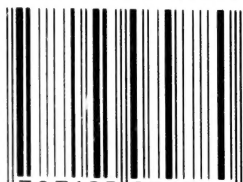
这是一本由研发者的科研论文、科研通信、应用实例组成的，从不同视角介绍超球面模型的论文集。超球面模型是为草原监测研发的一个新的多元分析工具。在它的研发初期，就指出了中国草原的退化趋势和北京地区大风日数增加的趋势。后来应用到美国土地条件趋势分析（LCTA）课题、模型的植被分类结果和土地条件趋势分析报告等，其结果与土地管理员的经验相吻合。为了进一步验证完善数学模型，模型被转用到股市数据分析，并直接用来指导个人退休金的投资管理。六年投资的年平均回报率达到18%，远远超过同期的道·琼斯、纳斯达克和标准普尔五百这三个著名的指数。材料引起美国联邦政府证券会中央地区办公室的注意和调查。他们审查的结论是，超球面公司可以用这一实验结果为数据分析工作做广告。研发者将视线转回到中国草原生态研究之后，认为中国沙尘暴的起因有深刻的文化层面的原因，是农耕文化排斥游牧文化的结果，而直接的原因是草原植被退化。只有旱生草本植被的庞大根系才能网住表层土壤，相对而言，树木的生物量却主要分布在空中和地下。研发者建议，厉行《十一律》，保护恢复草原植被才是保证中国天蓝、地绿、水清的根本。

责任编辑：松布尔

责任校对：乌日娜

封面设计：周 方

ISBN 7-105-07978-9



9 787105 079780 >

ISBN 7-105-07978-9/N · 81

(汉10)

定价：17.80元

