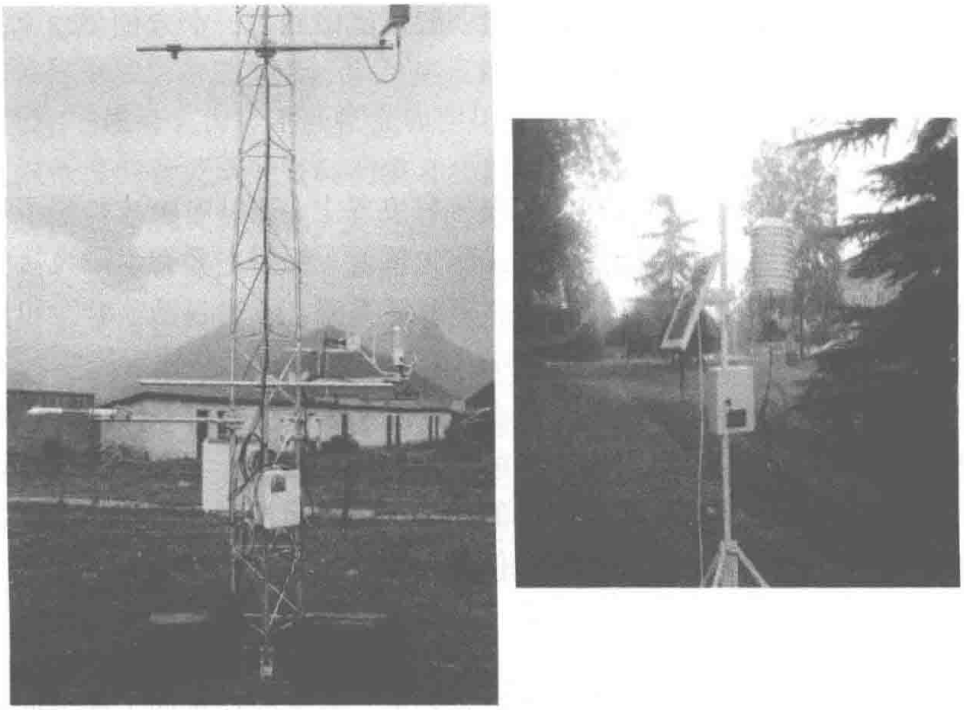


图 2-2　温度、降水、风速等多参数自动气象观测系统

该类设备的主要功能及特点有：①采集系统与计算机连接后，可实现区域性气象数据的整点自动采集、处理和储存，具有数据屏幕显示，同时可将小气候数据导出到 Excel 进行编辑和分析，按需要生成图表；②自动采集各类参数数据，通过接口可以将数据通过软件下载到计算机中，在主机大屏幕中文显示，可实时显示采样日期和时间（年、月、日、时、分、秒）、组数、温度、湿度、光照强度、风速风向、降雨量等参数，仪器小巧美观，应用方便；③可增加和更改测量参数组合，如光合有效辐射、风向风速、雨量、土壤水势、土壤 pH、GPS 定位信息等。

（三）历史气象数据的获取

为了观测气象资料，世界各国都建立了各类气象观测站，如地面站、探空站、测风站、火箭站、辐射站、农气站和自动气象站等。我国自新中国成立以来，已建成了类型齐全、分布广泛的台站网，台站总数达到 2000 多个。

如果需要研究较大空间范围及长时间序列的气候变化动态，如气候与自然植被的时空分布格局及其互作机制、潜在植被的演替动态等，就需要收集整理已有的气象观测数据和气候预测数据。

1. 气象观测数据

区域及地区空间尺度的气象数据，可以通过相关的机构获取。例如，中国历史气象数据，可通过中国气象数据网（http://data.cma.cn/）下载 1981～2010 年全国 824 个基本及基准气象站点的逐日气象观测数据。全球气象数据，可通过联合国世界气象组织（World Meteorological Organization，https://www.wmo.int/pages/index_zh.html）下属的世界气象数据中心（the WDC for Meteorology，http://www.ncdc.noaa.gov/wdc/），免费获取

全球 1948 年 5 月至今各月的温度、降水等数据（https://www.ncdc.noaa.gov/）。该数据共享系统由美国商务部（the U.S. Department of Commerce）和国家海洋与大气管理局（National Oceanic and Atmospheric Administration，NOAA）维护，由美国国家环境信息中心管理运作。也可通过全球气候数据网站（http://www.worldclim.org）下载全球空间分辨率为 30s（约 1km）的近 50 年（1950～2000 年）的各月降水量及温度等数据。该数据库主要由美国加利福尼亚大学伯克利分校（University of California at Berkeley）建立和维护，主要由 Hijmans 等（2005）根据 1950～2000 年全球气象台站观测的数据，采用薄板平滑样条插值方法建立空间数据库，通过多种插值技术的比较检验，证明这些数据库具有良好的拟合精度，已广泛用于全球气候变化的研究中。

2. 气候预测数据

通过气候变化与农业及食品安全网站（Climate Change Agriculture and Food Security Website）（http://www.ccafs-climate.org/data/），可以下载全球气候预测数据库。该数据集包括加拿大气候模拟与分析中心（Canadian Centre for Climate Modelling and Analysis，CCCMA）、澳大利亚联邦科学与工业研究组织（Commonwealth Scientific and Industrial Research Organisation，CSIRO）等多个机构的全球气候模型（Global Climate Model，GCM）和多种排放情景（Emission Scenarios，如 A2a 和 B2a 等）。联合国政府间气候变化专门委员会（Intergovernmental Panel on Climate Change，IPCC）发布的未来气候预测报告中使用了该数据库（IPCC，2007）。

三、土壤数据

土壤（Soil）是人类赖以生存和发展的物质基础，是陆地生态系统的核心组成部分。为了在全球尺度、国家尺度和区域尺度上解决资源、环境和生态等问题，采集土壤信息是草业科学研究必要的内容之一。土壤是草地生态系统的基础条件，土壤变化与草地变化关系十分密切，两者都受到自然因素与人类生产活动的影响，人们所熟知的草原上 3 个典型的土壤类型的形成、分布与生物气候带是相适应的，黑钙土是在温带半湿润草甸化草原下形成的，干旱的典型草原以栗钙土为主，而棕钙土则是荒漠草原环境的产物（李绍良等，2002）。因此，草地与土壤是相互作用相互影响的，二者之间一方的改变，都会引起另一方的变化，土壤数据的采集对开展草地生态环境评价、草地退化评价、草地生态风险预警等方面的研究有着非常重要的作用。

我国土壤资源丰富、类型繁多。土壤分类不仅是土壤地理学的重要理论基础，还是土壤科学发展水平的标志。土壤分类可以为土壤调查制图和草地资源评价及合理利用、改良土壤提供科学依据。目前我国土壤系统分类按照土纲、亚纲、土类、亚类、土科和土系 6 级分类单元分类。其中最高一级共分为 14 个土纲，39 个亚纲，138 个土类和 588 个亚类（龚子同和张甘霖，2006）。全国主要土壤类型有砖红壤、赤红壤、红壤和黄壤、黄棕壤、棕壤、暗棕壤、褐土、黑钙土、栗钙土、棕钙土、黑垆土、荒漠土、高山草甸土和高山漠土。

（一）土壤数据采集指标

主要包括土壤养分、土壤水分和土壤温度等指标。

1. 土壤养分

土壤养分是指由土壤提供的植物生长所必需的营养元素，能被植物直接或转化后吸收。土壤养分可大致分为大量元素、中量元素和微量元素，包括氮（N）、磷（P）、钾（K）、钙（Ca）、镁（Mg）、硫（S）、铁（Fe）、硼（B）、钼（Mo）、锌（Zn）、锰（Mn）、铜（Cu）和氯（Cl）等元素。在自然土壤中，土壤养分主要来源于土壤矿物质和土壤有机质，其次是大气降水、坡渗水和地下水。在耕作土壤中，还来源于施肥和灌溉。

根据在土壤中存在的化学形态，土壤养分的形态可以分为：①水溶态养分，土壤溶液中溶解的离子和少量的低分子有机化合物；②代换态养分，是水溶态养分的来源之一；③矿物态养分，大多数是难溶性养分，有少量是弱酸溶性的（对植物有效）；④有机态养分，矿质化过程的难易程度不同。根据植物对营养元素吸收利用的难易程度，土壤养分又分为速效性养分和迟效性养分。土壤养分的总贮量中，有很小一部分能为植物根系迅速吸收同化的养分称为速效性养分。一般来说，速效性养分仅占很少部分，不足全量的 1%。除速效性养分外，土壤中其余绝大部分养分必须经过生物的或化学的转化作用才能为植物所吸收，这部分养分称为迟效性养分。

2. 土壤水分

土壤水分是植物吸收水分的主要来源（水培植物除外），另外植物也可以直接吸收少量落在叶片上的水分。土壤水分主要来源于大气降水和灌溉水，此外，地下水上升和大气中水汽的凝结也是土壤水分的来源。水分由于在土壤中受到重力、毛管引力、水分子引力、土粒表面分子引力等各种力的作用，形成不同类型的水分并反映出不同的性质。其中，固态水为土壤水冻结时形成的冰晶，气态水主要存在于土壤和空气中，束缚水包括吸湿水和膜状水，自由水包括毛管水、重力水和地下水。

3. 土壤温度

土壤温度指地面以下与植物生长发育直接有关的浅层土壤的温度。土壤温度影响植物的生长、发育和土壤的形成。土壤中各种生物化学过程，如微生物活动所引起的生物化学过程和非生命的化学过程，都受土壤温度的影响。草地植被生长、发育同样受到土壤温度的影响。土壤温度在一定的范围内，直接影响农牧业生产，土壤温度越高，作物的生长发育越快。与气温相比，对种子发芽和出苗的影响，土壤温度要直接得多。但是，土壤温度随地形、土壤水分、耕作条件、天气影响而变化。适宜的土壤温度能促进作物的营养生长和生殖生长。土壤温度间接影响环境条件中的其他因子，从而影响作物的生长发育。土壤温度对微生物活性的影响极其明显。大多数土壤微生物的活动要求有 15～45℃的温度条件。超出这个范围（过低或过高），微生物的活动就会受到抑制。土壤温度对土壤的腐殖化过程、矿质化过程及植物的养分供应等都有重要意义。

4. 土壤容重

土壤容重是指土壤在未受到破坏的自然结构状况下，单位体积中的重量，通常以 g/cm^3 表示。土壤容重的大小与土壤质地、结构、有机质含量、土壤紧实度、耕作措施等有关。

（二）土壤样品采集方法

土壤样品的采集和处理是土壤分析工作的一个重要环节（陈自胜等，1992）。采集有代表性的样品，是测定结果如实反映其所代表的区域或地块客观情况的先决条件。一般依据研究目标的不同，采用分层取样分析方法，获取土壤相关数据。部分指标可以在外业直接观测记录，但大部分指标需要室内使用专用化学分析设备进行检测。为了在室内对土壤进行详细研究、观察或作为展览陈列和教学示范，通常将部分采集的土壤样品作为标本使用（郑慧莹和李建东，1999）。土壤标本主要分为散装土壤标本、分类纸盒标本、整段标本和土壤容重 4 类（柳维杨等，2007）。

1. 散装土壤标本采集法

研究人员在不同成土条件下选择具较强代表性的土壤样品，采集一定量后用于展览陈列。根据剖面层次，分层取样，依次由下而上逐层采取土壤样品，装入布袋或塑料袋，每个土层选典型部位取其中 10cm 厚的土样，一般为 0.5～1kg，要记载采样的实际深度，用铅笔填写标签，一式二份，一份放入袋中，一份挂在袋外，标签内容一般包括土壤剖面编号、土壤类型名称、采样地点、日期、采集人。

2. 分类纸盒标本采集法

根据土壤剖面层次，由下而上逐层采集原状土，挑出结构面，按上下装入纸盒，结构面朝上，每层装一格，每格要装满，标明每层深度，在纸盒盖上写明采集地点、地形部位、植物、母质、地下水位、土壤名称、采集日期及采集人。

3. 整段标本采集法

选择具有较强代表性的土壤剖面，按土壤的发生层次，采集客观真实、连续完整的土壤样品，进行加工制作而成的土壤制品（柳维杨等，2007）。在已挖好的土壤剖面上挖一个长方体土柱，其规格为 100cm×17cm×8cm，然后将采土器套在土柱上，顶部空出 3～5cm，采土器上端用螺丝杆固定；用削土刀在采土器下端切开，并削去多余的土体；在纤维板涂上胶水和采土器粘在一起，翻转放在地面上，松开螺帽，将采土器取下反折，放在三根螺丝杆上，然后将土体和纤维板一同放入采土器的三根螺杆上，削平并用毛刷洒上白乳胶，整段标本即采制成功。记录采集地点、地形部位、植物、母质、地下水位、土壤名称、采集日期及采集人。

4. 土壤容重取样方法

采用环刀法，在测量生物量样方处，挖出一个长宽高 0.5m×1m×1m 的剖面。清理干

净样方土壤表面的残留物和杂质，用环刀按照 0cm、10cm、20cm、40cm、60cm、80cm、100cm 的深度取容重土样，每层 2 个重复。具体步骤如下：①将环刀托套在环刀无刃的一端，刀刃朝下，用力均衡地压环刀托把，将环刀竖直压入土中。如土壤较硬，环刀不易插入土中时，可轻轻敲打环刀托把，待整个环刀全部压入土中，且土面即将触及环刀托的顶部时，停止下压。②用小铁铲把环刀周围土壤挖去，在环刀下方切断，并使其下方留有一些多余的土壤。取出环刀，将其翻转过来，刃口朝上，用削土刀迅速刮去黏附在环刀外壁上的土壤，然后从边缘向中部用削土刀削平土面，使之与刃口齐平。③盖上环刀顶盖，再次翻转环刀，使已盖上顶盖的刃口一端朝下，取下环刀托。同样削平无刃口端的土面并盖好底盖，将环刀盖好铝盖。④依不同层次取好土样，放于自封袋中并做好样方标记，带回室内在 105℃烘箱中烘干至恒重，称重并记录。⑤挖掘结束并拍完剖面照后，按土壤层次重新回填土壤，尽量恢复原状。

（三）土壤数据采集设备

在土壤样品采集时，一般需要准备的采样器具包括以下几类。

1）工具类：包括铁锹、铁铲、圆状取土钻、螺旋取土钻及适合特殊采样要求的工具等。

2）器材类：包括手持 GPS、罗盘、照相机、胶卷、卷尺、铝盒、样品袋、样品箱等。

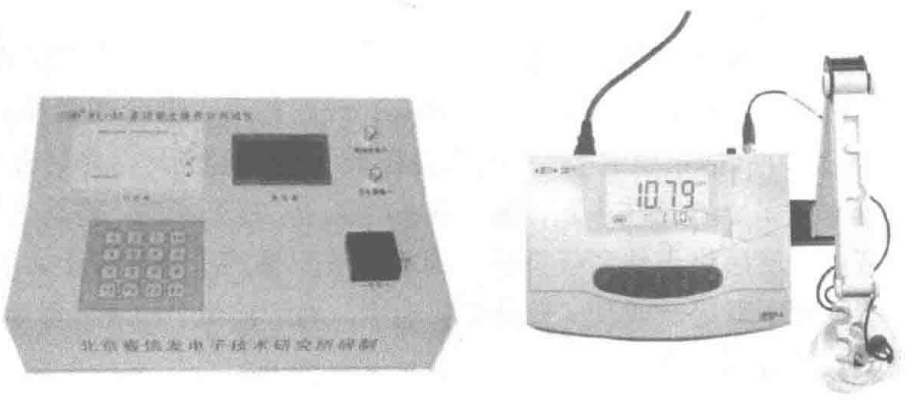
3）文具类：包括样品标签、采样记录表、铅笔、资料夹等。

4）安全防护用品：包括工作服、工作鞋、安全帽、药品箱等。

除上述这些传统器具外，还可以使用许多自动化程度较高的速测仪器。

1. 土壤养分速测仪

能检测土壤、植株、化学肥料、生物肥料等样品中的速效氮、有效磷、速效钾、有机质含量，以及植株中的全氮、全磷、全钾，土壤酸碱度、土壤含盐量、水分、紧实度（硬度）、温度。该类设备具有体积小、重量轻、便于携带、液晶显示、可以使用交流电、直接检测、查看检测数据等优点（图 2-3a）。



(a)　　(b)

图 2-3　土壤养分和 pH 速测仪

2. 多通道土壤温度记录仪

可以随时将测量时每次采样的数据存储到主机上，RS232 接口可与计算机连接将数据导出，具有存储、打印功能，内置 GPS 定位系统，可实时显示测量点的位置信息（经纬度），并可利用此定位数据在计算机中绘制土壤温度分布图。

3. 其他设备

包括土壤 pH 测定仪（图 2-3b）、土壤盐分速测仪和 GPS 土壤紧实度速测仪等，其中土壤紧实度速测仪可直接测量土壤紧实度，单位为 kg 或 kPa。

（四）历史土壤数据的获取方法

土壤类型的分类方法较多，不同国家通常采用不同的分类系统。无论在全球还是区域尺度上，目前许多机构均积累有大量的土壤数据。

中国科学院南京土壤研究所（http://www.issas.ac.cn/）建立了一个较为系统的基于 WebGIS 的中国土壤信息系统（Soil Information System of China，SIS China）。通过该系统，可以查询全国不同尺度的土壤空间数据、土壤剖面属性数据、土壤空间与属性融合后的土壤专题区域空间化数据、土壤类型参比数据及应用国际土壤主流分类方法建立的中国土壤分布特征数据。

通过世界土壤信息中心（http://www.isric.org/）、中国地球系统科学数据共享网（http://www.geodata.cn/）可以下载世界土壤数据库（HWSD v1.1）。该数据库是由联合国粮食及农业组织（FAO）、国际应用系统分析研究所（IIASA）、世界土壤信息中心（ISRIC）、中国科学院南京土壤研究所（ISSCAS）、欧洲委员会联合研究中心（JRC）于 2009 年 3 月共同发布的 1km 格网的世界土壤数据库（HWSD）。该数据库提供了各个格网点的土壤类型（FAO-74、85、90）、土壤相位、土壤（0～100cm）理化性状（16 个指标）等信息。

四、生物数据

牧区生物种类繁多，包括多种植物和动物，其中草地及家畜在牧区社会经济发展中具有极其重要的作用。草畜系统的数据采集是进行草地畜牧业资源评价、生产发展规划和生态环境保护的重要基础。特别是草地生产力不但体现草地生态系统的稳定性及生物种群的多样性，而且是制订畜牧业生产规划的基础（李建龙和蒋平，1998）。能否及时准确地掌握大范围的草地产量资料，对科学计算草地载畜量和合理安排草畜生产，提高草地畜牧业生产力，维护草地生态系统的持续稳定，都具有十分重要的意义。

（一）生物数据采集指标

依据研究地区及研究目标的不同，生物数据（Biological Data）的采集指标也有较大差异。就草地植被而言，一般需要采集其生物物理指标（如物种组成、物候期、生产力、盖度、高度等）和生物化学指标（如植物叶片的养分含量等）。就草地放牧家畜而言，

需要采集家畜种类、畜群数量、畜群结构（幼畜、成年畜、母畜的数量和比例）、种群密度（载畜力）等信息。从草地生物灾害防治角度出发，一般需要采集鼠害、虫害、病害及毒害草等方面的数据。此外，从微观尺度上，还需要采集生物信息，如微生物、植物与动物组织的显微结构、基因测序等方面的数据。限于篇幅，下面重点对草地生产力相关的数据采集方法及主要仪器设备等内容进行介绍。

（二）草地生产力数据的采集技术

草地生物数据的采集，特别是草地植被生产力数据的采集往往需要结合网络和“3S”等技术。通过网络技术，可以获取大量的数据共享资源。GPS 技术可以精确观测数据采集的地理位置、采集航迹、海拔等信息。利用 GIS 技术，可以采集已有的地图及相关属性数据。遥感技术则是动态监测草地生产力数据的关键技术。

遥感植被指数（Vegetation Index，VI）是表征地表植被特征的重要指标。在草地资源动态监测中，常用的遥感资料有陆地资源卫星数据（如 Landsat、SPOT 等）、气象卫星数据 NOAA/AVHRR 及 NASA/MODIS 等遥感数据。SPOT 卫星和 Landsat 等资源卫星数据具有多波段和较高的空间分辨率等特点，便于对水体和绿色植被等地物的识别。但是，陆地资源卫星数据由于其覆盖周期较长，不利于对研究区域实施快速的动态监测。气象卫星 NOAA/AVHRR 相对于陆地资源卫星资料，其空间分辨率较低，但时间分辨率高，时间序列较长，NOAA/AVHRR 系列卫星已有 20 多年的全球 NDVI 数据积累，便于进行草地生物量和 NPP 的长期动态监测。

新一代对地观测卫星 Terra 和 Aqua 遥感数据 MODIS 与 AVHRR 相比较，在波段和时空分辨率方面都有较大改进。MODIS 光谱分辨率较高，波幅窄，避免了几个大气吸收带，在计算植被指数时有更严格的去云算法和比较彻底的大气校正。MODIS-NDVI 是已有 20 年积累的 NOAA-NDVI 系列的延续，MODIS 植被指数可以更好地反映植被的时空变化特征，已成为当前植被及其变化动态宏观研究的主要遥感资料。MODIS-NDVI 与 AVHRR-NDVI 反映植被的趋势大致相同，但 MODIS-NDVI 比 AVHRR-NDVI 对植被的响应更敏感，NDVI 值的范围也更宽。另外，MODIS-EVI 比 MODIS-NDVI 有较大的改进，在植被监测方面，利用 MODIS 辐射仪的优点及 EVI，修正地表反射率，可以提高对高生物量区的敏感性，并通过叶冠背景信号耦合和减少大气影响可提高植被监测精度。此外，近年来针对植被监测还研发出基于 MODIS 数据的叶面积、NPP 等遥感产品。

草地生产力监测方法可以归纳为基于生物气候指标（如温度、降水等）或遥感植被指数（如 NDVI、EVI 等）的统计模型法和基于植物光合作用、有机物分解及营养元素的循环等生理过程的过程模型法两种类型。其中，统计模型法的缺点是缺乏严密的植物生理生态学机制，同时由于受到取样密度的影响，由点到面甚至向区域外推时产生的尺度转换问题会影响 NPP 的估测精度，但这类模型的输入参数简单易得，结合遥感技术可以对不同陆地生态系统 NPP 进行大范围的估算和预测，是目前应用较为广泛的一种方法。

草地生产力监测的重点之一是研究遥感植被指数与地上生物量的关系，其特点表现为所用遥感数据多样化、植被指数多样化、模型多样化，存在的主要问题表现在天地数

据的时空匹配及精度等方面。依据遥感资料及地面观测数据，建立植被指数与实地样点测产值之间的数学模型，以遥感植被指数值计算单位面积的生物量；逐个像元生物量累加，计算区域的总生物量及理论载畜量；辅以其他草地畜牧业资料（如草地面积、实际放牧家畜数量、补饲量等），可以分析草畜平衡状况，为草地生产力与生长状况评价、草畜平衡决策、生物灾害防治等研究提供基础资料。

在建立基于遥感植被指数的草地生物量统计模型时，外业观测的样地和样方的设置及其空间分布、取样密度等因素对反演模型的精度均有较大的影响。一般情况下，草地的现场取样技术应注意以下规范。

1. 样地

样地是用来描述与记载草地群落的生态环境、基本特征的典型地段，是进行草地定性与定量分析常用的方法。记载样地主要是用来对草地进行群落特征分析和确定草地类型。现场取样，首先是根据草地群落的植物种类组成、结构特征和分布的均匀性确定样地面积，面积的大小以能反映与代表所调查类型各项特征的最小面积为宜。一般的原则是，以生长草本植物为主的草地，样地面积要比木本植物为主的群落小些；群落草层低矮、结构简单、分布均匀，样地可小些，反之要大些。以往调查工作中，用于群落特征分析与记载的样地，以草本植物为主的群落，一般面积为 10m×10m；在植被稀疏的荒漠和灌丛草地，样地面积可扩大到 10m×100m 或 10m×150m。

2. 取样方法

在草地资源调查中，一般采用抽样调查的方法进行现场取样。常用的取样方法有两种，一种是典型取样法，另一种是用概率统计的方法取样。概率取样法包括随机取样、系统取样和分层取样。典型取样法是最常用的方法。在确定要调查的草地类型上，选择能够代表该类型自然与经济特征的典型地段设置样地，一般要求草群生长发育正常，未受或受家畜或其他活动干扰较小。具体做法是：①确定出样地的一个边界点，并将测绳或皮尺的一头用插钎固定，依据样地大小，圈出样地的范围。在样地圈定之后，仔细观察样地的代表性与典型性，如有缺陷，可再度移动调整。样地的数量在一个点上一般只取一个样地。如需要重复样地，在出现相同类型的其他地段再选取。②样地形状一般采用正方形。根据草地植物种类组成和分布的均匀程度，确定样地面积。最常用的是 10m×10m，也可以用 10m×20m、10m×100m 等。样地的记载内容分为一般情况和生境特征的描述与记载，以及草地群落特征的调查与记载。③草地生物量野外调查取样需要考虑样方的定位、样方的大小、样方的数量等问题。一个样地内一般取 3 个样方，采用随机取样法、五点式、对角线式、棋盘式、平行线式等方法设置样方。就均匀分布的情况而言，只要采取简单的随机取样方法即可，因为种群中每个个体都有相等的被选择机会；而对于其他的分布情况如系统分布、分层分布、梯度分布等，采取简单的随机取样方法未必能获得有代表性的样本。为了避免简单随机取样方法的不足，可以采用分层随机取样法，即将区域分为若干亚区域，然后在亚区域上进行随机取样。样方大小应视调查对象的大小和分布情况而定。一般而言，对于分布不规则的种群，多个小样方取样比

少数大样方取样效果更好。样方的数量越多，整体取样花费的时间就越多、精力就越大。为了提高取样的效率，人们总结了一些系统取样的经验，如五点取样法和等距取样法等。

（三）草地植被数据采集设备

目前有许多类型的观测设备，下面主要介绍一些外业观测的常用仪器设备。

1. 叶绿素测定仪

植物叶片中的叶绿素含量可指示植物本身的生长状况，长势良好的植物的叶子会含有更多的叶绿素，叶绿素的含量与叶片中氮的含量也有很密切的关系。叶绿素测定仪可以即时测量植物的叶绿素相对含量或“绿色程度”（图 2-4a）。一般通过测量叶片在两种波长范围内的透光系数来确定叶片当前叶绿素的相对数量，也就是在叶绿素选择吸收特定波长光的两个波长区域，根据叶片透射光的量来计算测量值。

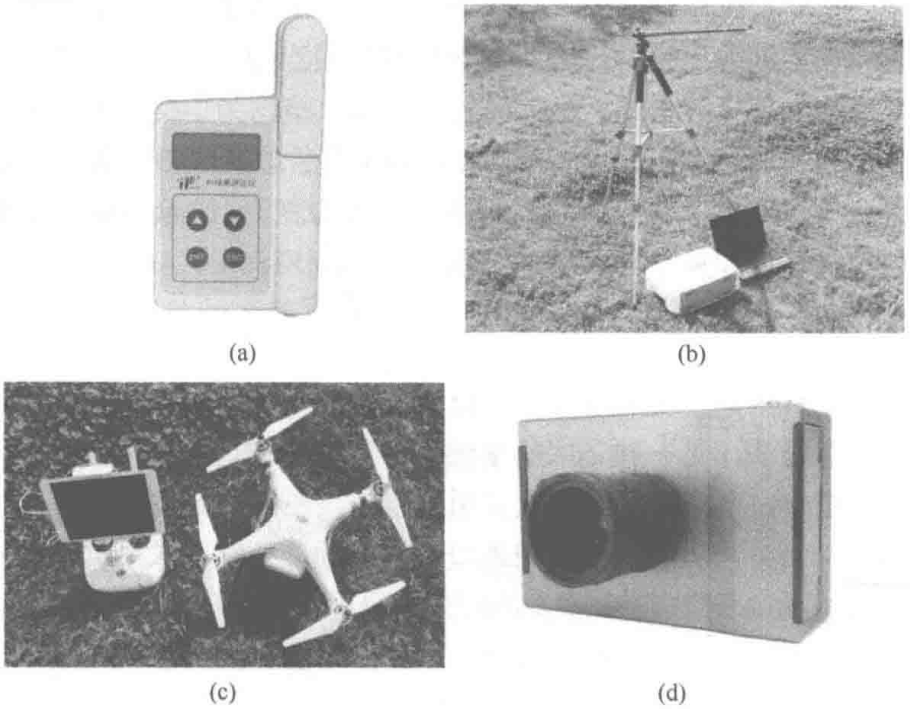


图 2-4　叶绿素测定仪（a）、地物光谱仪（b）、无人机（c）和农业多光谱相机（d）

2. 活体叶面积测定仪

叶面积测定仪是一种使用方便、可以在野外工作的便携式仪器。可测量被测叶片的叶面积，并将其面积值在仪器上显示和保存。如需细致的科研分析，还可将保存的数据导入电脑中用软件进行统计分析，能直观显示叶片的大小和形状，能分析单叶片的面积、虫斑数和虫斑面积，以及多张叶片的面积和虫斑面积等参数。此设备广泛应用于农业、草业、气象、林业等部门。

3. 植物光合作用测定系统

植物光合作用测定系统分为便携式光合作用测定系统和全自动便携式光合作用测定系统。测量参数包括 CO_2 浓度、净光合速率、蒸腾速率、胞间 CO_2 浓度、气孔导度、

大气湿度、空气温度、叶片温度、蒸汽压亏缺、大气压、光照强度等，并通过系统自带的自动测量程序测定植物的 CO_2-光合响应曲线、温度-光合响应曲线、湿度-光合响应曲线等各种响应曲线，还可以测量植物的叶绿素荧光效能。

4. 地物光谱仪

光谱测量范围一般为 350～2500nm，可采集研究区土壤、草地、水体等地表物体的光谱数据，与多光谱和高光谱传感器配合使用，为草地遥感算法构建、遥感模型验证、传感器通道设置与调整等提供基础（图 2-4b）。

此外，无人机（图 2-4c）、普通数码相机和农业多光谱相机（图 2-4d）在草地植被盖度、生长状况等方面的数据采集中，也发挥着越来越重要的作用，具有广阔的应用前景。

（四）历史生物数据的获取方法

在草地植被分类、草地生产力遥感动态监测、草地灾害等方面，世界各地许多机构已积累了大量数据。大多数数据通过相应网站均可以进行查询、检索、订购、下载。下面对一些重要网址及内容作一简介。

1）中国草业开发与生态建设网（http://www.ecograss.com.cn/），包括我国天然草地类型、栽培牧草品种、草地鼠害、虫害、病害及毒害草等方面的数据，以及青藏高原地区雪灾遥感监测与预警、甘南地区草畜平衡决策支持、基于综合顺序分类系统的世界草地分类等方面的数据。

2）寒区旱区科学数据中心（http://westdc.westgis.ac.cn/），可以订购全国及部分地区的遥感植被指数和植被类型等方面的数据。

3）通过中国科学院遥感卫星地面站（http://www.rsgs.ac.cn/）、美国 NASA MODIS 数据中心（http://modis.gsfc.nasa.gov/）、美国国家冰雪数据中心（http://nsidc.org/data/modis/）、美国马里兰大学 MODIS 地表反射率研究中心（http://modis-sr.ltdri.org/）和美国地质调查局（USGS）（http://glovis.usgs.gov/）等网站，可以下载遥感数据和部分地图空间数据。

4）中国科学院数据云（http://www.csdb.cn/），可查询下载中国生物资源、生态系统与生态功能区划等数据。

五、人文数据

人文数据（Social Economic Data）主要指社会经济与人类活动等方面的数据，主要包括行政区划、人口、经济、交通、通信等数据。其中，人口及社会经济发展状况数据的采集大多采用抽样调查统计方法；历史时期的数据一般可以通过收集区域统计年鉴、社会经济发展报告等资料进行采集，部分数据也可以通过国家统计局和国家商务部等机构，以及中国科学院地理科学与资源研究所开发的人地系统主题数据库（http://www.data.ac.cn/index.asp）、美国密歇根州立大学中国信息研究中心（http://umchina.umich.edu/）等相关网站进行订购。

六、其他数据

草业生产过程和草地生态环境变化动态都与人类生存和发展的水资源、土地资源、生物资源及气候资源的数量与质量状况息息相关。草业科学研究面临的草地退化、灾害防治等复杂系统问题的解决，不仅需要采集地形地貌、土壤、气象、生物和人文数据，还需要采集一些与自然及人文因素均密切相关的数据，如栽培牧草区划、草业生态经济分区、土地利用与土地覆盖等数据，这类数据也可以通过网络和“3S”等技术进行采集。

思 考 题

1. 简述草业数据类型及其基本特征?
2. 简述草地生产力数据的采集方法及地面观测设备？
3. 举例说明 GPS 技术在草业数据采集方面的作用？
4. 论述草业数据采集的基本技术？
5. 论述“3S”技术在草地植被动态监测数据采集方面的应用？

参 考 文 献

曹卫星. 2005. 农业信息学[M]. 北京: 中国农业出版社.

陈自胜, 赵明清, 刘文志, 等. 1992. 羊草草地松土施肥效果的研究[J]. 中国草地, (2): 28-32.

龚子同, 张甘霖. 2006. 中国土壤系统分类: 我国土壤分类从定性向定量的跨越[J]. 中国科学基金, (5): 293-296.

李建龙, 蒋平. 1998. 遥感技术在大面积天然草地估产和预报中的应用探讨[J]. 武汉测绘科技大学学报, 23(2): 153-158.

李绍良, 陈有君, 关世英, 等. 2002. 土壤退化与草地退化关系的研究[J]. 干旱区资源与环境, 16(1): 92-95.

柳维杨, 吕双庆, 姜益娟, 等. 2007. 土壤标本及整段标本采集制作方法分析[J]. 安徽农业科学, 35(32): 10394, 10399.

伍光和. 2000. 自然地理学[M]. 3 版. 北京: 高等教育出版社.

郑慧莹, 李建东. 1999. 松嫩平原盐生植物与盐碱化草地的恢复[M]. 北京: 科学出版社.

Hijmans R J, Cameron S E, Parra J L, et al. 2005. Very high resolution interpolated climate surfaces for global land areas[J]. International Journal of Climatology, 25(15): 1965-1978.

IPCC. 2007. Climate change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change[R]. Geneva, Switzerland.

第三章　草业信息统计分析与处理方法

在草业科学领域中通过科学试验、实地调查和遥感等可以获得大量数据，如何从海量数据中获取有价值的信息，数据的采集、处理、分析及专业的评估和预测报告等工作变得越来越重要。

本章主要介绍草业信息处理常用的统计分析方法、常用统计分析软件简介［包括Excel、SAS 系统（Statistics Analysis System）、SPSS（Statistical Product and Service Solutions）、MATLAB（MATrix LABoratory 矩阵实验室）］、草业信息处理的地统计学方法、多目标规划方法，以及相关方法的应用。

第一节　草业信息处理常用统计分析方法及软件简介

随着计算机技术的迅猛发展，相继涌现出各种版本的统计分析软件，使科研工作者用成熟的统计分析软件来处理数据成为可能。这不仅可以充分挖掘和利用数据中的信息，而且还可以提高统计分析的效率。

一、常用统计分析方法概述

（一）相关分析

相关分析就是要研究变量之间相互关系的密切程度，并能通过已经获得的样本数据推断总体是否相关。其显著特点是变量不分主次，被置于同等地位。

1. 直线相关分析

直线相关（或线性相关）是指在相关关系中设有自变量 x 和因变量 y，当自变量 x 发生变化时，因变量 y 值随之发生大致均等的变化，在图像中近似地可以用一条直线来表示，这种相关通称为直线相关。相关系数是反映两个变量 x、y 之间线性关系密切程度的指标，记作 r，其值介于–1～+1。如果 x 和 y 呈正相关，r 为正值，当 r=1 时称为完全正相关；如果 x 和 y 呈负相关，r 为负值，当 r=–1 时称为完全负相关。无论是完全正相关还是完全负相关，观测点几乎都落在直线回归线上。而当观测点在直线回归线周围的分布越离散时，r 的绝对值越小。当 r 与+1 或–1 越接近，则相关性越紧密。与 0 越接近，相关性越不紧密，特别当 r=0 时，反映出自变量 x 和因变量 y 之间无直线相关关系。

Pearson 简单相关系数用来反映两个等距型变量之间的线性相关密切程度，如在测度草原产草量与降水量的线性相关关系时，可以使用 Pearson 简单相关系数，它是形成其他相关系数的基础。

Pearson 简单相关系数的计算公式为

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2(y_i-\overline{y})^2}}$$

其中，n 为样本数，x_i 和 y_i 分别为两个变量的变量值。

Pearson 简单相关系数的检验统计量为 T 统计量，其定义为

$$T=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

其中，T 统计量服从自由度为 n–2 的 T 分布。

2. 多元线性相关分析

将一元线性相关分析扩展到多元的情况，就是多元相关分析或复相关分析。多个变量之间的相关关系是相当复杂的，由于受其他变量的影响，相关系数不能准确地反映出两个变量之间的线性相关程度。

在具有多元相关关系的变量中，复相关系数可以用来衡量因变量 y 与一组自变量 $x_1,x_2,\cdots,x_m$ 之间的相关程度。和简单相关系数相同，其值仍然介于–1～+1。复相关系数的计算公式为

$$r_{y,123\cdots m}=\sqrt{1-\frac{S_E}{S_T}}=\sqrt{1-\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\overline{y}_i)^2}}$$

式中，S_E 为剩余离差平方和，即 $\sum(y_i-\hat{y}_i)^2$，S_T 为总离差平方和，即 $\sum(y_i-\overline{y}_i)^2$。

偏相关分析是指当其他变量固定不变时，分析两个变量之间的线性相关，用偏相关系数来度量关联的强度。

例如，设有三个变量 x_1, x_2, x_3，如果在 3 个变量中，剔除其中一个的影响，可计算其他两个对该变量的偏相关系数，其计算公式为

$$r_{12,3}=r_{x_1x_2,x_3}=\frac{r_{x_1x_2}-r_{x_1x_3}\cdot r_{x_2x_3}}{\sqrt{1-(r_{x_1x_3})^2}\sqrt{1-(r_{x_2x_3})^2}}$$

$$r_{13,2}=r_{x_1x_3,x_2}=\frac{r_{x_1x_3}-r_{x_1x_2}\cdot r_{x_3x_2}}{\sqrt{1-(r_{x_1x_2})^2}\sqrt{1-(r_{x_3x_2})^2}}$$

$$r_{23,1}=r_{x_2x_3,x_1}=\frac{r_{x_2x_3}-r_{x_2x_1}\cdot r_{x_3x_1}}{\sqrt{1-(r_{x_2x_1})^2}\sqrt{1-(r_{x_3x_1})^2}}$$

偏相关系数的检验统计量为 T 统计量，其定义为

$$T=\frac{r\sqrt{n-q-2}}{\sqrt{1-r^2}}$$

式中，r 为偏相关系数，n 为样本数，q 为阶数。T 统计量服从自由度为 n–q–2 的 T 分布。

实例分析：2000～2010 年石羊河流域 NPP 时空变化及驱动因子（李传华和赵军，2013）。

基于MODIS卫星遥感数据，以NPP变化与气候变化和人为影响之间的关系为研究内容，利用ArcGIS软件中的IDW插值方法，得到2000～2010年降水与气温分布图，像元大小为1km×1km。使用SPSS软件分别对NPP与降水和气温逐像元进行相关分析，计算每个像元的NPP分别与降水和气温的相关系数，根据NPP像元ID进行属性连接，完成相关系数的空间化（图3-1），分析了2000～2010年石羊河流域植被NPP变化的人为影响的作用（图3-2）。

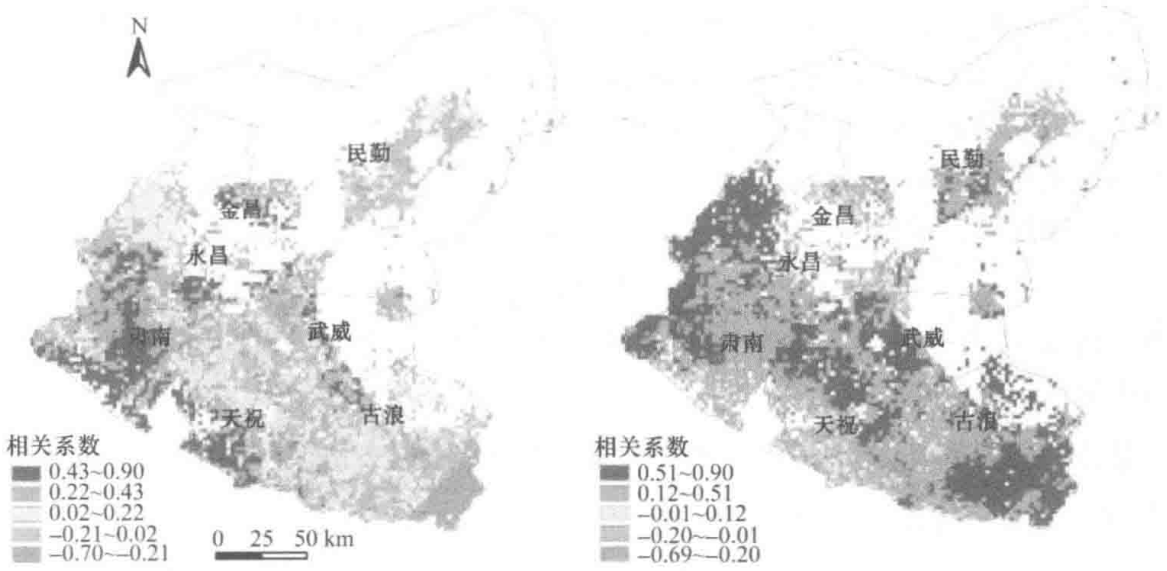


图3-1　NPP与年均温（左）和年降水（右）相关系数示意图

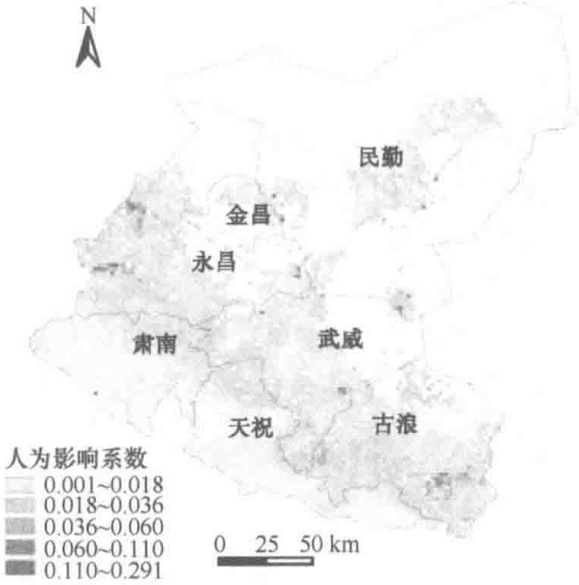


图3-2　2000～2010年植被NPP变化的人为影响空间分布示意图

（二）回归分析

由于相关关系不能反映出变量间的因果关系，因此要明确因果关系必须借助于回归分析。回归分析是研究一个因变量或多个因变量与一个自变量之间是否存在某种线性或

非线性因果关系的统计方法。其主要任务是在已有的相关分析基础上，对具有相关关系的两个或多个变量之间进行一般关系的数量变化测定，建立一个数据模型，从而能够通过一个已知量推断另一个未知量。也就是说，根据数据估计参数，建立模型，对参数和模型进行检验和判断，并进行预测等。与相关分析的主要区别是，研究目的和模型假设不同（贾俊平，2008）。

1. 一元线性回归

即直线回归分析，就是从相关关系的具体形态出发，选择恰当的直线回归模型，来表示两个数量变量之间的依存关系。在直线回归分析中因变量是依赖自变量而变化的，通常分别用 y 和 x 来表示。其中，x 可以是非随机变量（即变化规律或人为确定的数值），也可以是随机变量。

一元线性回归模型为

$$y = \beta_0 + \beta_1 x + \varepsilon$$

式中，β_0、β_1 为参数，$\beta_0 + \beta_1 x$ 为由变量 x 变化引起的 y 值变化的线性部分，ε 为其他随机因素所引起的误差项。

总体回归参数 β_0 和 β_1 是未知的，必须利用样本数据去估计，可用样本统计量 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 代替。从而，估计的回归方程为

$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$

可以看出，一元线性回归是忽略其他因素或设定这些因素为已知时，分析变量 x 是如何线性影响 y 的，是一种相对理想化的分析。

2. 多元线性回归

在处理实际问题的过程中，一元回归分析只是一种特殊情形，很多时候还是要讨论一个随机变量与若干个变量之间的相关关系，这就需要借助多元回归分析的方法来解决。由于多元回归分析的复杂性，仅简要介绍多元线性回归分析。

多元线性回归分析（或复线性回归分析）是一元线性回归分析的进一步推广，主要研究一个因变量如何受一组自变量的直接影响。其中的自变量 x 是指能独立自由变化的变量，而因变量 y 是非独立的、受其他变量影响的变量。因其仅涉及一个因变量，所以也被称为单变量线性回归分析。

多元线性回归模型为

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

式中，$\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$ 为由 p 个变量 x 变化引起的 y 值变化的线性部分，ε 为其他随机因素所引起的误差项，$\beta_0, \beta_1, \cdots, \beta_p$ 均为模型中的偏回归系数。

多元线性回归模型的回归方程为

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$$

从统计意义上讲，多元线性回归分析可在忽略其他自变量的影响后，分析每一个自

变量的变化是否能引起因变量的变化，并且在固定其他自变量不变的情况下，估算出每个自变量对因变量影响大小的数值。

实例分析：2001～2010 年青藏高原草地生长状况遥感动态监测（冯琦胜等，2011）。

基于 1970 个地面实测数据，结合 MODIS-EVI 和 NDVI 数据，利用留一法交叉验证方法（LOOCV）确定了适合青藏高原地区草地生长状况的遥感反演模型，估算了 2001～2010 年草地生物量的干重空间分布格局，分析了近 10 年草地生物量变化动态。

利用 2005 年和 2006 年青藏高原 1970 个草地样方数据，结合同时相的 MODIS-EVI 和 NDVI 数据，构建了青藏高原草地风干重和盖度与两种植被指数之间的线性、指数、对数和乘幂函数模型，并用 LOOCV 比较分析了不同模型的精度（表 3-1）。

表 3-1　草地风干重和盖度模型精度评价

指数	模型	风干重/（kg/hm^2）			盖度/%		
		R^2	RMSEP	r	R^2	RMSEP	r
EVI	线性	0.36	683.29	0.6	0.42	18.63	0.65
	指数	0.46	687.92	0.59	0.38	19.51	0.61
	对数	0.32	734.83	0.57	0.46	19.82	0.68
	乘幂	0.47	815.63	0.6	0.46	23.47	0.67
NDVI	线性	0.37	678.94	0.61	0.47	17.7	0.69
	指数	0.49	671.80	0.62	0.44	19.06	0.67
	对数	0.31	717.33	0.56	0.48	17.82	0.69
	乘幂	0.47	693.68	0.6	0.49	17.54	0.7

结果表明：EVI 与风干重的指数模型表现最好。在 EVI 与盖度的 4 种模型中，指数模型的 R^2 最低。对数模型和乘幂模型的 R^2 相等，但对数模型的 RMSEP 较低，所以 EVI 与盖度的对数模型精度最好。NDVI 与风干重的指数模型总体精度最好，R^2、RMSEP 和 r 依次是 0.49、671.80 和 0.62；而 NDVI 与盖度回归模型中，乘幂模型表现最好。

总体而言，在青藏高原地区 NDVI 植被指数较 EVI 有更好的表现，指数模型模拟的风干重精度最高，而盖度适合使用乘幂模型反演。

（三）时间序列分析

时间序列指将同一现象在不同时间上取得的观测值按时间顺序排列而成的序列。按照连续性可分为离散型和连续型。在实际问题中，为了研究和叙述的方便，我们一般考察的时间序列都是离散型随机过程和时间序列，即观测值是从相同时间间隔点上得到的。而所谓时间序列分析，是一种根据动态数据揭示系统动态结构和规律的统计方法。其基本思想：根据系统的有限长度的运行记录（观察数据），建立能够比较精确地反映序列中所包含的动态依存关系的数学模型，并借以对系统的未来进行预报（王振龙，2000）。时间序列分析有两种目的：一是揭示其发展规律，预测未来发展趋势。采用的方法有平滑预测法（滑动平均、移动平均、指数平滑）、趋势线预测法（直线、指数型曲线、抛物线）、季节型预测法自回归模型。二是研究趋势变化，消除异常值和噪声污

染。采用的方法有阈值法（The Best Index Slope Extraction，BISE）、滤波法（Savitzky-Golay；Mean-value Iteration Filter）、函数拟合法（Fourier Transform；Harmonic Analysis of Time Series；Asymmetric Gaussian Model Function；Double Logistic Function）。

实例分析：MODIS 植被指数时间序列 Savitzky-Golay 滤波算法重构（边金虎等，2010）。

利用 Savitzky-Golay（S-G）滤波方法对若尔盖高原湿地区 2000～2009 年 MODIS 16d 最大值合成的 NDVI 时间序列数据进行了重构，并与中值迭代滤波法、傅里叶变换法进行了比较。结果表明，基于 S-G 滤波的时间序列重构方法重构后的 NDVI 时间序列在直观及像元的时间序列曲线上均取得了较好的效果，对提高该数据产品质量有很大帮助，通过该方法重构后的高质量的 NDVI 时间序列为利用该数据源对若尔盖湿地生态系统监测提供了良好的基础。

对全部数据进行预处理，首先采用 MODIS 网站上提供的 MRT（MODIS Reprojection Tool）工具对研究区影像进行重投影，影像的原始投影方式为 Sinusoidal 投影，重投影为 Albert 标准投影。其次对影像中填充值进行修补，最后乘以尺度因子以获取每个像元的 NDVI 值。以研究区土地利用为基础，结合预处理后的 NDVI 时间序列随机选取 10 个点作为 NDVI 曲线验证点，部分验证点 NDVI 曲线如图 3-3 所示。由图可以看出，验证点的时间曲线并不是十分圆滑，且部分时间段 NDVI 值发生突降。部分突降点对应其像元可信度中的冰雪像元。这些突降点在植被生长的周期中是不合理的，应该作为噪声点给予修正。

（四）系统聚类分析

聚类分析是一种多元统计分类方法，根据样本自身属性，按照某种相似性或差异性指标，定量地确定样本之间的相似性，以此对样本进行聚类。在聚类分析过程中，首先，要寻找到度量事物相似性的统计量，聚类分析中用来衡量样本个体之间属性相似程度的统计量和用来衡量指标变量之间属性相似程度的统计量是不同的，前者用的统计量是距离系数，后者用的统计量是相似系数。距离系数的定义有很多，如欧氏距离、极端距离、绝对距离等。相似系数的定义也很多，如相关系数、列联系数等。其次，要寻找到恰当的分类方法。聚类分析的方法很多，常见聚类方法有系统聚类、动态聚类等。系统聚类法适用于事先并不清楚应分成多少类型的问题，形成的类只是分析的结果，即小样本的样本聚类或指标聚类。动态聚类法适用于解决事前已初步确定大致可以分成多少类的问题，在研究中需对何种样本分类，即大样本的样本聚类（汪海波等，2013），在这里我们主要介绍系统聚类分析法。

系统聚类法的一般步骤如下。

1）数据标准化对所获得的原始数据，根据样本各变量的观察值予以分类，包括总和标准化、标准差标准化、极大值标准化、极差标准化等。

2）距离计算，在做聚类分类前必须计算出类与类之间的距离。常用的计算方法有绝对距离、欧氏距离、明科夫斯基距离、切比雪夫距离等。

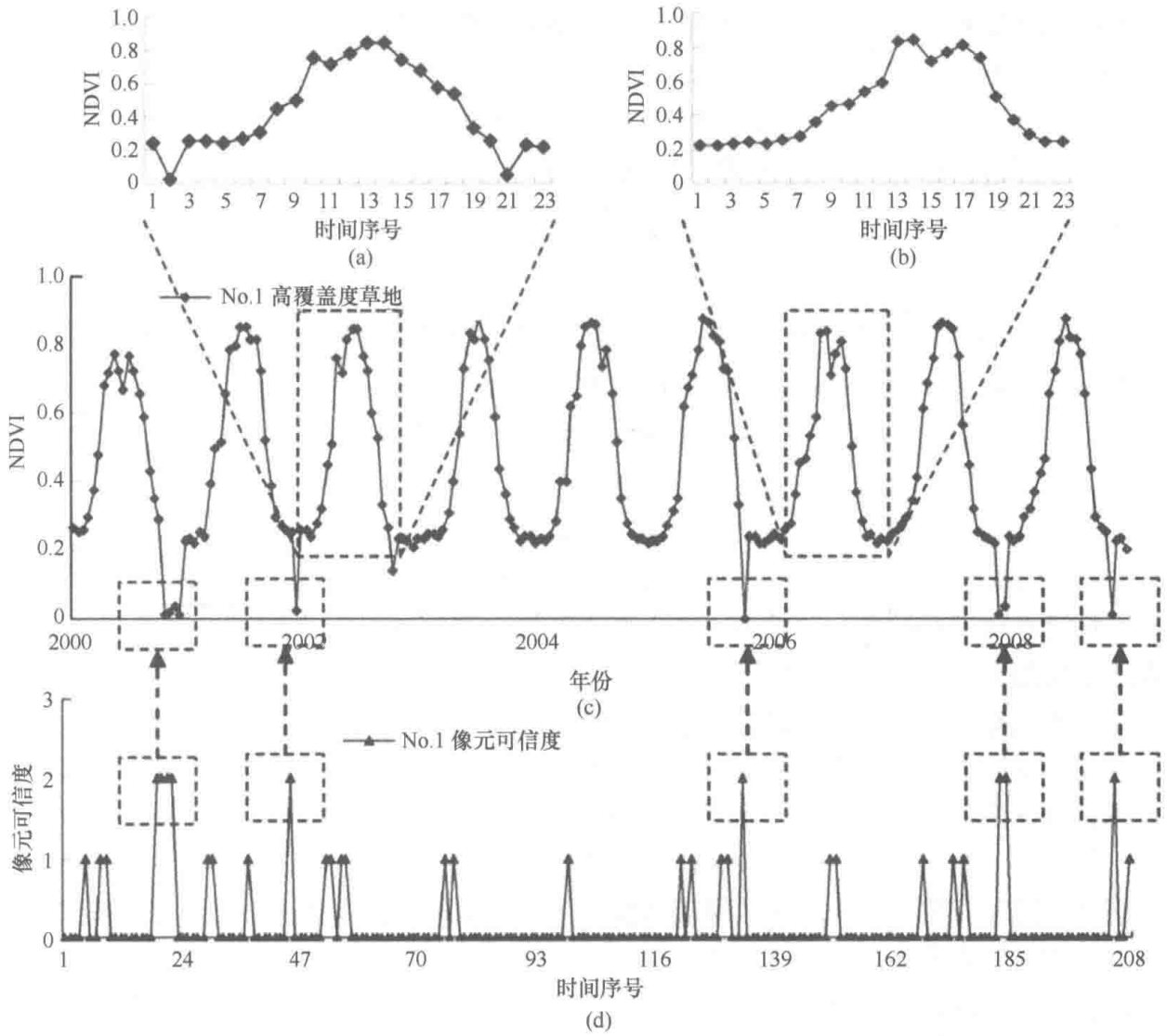


图 3-3　验证点的 NDVI 曲线及其像元可信度

（a）验证点 2002 年 NDVI 的时间序列曲线；（b）验证点 2006 年的 NDVI 时间序列曲线；（c）验证点 2000～2009 年的整体时间序列曲线；（d）该验证点时间序列曲线对应的像元可信度。其中，第 1 期 NDVI 时间序号为 1，最后一期 NDVI 时间序号为 208

3）分类，先将 n 个样本各自看成一类，并规定样本与样本之间的距离和类与类之间的距离。开始时，因每个样本自成一类，类与类之间的距离与样本之间的距离是相同的。然后在所有的类中，选择距离最小的两个类合并成一个新类，并计算出所得新类和其他各类的距离；重复前面的步骤，共进行 n–1 次，每次减少一类，直至将所有的样本都合并成一类为止。这样一种连续并类的过程可用一种类似于树状结构的图形即聚类谱系图（俗称树状图）来表示，由聚类谱系图可清楚地看出全部样本的聚集过程，从而可做出对全部样本的分类（陈正昌等，2005）。

（五）主成分分析

在实际问题中，变量之间存在一定的相关性，如何找到较少的几个彼此独立的综合指标来反映原有众多变量的信息。主成分分析正是这种综合性的分析方法，它将一组彼此相关的指标变量转化为一组新的彼此独立的指标变量，并用其中较少的几个新指标变

量就能综合反映原多个指标变量中所包含的主要信息，核心是进行降维处理（Jolliffe，1986）。

假设有 k 个指标 $x_1, x_2, \cdots, x_k$，每一个指标有 n 个观测值，它们的标准化指标变量是 $X_1, X_2, \cdots, X_k$，对原始指标数据进行标准化变换

$$X_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j}, j = 1,2,\cdots,k$$

式中，$\overline{x}_j$ 是第 j 个指标的均值，s_j 是第 j 个指标的标准差。它们的综合指标（新变量指标）为 $z_1, z_2, \cdots, z_m$（$m \leqslant k$），则进行线性变换

$$\begin{cases} z_1 = l_{11}X_1 + l_{12}X_2 + \cdots + l_{1k}X_k \\ z_2 = l_{21}X_1 + l_{22}X_2 + \cdots + l_{2k}X_k \\ \vdots \\ z_k = l_{k1}X_1 + l_{k2}X_2 + \cdots + l_{kk}X_k \end{cases}$$

将标准化指标变量 $X_1, X_2, \cdots, X_k$ 转换成新的变量 $z_1, z_2, \cdots, z_k$，则 $z_1, z_2, \cdots, z_k$ 是 $X_1, X_2, \cdots, X_k$ 的 k 个主成分，其中 z_1 是第一主成分，z_2 是第二主成分，……，z_k 是第 k 主成分。称 l_{ij} 为第 i 个主成分在第 j 个标准变量上的得分系数。将每一个样本的标准化观察值代入计算公式中，计算得到每一个样本的 k 个主成分值（汪海波等，2013）。

实例分析：基于近 20 年遥感数据的藏北草地分类及其动态变化（毛飞等，2007）。

用 1982～2000 年 NOAA / AVHRR 的旬合成归一化植被指数（NDVI）资料，采用主成分分析和非监督分类方法对藏北那曲地区植被进行分类，分析不同草地类型代表像元的 NDVI 年内和年际变化特征；定义那曲地区牧草主要生长期平均 NDVI≥0.1 的地区为植被区，NDVI<0.1 的地区为植被稀少区，进一步分析植被区每个像元 NDVI 的时空变化特征。结果表明，该地区草地类型可分为高寒草甸、高寒草甸草原、高寒草原和高寒荒漠，分类结果与实际情况相符。

用于草地分类的 NDVI 数据集有 18 维，将这 18 个旬（18 维）的多年平均 NDVI 值看作 18 个变量（x_1：5 月上旬，x_2：5 月中旬，x_3：5 月下旬，…，x_{16}：10 月上旬，x_{17}：10 月中旬，x_{18}：10 月下旬），样本量是那曲地区的像元数，为 6513（古德投影，网格尺寸 8km×8km），以此建立一个样本数据阵（6513×18）。由该样本数据阵，用主成分分析方法得到样本相关阵的特征向量（表略）、特征根、贡献率和累积贡献率（表 3-2 给出前 10 个主成分）。由表 3-2 可以看出，当取前 3 个主成分时，其累积贡献率达 99.76%，其中第 1 主成分的贡献率为 94.9959%，即通过主成分分析后得到的第 1 主成分（1 个新变量）就包含了原来 18 个变量的 95.00%的信息。因此，下面只讨论第 1 主成分的表达式。

如果样本数据阵的主成分用 $y_1, y_2, \ldots, y_{18}$ 表示，则第 1 个主成分的表达式为

$$\begin{aligned} y_1 = {} & 0.1203\,x_1 + 0.0131\,x_2 + 0.1477\,x_3 + 0.1701\,x_4 + 0.1987x_5 \\ & + 0.2318\,x_6 + 0.2665\,x_7 + 0.2904\,x_8 + 0.3051x_9 + 0.3140\,x_{10} \\ & + 0.3117\,x_{11} + 0.3015\,x_{12} + 0.2850\,x_{13} + 0.2621\,x_{14} + 0.2298\,x_{15} \\ & + 0.1964\,x_{16} + 0.1659\,x_{17} + 0.1431\,x_{18} \end{aligned}$$

表 3-2　样本相关阵的特征根、贡献率和累积贡献率

序号	特征根	贡献率/%	累积贡献率/%
1	3266.5990	94.9959	94.9959
2	154.4950	4.4929	99.4888
3	9.3737	0.2726	99.7614
4	4.2585	0.1238	99.8852
5	1.3957	0.0406	99.9258
6	0.6014	0.0175	99.9433
7	0.3173	0.0092	99.9525
8	0.2343	0.0068	99.9593
9	0.2023	0.0059	99.9652
10	0.1780	0.0052	99.9704

从第 1 主成分 y_1 来看，它是变量 $x_1, x_2, \cdots, x_{18}$ 的线性函数，且 7 月上旬至 9 月上旬（$x_8 \sim x_{13}$）的系数比较大，说明对第 1 主成分的贡献率比较大，其中系数最大的 x_{10} 在 8 月上旬。这期间正是那曲地区牧草生长旺季，是一年中 NDVI 值最大的时段。分析发现，那曲地区 NDVI 年变化曲线呈单峰型，最高值一般出现在 8 月。在贡献率最大的第 1 主成分中，变量 $x_8 \sim x_{13}$ 的系数最大，其 NDVI 值也最大，由此可以认为，那曲地区 7 月中旬至 9 月上旬，尤其是 8 月上旬是研究植被变化的关键期，这一时期 NDVI 的变化能很好地代表植被的年际变化。

二、常用统计软件分析模块简介与应用实例

下面介绍几种常用的统计软件分析模块，以便为以后应用统计方法解决草业科学中的实际问题奠定一定基础。

（一）统计分析软件 Excel 的数据分析方法

Microsoft Excel 是美国微软公司开发的 Windows 环境下的数据处理软件，具有强大的计算功能，其丰富的函数使其具有一定的统计计算功能。尽管 Excel 并不属于专用统计软件，利用 Excel 的统计计算功能对数据进行统计分析，既解决了手工计算时出现的工作量大、繁琐、费时、容易出错等问题，又避免了使用专业统计分析软件时出现的操作复杂、编程困难等弱点。Excel 可以完成求平均数、次数分布、标准差和变异系数、区间估计、方差分析、相关和回归分析、F 检验、回归分析、直方图、t 检验等常规统计分析，甚至做多元非线性回归（龚江等，2011）。

但是 Excel 在统计分析方面还是有一定的局限性。例如，Excel 2003 的最大行数为 65 536 行，最大列数为 256 列，Excel 2007 的最大行数为 1 048 576 行，最大列数为 16 384 列，这限制了数据的使用范围。再如 Excel 没有直接提供相关系数的 p 值、方差分析中的多重比较、非参数检验方法等一些常用的统计方法，而且部分计算结果不够稳定。因此，在实际应用时主要用于对数据的初步处理，以便及时发现试验研究过程中可能存在的数据采集、录入等问题，应该避免采用 Excel 进行复杂计算。

（二）统计分析软件 SAS（Statistics Analysis System）的数据分析方法

SAS 被誉为统计分析的标准软件，是目前在数据处理和统计分析领域中最为流行的一种大型统计分析系统。它由最初的用于统计分析经不断发展和完善而成为大型集成应用软件系统，具有完备的数据存取、管理、分析和显示功能。SAS 系统具有功能齐全、使用便捷、编程语言完善、数据处理和统计分析融为一体、适用性广泛等特点，以数据管理和数据分析融为一体为 SAS 系统的最大特点（罗纳德·科迪和杰弗里·史密斯，2011）。

SAS 是一个模块软件系统，它由多个功能的模块组合而成，其核心软件模块有以下几方面。

1. 基础部分——BASE SAS

SAS 系统的基础，主要功能是数据管理和数据加工处理，并有报表生成和描述统计的功能。BASE SAS 软件可以单独使用，也可以同其他软件产品一起组成一个用户化的 SAS 系统。

2. 统计分析计算部分——SAS/STAT

完整的统计分析模块，主要包括回归分析、方差分析、属性数据分析、多变量分析、判别分析、聚类分析、残存分析、心理测验分析和非参数分析等 8 类方法，共 40 多个过程。

3. 绘图部分——SAS/GRAPH

该模块能够完成多种绘图功能，如直线图、二维和三维图、直方图、饼图、星形图、地理图和各种映像图。除此之外，还具有对图进行全屏幕编辑和修改等功能。

4. 矩阵运算部分——SAS/IML

该模块用于矩阵的运算，功能完善的编程语言，使得用户可根据自己的需要编写各种矩阵运算的程序。

5. 运筹学和线性规划——SAS/OR

该模块用于运筹学和工程管理，可以帮助用户实现对人力、时间及其他资源的最佳利用。模块中包括通用的线性规划、正数规划、混合整数规划和非线性规划方法。

6. 经济预测和时间序列分析——SAS/ETS

主要用于计量经济与时间序列分析，利用该模块可建立各种统计模型，并进行相应的模拟和预测。此外，模块中还提供许多处理时间序列数据的实用程序，如时间频率转换和季节调整等。

我国很多研究机构如国家信息中心、国家统计局、中国科学院等都是 SAS 系统的大用户。SAS 可以同时处理多个数据集，尽管提供了菜单式操作界面，仍然需要通过学习掌握一定的程序编译方法才可以使用。因此，该统计软件主要适合于统计工作者和科

研工作者使用。

（三）统计分析软件 SPSS（Statistical Product and Service Solutions）的数据分析方法

SPSS 是一种运行在 Windows 系统下的统计软件工具包，是世界上最早的统计分析软件。SPSS 采用类似 Excel 的数据编辑窗口作为操作界面，可以自己定义数据属性，最多可使用 4096 个变量。统计分析方法涵盖面广，其基本功能包括数据管理、统计分析、图表分析、输出管理等。SPSS 的统计分析过程包括描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、聚类分析、数据简化、生存分析、时间序列分析、多重响应等几大类，每类中又分几个统计过程，如回归分析中又分线性回归分析、曲线估计、Logistic 回归、Probit 回归、加权估计、两阶段最小二乘法、非线性回归等多个统计过程。它也包括多元统计技术，如多元回归分析、聚类分析、判别分析、主成分分析和因子分析等方法。并且，随着它的功能不断完善，统计分析方法不断充实，大大提高了统计分析工作的效率。但要注意的是，SPSS 在数据管理功能方面相对较弱，每一个 SPSS 过程只能同时打开一个数据文件，这使得在使用过程中缺乏可操作性（薛微，2013）。

由于 SPSS 功能齐全，操作简便，价格合理，因此很快地应用于通信、银行、医疗、科研教育等多个领域和行业，是目前应用最广泛的专业统计软件之一。SPSS 软件对于非统计学专业的使用者而言，只要关心某个具体问题需采用何种统计方法，并基本了解其计算结果的解释，就能在使用手册的帮助下定量分析数据。因此，对于非统计工作者是很好的选择。

实例分析：基于 TM 数据的雅鲁藏布江源区草地植被盖度估测（孙明等，2012）。

采用 Landsat-5 TM 数据，以其派生数据 NDVI、RVI、VI3、PVI、DVI、MSAVI、SAVI、TM4/TM5 为主要分析因子，结合野外植被样地调查数据，选取相关性最高的因子与植被盖度实测值，通过 SPSS 建立回归模型，然后利用该模型反演源区的植被盖度。

将草地盖度实测数据与对应样地的各种植被指数及 TM4/TM5 指数进行相关性分析，结果如表 3-3 所示。

表 3-3　草地盖度实测数据与植被指数的相关性

植被指数	相关系数	植被指数	相关系数
NDVI	0.480 208	SAVI	0.472 967
RVI	0.560 418	MSAVI	0.404 473
DVI	0.454 968	VI3	0.716 828
PVI	0.454 951	TM4/TM5	0.723 452

由表 3-3 可见，不同指数相关性排序为：TM4/TM5＞VI3＞RVI＞NDVI＞SAVI＞DVI＞PVI＞MSAVI。将实测值与 TM4/TM5 值的变化趋势绘制成曲线。通过对比发现，二者之间的变化趋势基本同步（图 3-4）。

用 TM4/TM5 作为自变量，植被盖度作为因变量，将样地数据分为 2 组，第一组样地数据（41 个）参与分析，第二组（10 个）进行精度检验；然后利用第一组样地数据

通过 SPSS 构建回归模型，对草地植被盖度与 TM4/TM5 的关系进行回归分析，拟合结果如表 3-4 所示。

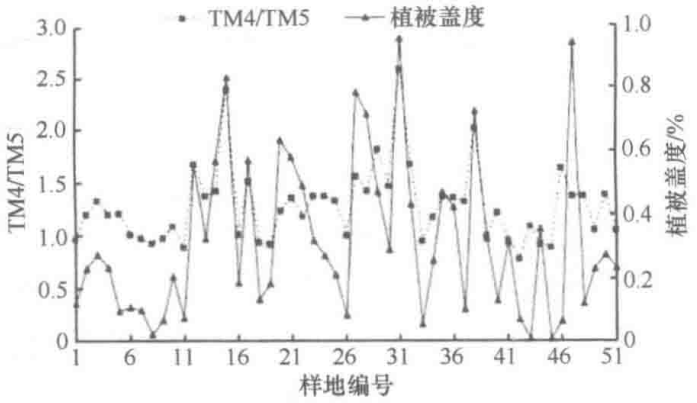


图 3-4　草地植被盖度与 TM4/TM5 曲线

表 3-4　草地植被盖度与 TM4/TM5 的线性及非线性拟合模型参数表

模型	模型概述					参数估计			
	相关系数 R^2	F 检验	第一自由度	第二自由度	显著性检验	常量	变量 1	变量 2	变量 3
线性模型	0.409	27.003	1	39	0	–0.315	0.509		
对数模型	0.396	25.586	1	39	0	0.185	0.701		
抛物线模型	0.409	13.164	2	38	0	–0.351	0.56	–0.017	
一元三次函数	0.411	8.593	3	37	0	0.051	–0.294	0.549	–0.116
幂函数	0.358	21.736	1	39	0	0.12	2.778		
指数模型	0.323	18.584	1	39	0	0.019	1.886		

从表 3-4 可以看出，草地植被盖度和 TM4/TM5 评价指标之间有 6 个回归模型，回归模型中，R^2 最高的是一元三次函数模型，其次为抛物线模型，但这 2 个模型的 F 检验值很低。就 TM4/TM5 与草地植被盖度的相关分析而言，最适合用于草地植被盖度拟合反演的是线性模型，不但 R^2 值较高（0.409），且 F 检验值也最大（27.003）。

根据以上分析可见，以 TM4/TM5 为自变量的线性函数拟合草地植被盖度变化的效果比较好，利用 SPSS 统计软件进行线性回归分析，得出线性回归模型

$$y = 0.509x - 0.315$$

式中，y 为草地植被盖度，x 为 TM4 /TM5，相关系数 $R^2 = 0.64$。

（四）统计分析软件 MATLAB（MATrix LABoratory 矩阵实验室）的数据分析方法

MATLAB 是由美国 Mathworks 公司发布的主要面对科学计算、可视化及交互式程序设计的高科技计算环境。它将数值分析、矩阵计算、科学数据可视化及非线性动态系统的建模和仿真等诸多强大功能集成在一个易于使用的视窗环境中，为科学研究、工程设计及必须进行有效数值计算的众多科学领域提供了一种全面的解决方案（谢中华等，2012）。

MATLAB 的基本数据单位是矩阵，它的指令表达式与数学、工程中常用的形式十分相似，故用 MATLAB 来解算问题要比用 C、Fortran 等语言完成相同的事情简捷得多。

在新的版本中也加入了对 C，Fortran，C++，Java 的支持。在草业科学研究的许多实际问题中不但需要对数据进行统计分析，而且希望能建立合适的模型进行模拟，MATLAB 在应用方面具有操作简单、接口方便、扩充能力强等优势，这是其他软件不可比拟的地方。因此，MATLAB 在统计应用上的地位显得越来越重要。

（五）统计分析软件 R 的数据分析方法

R 是一套完整的数据处理、计算和制图软件系统。其功能包括数据存储和处理系统，数组运算工具（其向量、矩阵运算方面功能尤其强大），完整连贯的统计分析工具，优秀的统计制图功能，简便而强大的编程语言，可操纵数据的输入和输出，可实现分支、循环，用户可自定义功能。R 不但提供若干统计程序，而且操作者只需指定数据库和若干参数便可进行一个统计分析。R 可以提供一些集成的统计工具，但更重要的是它能提供各种数学计算、统计计算的函数，从而使使用者能灵活机动地进行数据分析，甚至创造出符合需要的新的统计计算方法。因此，可将 R 软件的优点归结为免费、能应用最新的算法软件包、因编程而易于与同行交流。而其最大的弱点在于可信度没有收费软件 SAS、SPSS 等有保障，所以使用起来还是有一定的局限（薛毅和陈立萍，2007）。

第二节　草业信息分析处理的地统计学方法

地统计学（Geostatistics），又称为地质统计学，是一门以空间或者时空数据集为研究对象的统计学分支。设计之初是为了采矿业矿石等级的分布预测，目前已广泛应用于石油地质、水文地质、水文、气象、海洋、地球化学、地理、林业、环境控制、景观生态学、土壤学、农业及草业科学。

在草业地理信息中地统计学分析主要应用于空间插值的计算，利用空间上测量的离散数据计算曲面上（空间）的连续数据，以监测草地植被及其生长等相关参数在不同空间尺度上的分布和动态变化过程。因此，草业地理信息中的空间插值主要包括已知的离散点数据和用来计算未知曲面连续数据的插值方法。已知的离散点数据即已知点，又称采样点、控制点，为空间差值过程提供了必要的数据基础。控制点的数量、性质、分布特点、趋势等直接决定了插值所适用的方法、插值过程的复杂度和最终结果的精确度。分析控制点数据是插值前的方法预判的重要步骤。

空间插值的方法有很多种：①精确和近似插值，依据用于插值的控制点是否经过最终插值曲面，空间插值分为精确插值和近似插值（Chang，2010；Lam，1983）。精确插值结果中控制点位置上的插值估计值与采样值结果相同，而近似插值中插值结果的估计值替代了原来的采样值，对原始数值进行了误差平滑（图 3-5）。②确定性和随机性插值，可根据插值方法中除数学函数外是否引入了统计学方法，空间插值分为确定性插值法和可提供预测误差（统计学模型）的随机性插值法。③整体插值和局部插值，根据其利用控制点选取范围的不同而划分为整体空间插值和局部空间插值两类。下面就整体插值和局部插值分类法具体介绍草业工作中常用的几种插值方法（表 3-5）。

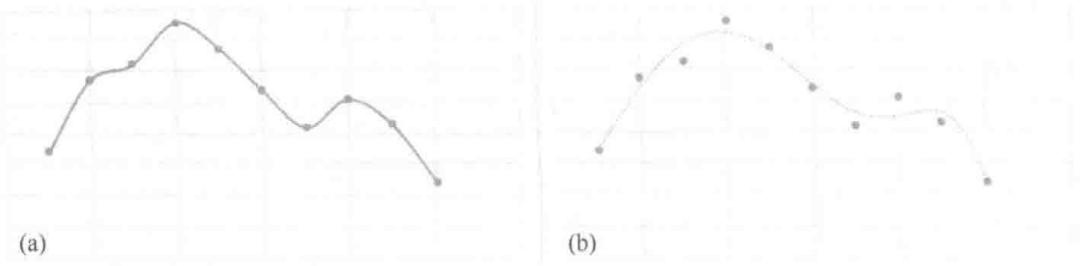


图 3-5　精确插值（a）和近似插值（b）

表 3-5　草业地信中常见几种空间插值方法分类表

	整体空间插值	局部空间差值
近似插值	全局多项式插值	—
精确插值	—	反距离加权插值法
	—	样条回归插值法
	—	泰森多项式插值法
	—	克里格插值法

一、整体空间插值分析方法

整体空间插值法是利用已知所有的样点值预测估算空间曲面上的未知数据，其中最主要的方法为全局多项式插值法。

全局多项式插值法：也称为趋势面分析法，是结合已有数据用数学多项式的方法拟合一个光滑的预测趋势面（近似插值），以获取未知点数据的非精确插值方法，也是利用数学曲面模拟地理系统要素在空间上的分布及变化趋势的一种数学方法，实质上是通过回归分析原理，运用最小二乘法拟合一个二元非线性函数，模拟地理要素在空间上的分布规律，展示地理要素在地域空间上的变化趋势。它是根据采样点的地理坐标 X，Y 值与样点的属性 Z 值建立多元回归模型，前提假设是 Z 值是独立变量且呈正态分布，其回归误差与位置无关。根据自行设置的参数可建立线性、二次……或 n 次多项式回归模型，从而得到不同的拟合平面，可以是平面，也可以是曲面。精度以最小二乘法进行验证。它的缺点在于：当研究区域范围较大、地形很复杂时，需要用高阶多项式拟合以提高精度，但高阶将增加其计算成本。

以给定 N 个采样点为例，多项式公式如下（Lam，1983）

$$f(x,y)=\sum_{i,j=0}^{N} a_{ij}x^i y^j$$

式中，系数 a_{ij} 由方程组决定

$$f(x_i,y_i)=z_i, i=1,\cdots,N$$

随着多项式阶数的增加，可以拟合逼近更加复杂的趋势面数据，阶数越高曲面越起伏多变，系数则决定了曲面的弯曲程度。但其所能代表的空间物理意义越来越不显著且计算量庞大，一般来说用来拟合地貌特征只需要 3～5 阶的回归模型即可。因此全局多

项式插值法适用于模拟变化趋势缓慢、物理特征明显的空间变化趋势面。整体空间插值方法适用于研究整个插值区域的趋势特征。

二、局部空间插值分析方法

局部插值法用相邻的一组样本来估算未知值，又称为局部多项式插值（Local Polynomial Interpolation），通常采用多个多项式对未知值拟合。每个多项式都只在特定重叠的邻近区域内有效，通过设定搜索半径和方向来定义邻近区域。显然，局部多项式插值是对全局多项式，即趋势面拟合的一大改进。这里涉及一个搜索邻域的概念。从空间自相关性的概念可知，空间上越靠近，属性就越相似，相关性也越高。那么，两个样点间在多远的距离内所具备的相关性可以不考虑，或者其相关将消失呢？可以根据经验或专业背景找出这么一个阈值，作为邻近区域的半径。同时，如果其自相关性在不同的方向上消失的距离值也不同的话，将还需要设置一个方向值及长短两个半径值，此时的邻近区域将呈椭圆。如当属性值受风向影响较大时，应当将风向角度设置为搜索方向，即长半径所在的方向。通过半径和方向可以定义出一个以待估点为中心的区域（圆或者椭圆）。此外，还可以通过限制参与预测某待估点值的样点数来定义邻近区域，即参与某点预测的最多样点数和最少样点数。在由半径和方向决定的区域内包含到的样点数为0时，则扩大搜索区域使其达到最小样点数值。局部空间插值法主要包括反距离加权插值法、样条回归插值法、克里格插值法等（Béla，2010），均可在 ArcGIS 等软件中实现。

（一）移动平均插值法（Moving Average）

通过设定邻近区域，取该区域内样点的平均值作为待估点值。适用于样点分布均匀、密集，而且变化缓慢的情况下，对缺失值进行填补。主要用于消除随机干扰，即局部降噪功能。优势在于计算简便快速，但适用范围较窄。

（二）线性三角网法（Triangulation with Linear Interpolation）

是最佳的德劳内三角形，连续样点数据间的连线形成三角形，覆盖整个研究区域，所有三角形的边都不相交（即与构建 TIN 文件的原理一致）。线性三角网法将在整个研究区域内均匀分配数据，地图上的稀疏区域会形成截然不同的三角面。

（三）最小曲率法（Minimum Curvature）

非精确插值法。其插值基准是生成一个具有最小曲率（即弯曲度最小），且到各样点的 Z 值距离最小的曲面。影响最小曲率插值法精度的参数有：最大残差，通常允许残差在 1%～10%；最大循环次数，与栅格大小（Cell Size）有关，通常设置为生成的栅格数量的 1～2 倍。

（四）径向基函数插值法（Radial Basis Function）

所谓径向基函数，即基函数是由单个变量的函数构成的，是一系列精确插值法的统称。该插值法中的单个变量是指待估点到样点间的距离 H，其中每一插值法都是距离 H

的基函数。径向基函数是对最小曲率插值的改进，即属于精确的最小曲率插值法。径向基函数包括多种函数，主要有倒转复二次函数（Inverse Multi-quadric）、复对数（Multi-log）、复二次函数（Multi-quadric）、自然三次样条函数（Natural Cubic Spline）和薄板样条法函数（Thin Plate Spline）。径向基函数比同为精确插值法 IDW 的优点在于，它可以计算出高于或低于样点 Z 值的预测值。通常俗称的样条插值法即径向基函数插值法。在实际应用中，也发展出多种样条插值法，包括 GRASS 软件的 RST（Regularized Spline with Tension）和 ANUSPLINE 的薄盘光滑样条插值法，从而大大提升了样条插值的精度，如对气象要素进行插值时可以使用该方法，综合考虑多个协变量，减少结果的不确定性。径向基函数适用于样点数据集大、表面变化平缓的情况；当局部变异性大，且无法确定样点数据的准确性或样点数据具有很大不确定性时，不适用该技术。

（五）反距离加权插值法（Inverse-distance Weighted Method，IDW）

反距离加权插值法是一种普遍适用的插值方法（Lam，1983）。未知的插值点与样点距离越近，其相似性就越高，则赋予越高的权重，反之亦然。这种方法并不是依据控制点在局部区域内建立新的曲面而是在控制点取得极值，因此它也属于精确插值的一种。曲面上任一点和控制点之间都存在一定的联系，但距离越近的影响力越大。未知点的属性计算公式如下

$$P_i = \frac{\sum_{j=1}^{G} \frac{P_j}{D_{ij}^n}}{\sum_{j=1}^{G} \frac{1}{D_{ij}^n}}$$

式中，P_i 是所求的 i 点属性值，P_j 是样点 j 的属性值，D_{ij} 是 i 点到 j 点的距离，G 是控制点数目，n 是确定的幂。实际上 n 值控制着每个样点所影响的区域范围，n 值增加，影响范围则趋于无限减小，接近到控制点 i 点；n 值减小至无限接近 0，则等同于简单的样点值平均方法（Watson and Philip，1985），所以每个未知点都有一个取值范围，介于某个已知最大和最小值之间。

案例分析：Messina 等（2011）通过反距离加权插值法的应用，研究了疟疾在刚果共和国在空间尺度的暴发密度、传播途径等，选取了两大类因子，即个人因子和社会因子。个人因子包括年龄、性别、抗体、受教育程度、蚊帐拥有度等，社会因子包括水源距离、就医时间、距市区距离、人口密度、气候特征等。最终针对空间上和因子之间所呈现的传播特征，对疟疾在该国的传播提出了更加高效的预防措施（图 3-6）。

（六）样条回归插值法（Spline Regression Interpolation Method）

样条回归分析法是利用最小化表面总曲率的数学函数来估计值生成恰好经过控制点的平滑表面，也称最小曲率法。样条回归的线段由多个回归线片段在节点处连接而成，虽然没有突出的转折点，但线条仍可在节点处改变方向。样条回归法的发展避免了多项式回归中二阶、三阶或更高阶的多项式，使得其更加灵活且更容易产生多重共线性。

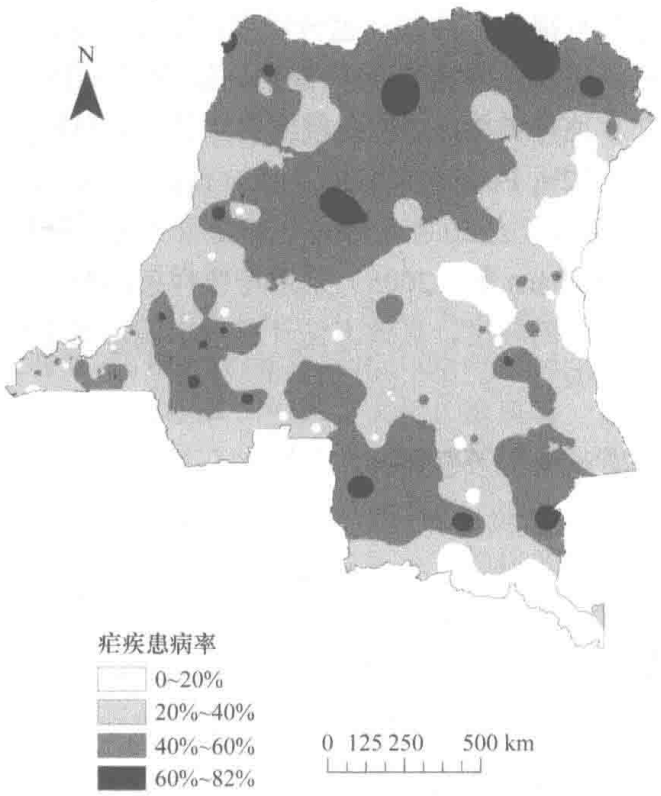


图 3-6　反距离加权插值法计算的疟疾在刚果共和国传播率空间分布图（Messina et al.，2011）

最常见的一种基本样条插值法称为薄板样条插值法（Thin-plate Splines），这个名字来源于该函数微分方程所表述的弯曲性是无限的，仅受限制于控制点的载荷力（Wabba，1981）。作为一种拟合模型，薄板样条插值在已知误差和排除小尺度特性的情况下具有很大的优势，公式为

$$Q(x,y)\sum A_i d_i^2 \log d_i + a + bx + cy$$

式中，Q 是未知点，x、y 是 Q 点的坐标值，$d_i^2 = (x - x_i)^2 + (y - y_i)^2$，$x_i$ 和 y_i 分别是控制点 i 的坐标，$a+bx+cy$ 为一阶线性趋势面，$d_i^2 \lg d_i$ 是生成最小化表面曲率的函数。系数 A_i、a、b、c 由以下线性方程组计算得出

$$\sum_{i=1}^{n} A_i d_i^2 \lg d_i + a + bx + cy = f_i$$

$$\sum_{i=1}^{n} A_i = 0$$

$$\sum_{i=1}^{n} A_i x_i = 0$$

$$\sum_{i=1}^{n} A_i y_i = 0$$

式中，n 为控制点数目，f_i 为控制点 i 的属性值，系数的计算要求解 n+3 个方程的方程组（Franke，1982；Chang，2010）。

样条插值法目前主要应用于估算（Hutchinson，1991）和表述（Wabba，1981）气

象因素，优势在于即使只有很少的控制点也可以生成精度较好的曲面数据。

案例分析：梁天刚、林慧龙等在 2012～2013 年，结合 GIS 技术依据草原综合顺序分类法制作了全国潜在草地分类图，并结合模型估算了相应草地植被潜在的初级净生产力（Liang et al.，2012a，2012b；Feng et al.，2013；Lin et al.，2013a，2013b；Lin and Zhang，2013）。研究过程中所用到的过去 50 年（1950～2000 年）月空间气象数据（降水、气温）都来自加利福尼亚大学 Hijmans 等（2005）采用薄板样条回归分析方法的插值结果，多种内插值的比较表现出良好的精度，且该方法计算效率高、易于操作，已经广泛应用于全球尺度的各类研究工作中。

（七）泰森多边形（Thiessen 或 Voronoi 多边形）

泰森多边形也称狄利克雷棋盘花纹或沃罗诺伊图，又称最近邻点插值法（Nearest Neighbor），属于精确差值的一种，未知点的属性值等于最邻近控制点值的插值。这种方法常用于气象分析，当某些地区缺乏气象台站数据时则近似使用最近台站的数据信息。即定义为每个独立的多边形是距该多边形内控制点最近的点的集合。

构建泰森多边形，首先基于控制点的邻近点连接所建立的不规则三角网格，即德劳内三角网，使每个控制点都成为三角形的顶点。

德劳内三角网的基本规则：①任何一个三角形的外接圆内不能包含其他任何控制点；②最小角最大准则，两个相邻的三角形构成凸四边形，在交换该四边形的对角线之后，6 个内角的最小值不再增加；③对这些三角网格的各个边作垂直平分线，用每个控制点周围的若干条垂直平分线连接围成一个多边形，最终形成泰森多边形（图 3-7）。

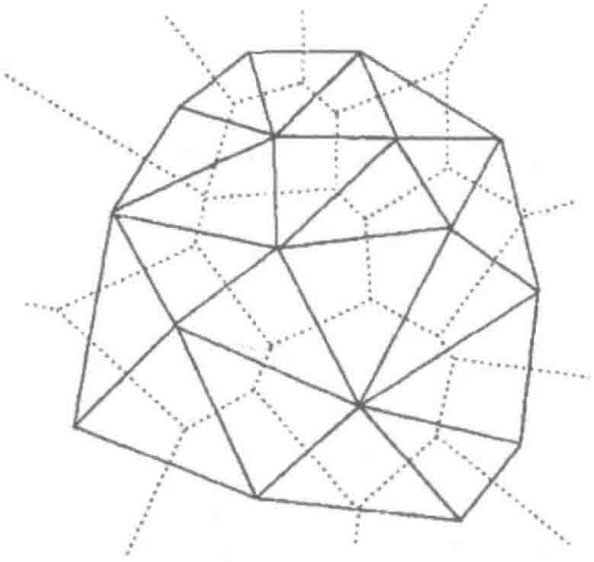


图 3-7　德劳内三角网（实线）和泰森多边形（虚线）

（八）自然邻近插值法（Natural Neighbor）

自然邻近插值法是对泰森多边形插值法的改进。它对研究区域内各点都赋予一个权重系数，插值时使用邻点的权重平均值决定待估点的权重。每完成一次估值就将新值纳入原样点数据集，重新计算泰森多边形并重新赋权重，再对下一待估点进行估值运算。

对于由样点数据生成栅格数据而言，通过设置栅格大小（cell size）来决定自然邻近插值中的泰森多边形的运行次数 n，即设整个研究区域的面积 area，则有 n=area/cell size。也可设置各向异性参数（半径和方向）来辅助权重系数的计算。

三、克里格插值分析方法

克里格插值法（Kriging）是由南非矿产学家 D. R. Krige 用于预测空间过程的地统计学方法，后由法国统计学家 G. Matheron 系统化（Oliver and Webster，1990）。目前已广泛地应用于各个学科中，如地球物理、生态、大气、地质等。克里格法的数学基础是变异函数理论，能够在有限区域内对区域化变量取值进行无偏最优化估计。一般当变异函数和相关分析的结果表明区域化变量存在空间相关性，可使用克里格法对空间未知点或区域进行估值计算。其实质是利用区域化变量的原始数据和变异函数的结构特点，对未知点的区域化变量的属性值进行最优线性、无偏估计。这种方法的优势在于不仅考虑待插值点与邻近控制点数据的空间位置，还考虑了各邻近点之间的位置关系，利用了控制点空间分布的结构特点，使其估计结果更加精确、符合实际，并可以有效避免系统误差产生的“屏蔽效应”（徐建华，2002；刘承香等，2004）。

克里格插值法是一个最优的无偏估计法。获得预测图并不要求数据呈正态分布，但当数据呈正态分布时，克里格插值法将是无偏估计法中效果最好的一种方法之一。因此，在进行克里格插值前，可先对非正态分布的数据进行转换，包括 Log 对数转换、Box-Cox 转换，使之呈正态分布，然后再进行插值。

（一）克里格法的分类

一般来说，克里格插值法包括了一系列的局部插值方法，如简单克里格、普通克里格、协同克里格（Co-Kriging）、泛克里格和析取克里格。

简单克里格假设，数据集的平均值是已知的，但在实际中均值一般很难得到。协同克里格是对普通克里格在逻辑上的发展，两个或两个以上的变量在空间上相互依存，把某一控制点上某一变量的观测值与在统计分析上依赖于相邻控制点上的另一变量的观测值，建立这两种变量之间的相关性，用一种变量的信息去弥补所遗漏或提供另一变量的信息。泛克里格是指空间变量还同时受到漂移和随机的影响，将空间过程划分为趋势项和残差项两个部分。而析取克里格则是一种非线性的插值方法，它在进行预测的同时提供了超过或不超过某一阈值的概率，适用于辅助决策（Oliver and Webster，1990；Chang，2010）。

根据样点数据统计特征的不同，可将克里格分成多种不同的插值方法。当样点数据是二进制值时，用指示克里格插值法进行概率预测；对样点数据进行了未知函数变换后，可用该变换函数进行析取克里格插值；当样点数据的趋势值 $\mu(s)$是一个未知常量时，用普通克里格插值；当样点数据的趋势可用一个多项式进行拟合，但回归系数未知时，用泛克里格插值法；当样点数据的趋势已知时，用简单克里格插值法；其中最常用的是普通克里格与泛克里格插值法；当加入了协变量进行插值时，则称为协同普通克里格插值

法和协同泛克里格插值法。

本节以普通克里格为例介绍克里格的基本数学原理。

（二）普通克里格

1. 变异函数

又称变差函数、变异矩阵，是区域化变量增量的方差（徐建华，2002）。当空间一点 x 在单一轴上变化时，区域化变量 $Z(x)$在点 x 和 $x+h$ 处的值 $Z(x)$与 $Z(x+h)$差的一半为区域化变量 $Z(x)$在 x 轴方向上的变异函数，$r(h)$表达式

$$r(x,h)=\frac{1}{2}\mathrm{Var}\left[Z(x)-Z(x+h)\right]$$

$$=\frac{1}{2}E\left[Z(x)-Z(x+h)\right]^2-\frac{1}{2}\{E\left[Z(x)-Z(x+h)\right]\}^2$$

二阶平稳假设满足

$$E\left[Z(x+h)\right]=E\left[Z(x)\right]$$

以上式子可改写为

$$r(x,h)=\frac{1}{2}E\left[Z(x)-Z(x+h)\right]^2$$

当 r 与 x 位置无关时，$r(x, h)$可改写为 $r(h)$，则有

$$r(h)=\frac{1}{2}E\left[Z(x)-Z(x+h)\right]^2$$

此时，$r(h)$称为半变异函数，$2r(h)$为变异函数。

2. 普通克里格

普通克里格中，不存在漂移，用半变异函数直接插值。估算某未知点 Z_0 值的方程为

$$Z_0=\sum_{i=1}^{n}\lambda_i Z_i$$

式中，Z_0 为未知点值，Z_i 为控制点 i 的已知值；λ_i 为第 i 个位置处的测量值的权重；n 为控制点数目。权重可由一组方程组求得。例如，由已知的三个控制点（1，2，3）估算一个未知点（0），则有以下方程组

$$\lambda_{1r}(h_{11})+\lambda_{2r}(h_{12})+\lambda_{3r}(h_{13})+v=r(h_{10})$$

$$\lambda_{1r}(h_{21})+\lambda_{2r}(h_{22})+\lambda_{3r}(h_{23})+v=r(h_{20})$$

$$\lambda_{1r}(h_{31})+\lambda_{2r}(h_{32})+\lambda_{3r}(h_{33})+v=r(h_{30})$$

$$\lambda_1+\lambda_2+\lambda_3+0=1.0$$

式中，$r(h_{ij})$为控制点 i 和 j 间的半变异函数，$r(h_{i0})$是第 i 控制点和未知点之间的半变异函数，v 是确保估算误差最小的拉格朗日系数。解出方程组后可算出未知点值

$$Z_0=Z_1\lambda_1+Z_2\lambda_2+Z_3\lambda_3$$

由该方程可看出，克里格插值中求未知值时不仅用到了位置点与控制点之间的半变异函数，而且还用到了控制点之间的半变异函数，使其与反距离权重法区别开来。通过

对每个未知点进行了变异量算来说明其可靠性，是它与其他局部插值法相比的重要优势（Oliver and Webster，1990；Chang，2010）。

案例分析：高远等（2009）利用普通克里格插值法对天津南郊地区的土壤污染物，主要指砷（As）进行空间分布研究。文章中首先对研究区内网格布点法下的163个实测样点值进行统计分析，了解数据的集中分散性分布特征，而后进行空间结构性分析，主要目的是筛选异常值。对排除异常值后的数据计算实验半变异函数，再通过拟合实验半变异函数获得半变异函数。最终通过交叉验证后确定了模型参数，结果表明各向异性的球形模型效果最好，取得了精度较高的插值结果，达到了预期的研究目的。

第三节　草业信息分析处理的多目标规划方法

在过去的长期理论与实践研究中，系统优化积累了大量的成功经验和方法。优化模型大体可分为数学规划模型和非数学规划模型两大类，其中应用最广泛的是基于数学规划技术的优化模型，如线性规划、非线性规划、动态规划和整数规划模型等。非数学规划模型大多数是基于经验和观察所总结的经验性方法。

本节不可能对数学规划的每一种方法都给予十分详尽的阐述，作为方法论上的提示，只能摘其要者概述。

一、数学规划的概念

数学规划，即在一些等式或不等式的约束下，求一个目标函数的极大（或极小）的优化模型。依据有无约束条件，数学规划可分为**约束数学规划**和**无约束数学规划**。约束数学规划的一般形式（施光燕等，2007）为

$$\begin{aligned} & \max f(x)[\text{或}\min f(x)] \\ s.t. \quad & h_i(x)=0,\ i=1,2,\cdots,l \\ & g_j(x)\geqslant 0,\ j=1,2,\cdots,m \end{aligned} \tag{3-1}$$

式中，x 为向量，即 $x=(x_1, x_2, \cdots, x_n)^T$。

由于系统优化总处于非平衡状态下，应认识到限制与干扰在量上具不等同性，其作用和价值会因时、因地而异。

在数学规划的讨论中，把满足所有约束条件的点 x[即满足 $h_i(x)=0, i=1,2,\cdots,l$ 和 $g_j(x)\geqslant 0, j=1,2,\cdots,m$]称为**可行点**（或**可行解**），所有可行点组成的点集称为**可行域**。

若把**可行域**记为 S，即

$$S=\{x \mid h_i(x)=0, i=1,2,\cdots,l, g_j(x)\geqslant 0, j=1,2,\cdots,m\}$$

数学规划便是求 $x^*\in S$，使得 $f(x^*)$在 S 上达到最大值（或最小值），称求得的 x^*为**最优点**（**最优解**），相应的 $f(x^*)$为**最优值**。

在数学规划中，所研究的问题只含有一个目标函数称为**单目标规划**。但是，在草业现实的生产实际中，往往需要同时考虑多个目标在某种意义下的最优化问题，我们称这

种含有多个目标的最优化问题为**多目标规划**。

特别需要说明的是任何系统优化都会包含时间尺度，并且在时间序列基础上，动态地改变着优化条件与优化内容，甚至于优化的目标。

二、线性规划问题的数学形式

在数学规划中，若目标函数和约束条件中的函数均为线性函数，则称为**线性规划**，否则就称为**非线性规划**。线性规划是研究多变量函数在变量受约束条件下的最优化问题，是运筹学中重要的分支之一。

第二次世界大战期间，由于军事、生产组织、物质调运等方面的迫切需要，以线性规划的形式提出的问题越来越多，这大大促进了运筹学的应用和发展。但当时，对线性规划问题还没有完善的解法。直到 1947 年单纯形法的提出，才为线性规划的发展奠定了基础。J. Von Neumann 和 O. Morgenstern（1947）提出了对偶性概念，进一步发展了单纯形法，提高了线性规划解决问题的能力，开拓了更广阔的应用领域。现在，无论从理论观点还是从应用和计算观点来看，线性规划都是发展得最成熟的优化技术方法。

求一组变量 X_j（j=1, 2, …, n）满足条件

$$\begin{cases} \sum^{n} a_{ij} X_{j} \geqslant b_i,\ b_i \geqslant 0, i = 1,2,\cdots,m \\ X_j \geqslant 0,\ \ j = 1,2,\cdots,n \end{cases} \tag{3-2}$$

$$\text{使函数 } f = \sum^{n} C_j X_j = \min \text{ （或 max）} \tag{3-3}$$

式中，X_j 为决策变量，农业系统中也叫活动方式，a_{ij} 为约束条件中的决策变量系数，b_i 为资源限制量，C_j 为目标函数中的决策变量系数，f 为决策目标。

同上，称式（3-1）为**约束条件**或**约束方程**，称式（3-3）为**目标函数**。把式（3-2）的任一组解称为线性规划问题的一个**可行解**，可行解的全体称为该问题的**可行解集合（可行域）**。使目标函数取得极值的可行解称为**最优解**。

求解线性规划模型的算法很多，除了单纯形算法外，还有两阶段法、大 M 法和 Karnarkar 法等，但单纯形法是应用最广的一种通用算法，经过多次迭代，求出线性方程组的最优解。

线性规划的具体求解过程及原理不是本节的重点，使用者可调用 LINDO System 公司开发的解线性规划的数学软件 LINDO，其使用方法请登录网站 http://www.lindo.com 进一步学习（任继周，2004）。

需要补充说明的是，我们所建的优化模型必须能比较真实精确地反映系统的实际状态，才能使所得到的解越来越接近最优解。草地农业信息系统的问题在一定程度上都是非线性的，完全靠线性的模型无法准确反映问题的实质，因此非线性规划模型正日益受到广泛的重视。

三、非线性规划问题的数学形式

非线性规划是研究在一组线性与（或）非线性约束条件下，寻求某个非线性或线性

目标函数的最大值或最小值问题。非线性规划问题通常可用数学模型表示为

目标函数 max（或 min）$f(x)$

约束条件

$$\begin{cases} h_i(x)=0,\ i=1,2,\cdots,m \\ g_j(x)\geqslant 0(\text{或}\leqslant 0),\ j=1,2,\cdots,n \end{cases}$$

$$x=(x_1,x_2,\cdots,x_n)^T\in E^n$$

在优化设计研究中，非线性规划模型应用较多，且有许多求解算法，如拟线性规划法、拉格朗日乘子法、梯度法（微分法）、广义简约梯度法、罚函数法及各种改进或组合算法等，在非线性优化模型的求解中都有不同程度的应用。有许多非线性规划模型的求解都是通过增加一些改进策略或措施，以求解线性规划模型为基础寻找有效的求解途径。

非线性规划优化方法是一类系统优化方法，已经被有力地应用于生态系统优化之中，这是近 30 年来在逐步发展起来的一类方法。由于它们在数学解析中的巨大障碍，因此仍处于探索之中。因为在生态系统中非线性问题的普遍存在，加上其他领域中对于非线性研究的巨大关注，因此这一类方法必将有更大的发展。在具体求解过程中也可调用 LINDO System 公司开发的解非线性规划的数学软件 LINDO，其使用方法请登录网站 http://www.lindo.com 进一步了解。

四、多目标规划（Multi-objective Programming，MOP）的数学形式

多目标优化问题最早由意大利经济学家 V. Pareto 提出，他把许多本质上不可比较的目标化成单目标寻优。此后，Kuhn 和 Tucker（1950）及 Zadeh（1963）等曾从不同角度讨论过多目标问题，并给出了一些基本概念和基本定理。到 20 世纪 70 年代，多目标规划的研究得到越来越多的重视。首先，由于许多实际问题本来就是多目标的，如果简化成单目标求解，可能无法得到令人满意的结果；其次，由于近年来计算机技术有了很大的发展，使得求解多目标规划问题成为可能（严祖梁，1988）。

多目标规划就是确定一组变量 $X=(x_1,x_2,x_3,\cdots,x_n)^T$，使 $p(p\geqslant 2)$ 个目标函数 $f_1(x),f_2(x),\cdots,f_n(x)$ 达到最优，即 $\max f(x)=[f_1(x),f_2(x),\cdots,f_n(x)],x\in X$。

由于存在绝对最优解不存在的情况，因此对于多目标规划还要引进一些解的概念。对于多目标规划：

$$(\text{VP})\qquad \min_{x\in D}[f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]^T$$

定义 1　对于可行点 $\boldsymbol{x}_0\in D$，若存在另一可行点 $\bar{\boldsymbol{x}}\in D$，使 $f_k(\bar{\boldsymbol{x}})<f_k(x_0)$，$k$=1，2，…，$p$，则称可行点 x_0 为（VP）的**劣解**。

定义 2　对于可行点 $\boldsymbol{x}_0\in D$，若不存在另一可行点 $\bar{\boldsymbol{x}}\in D$，使 $f_k(\bar{\boldsymbol{x}})\leqslant f_k(x_0)$，$k$=1，2，…，$p$，则称可行点 x_0 为（VP）的**有效解**。

定义 3　对于可行点 $\boldsymbol{x}_0\in D$，若不存在另一可行点 $\bar{\boldsymbol{x}}\in D$，使 $f_k(\bar{\boldsymbol{x}})<f_k(x_0)$，$k$=1，2，…，$p$，则称可行点 $\boldsymbol{x}_0$ 为（VP）的**弱有效解**。

当我们求解多目标规划时，最后得到的最优解应该是有效解，至少应该是一个弱有效解。

（一）多目标规划中不确定性问题的处理

处理多目标规划中的不确定性问题有多种方法，归纳起来主要分为以下几种（邹锐等，1999）。

1. 随机多目标规划（Stochastic Multi-objective Programming，SMOP）

较确定性规划而言，由于随机变量的介入，导致其复杂性大大增加。Stancu-Minasian编著的《随机多目标规划》一书中，对随机多目标规划的问题进行了详细的论述，为这一理论的发展奠定了基础。目前求解随机多目标规划常用思路有两种：一是利用问题中随机变量的某种（或几种）概率特性将随机多目标规划问题转化为确定性多目标规划问题；二是将随机多目标规划问题转化为单目标随机规划问题进行求解。但在实际问题中，有可能无法得出确定性规划的具体形式，或确定性规划的形式变得相当复杂，甚至以现有的数学手段难以求解。

2. 模糊多目标规划（Fuzzy Multi-objective Programming，FMOP）

L. A. Zadeh（1965）开创性地提出“模糊集合”，模糊数学逐步发展，并渗透到自然科学、社会科学、工程技术等众多领域。而 Sakawa 和 Yano（1989）提出的模糊多目标规划模型，较 SMOP 而言，不需要太多的数据以确定参数分布，因此在其应用上的限制要宽一些。但即便如此，在实际研究中要获得所有系统组分的隶属度信息也会遇到很大困难。

3. 不确定性多目标规划（Inexact Multi-objective Programming，IMOP）

为避免 SMOP 和 FMOP 方法缺陷，一条有效途径就是将不确定性等概念引入多目标规划的框架中，从而导出不确定性多目标规划（IMOP）的方法。IMOP 作为系统分析与系统优化的有力工具，能较好反映复杂系统的综合性、动态性、多目标性和不确定性特征，并科学高效地产生可操作的规划方案。该方法直接在普通 MOP 模型中引入代表不确定性信息的区间数，而无需考虑参数的概率分布信息或模糊隶属度信息，因此在数据获取、算法实现、计算要求及结果解译上比原有的其他方法都存在较为显著的优越性。IMOP 自提出以来，在国内外已被成功地应用于流域土地规划、固体废弃物管理规划、经济开发区规划和城市旅游圈规划等领域（邹锐等，1999；王丽婧等，2005）。

（二）利用 MATLAB 求解多目标规划问题举例

1. MATLAB 优化工具箱简介

MATLAB 有一套程序扩展系统和一组称为工具箱的特殊应用子程序。工具箱是MATLAB 函数的综合程序库，它们是为某类学科专业和应用定制的 MATLAB 运行环境。用户应用 MATLAB 工具箱可以快速获得准确答案，因为每个工具箱都建立在可靠的数值算法基础之上。所有工具箱对运行 MATLAB 的各种计算机平台兼容，而且各个工具

箱可以相互调用。用户还能够进入工具箱源代码修改、定制、扩展算法和工具箱的功能以适应个人需要（施光燕等，2007）。

MATLAB 现有 30 多个工具箱，优化工具箱（Optimization Toolbox）就是其中应用较为广泛、影响较大的一个。它的主要功能有求解线性规划和二次规划、求函数的最大值和最小值、多目标优化、约束条件下的优化和求解非线性方程等（表 3-6）。

表 3-6　MATLAB 优化工具箱常用函数

求解问题	函数	具体含义	语法
线性规划最优解	linprog	$\min c^T X$ $s.t.\ AX \leqslant b$	X = linprog（c，A，b）
无约束的一元函数最小值	fminbnd	$\min f(x)$ $s.t.\ x_1 \leqslant x \leqslant x_2$	x = fminbnd（'f'，x_1，x_2）
无约束多元函数的最小值	fminunc fminsearch	$\min F(X)$	X=fminunc（'F'，X_0） X=fminsearch（'F'，X_0）
有约束的多元函数最小值	fmincon	$\min F(X)$ $s.t.\ G(X) \leqslant 0$	X= fmincon（'FG'，X_0）
二次规划	quadprog	$\min xTHx+cTx$ $s.t.\ AX \leqslant b$	X=quadprog（H，c，A，b）
多目标优化问题	fgoalattain	$\min \gamma$ $s.t.\ F(x)-W\gamma \leqslant goal$	X=fgoalattain（'F'，x，$goal$，W）
求解极小极大问题	fminimax	$\min\{\max F(X)\}$ $s.t.\ G(X) \leqslant 0$	X=fminimax（'FG'，X_0）
求解非线性方程	fsolve	$F(X)=0$	X = fsolve（'F'，X_0）

注：以上函数的具体使用方法可查阅 MATLAB 帮助文档或访问 MATLAB 官方网站（http://www.mathworks.cn/）

2. MATLAB 求解多目标规划

问题举例：甘南牧区的草畜平衡状态与草原生态保护、牧民增收问题息息相关，草原生态保护与牧民的收益相互制约相互影响。若一味强调草原生态保护，可能无法使牧民收益有显著增长；而若不断扩大牲畜数量，虽然可能在短期内能够提高牧民收益，但可能导致超载过牧、天然草地退化等严重生态环境问题，影响牧区的可持续发展。为了研究甘南牧区草畜平衡状态，合理规划草地生产力与经济收益、生态环境影响力之间的关系，梁天刚等（2011）在构建天然草地适宜载畜量监测模型和草畜平衡监测模型的基础上，通过构建多目标优化模型的方法，以实现草畜平衡。

目前，在优化研究中一般采用的优化技术如线性规划、非线性规划、动态规划、多目标决策技术等进行实际问题的优化求解时，存在不同程度的“维数灾难”，即随着问题规模的扩大和约束条件的增加，求解时间的急剧增加而容易产生“组合爆炸”，最后导致不可解。

遗传算法和人工神经网络是两种新型的优化算法，已成功用于众多领域。遗传算法的全局寻优能力和隐含并行特性，使遗传算法特别适合于处理传统算法无法解决的复杂的非线性问题。人工神经网络的并行分布式计算结构和非线性动力学演化机制，使神经网络计算从本质上跳出了系统优化计算和数值迭代搜索算法的基本思想，为优化计算的快速实现提供了新途径。遗传算法的全局寻优能力和人工神经网络的非线性动力学特征

比传统的优化方法更适合于解决复杂的非线性问题。应用遗传算法和人工神经网络理论进行系统优化研究将是一个草地农业系统分析和设计、决策的新研究方向，尚处于探索和尝试阶段，许多问题有待于深入研究。把具有全局寻优能力的遗传算法和快速计算能力的人工神经网络有机结合起来，进行系统优化布置和优化设计研究，是改善和提高系统优化水平的一个新突破点。因此，综合应用遗传算法和人工神经网络进行系统优化研究，可以为系统优化理论和算法开辟一条新途径，同时也可以拓宽遗传算法和人工神经网络优化理论的应用领域。

思 考 题

1. 简述常用的统计分析功能有哪些？
2. 什么是局部空间插值分析方法？
3. 简述克里格插值分析方法？
4. 论述在草业科学中如何应用 MATLAB 求解多目标规划问题，可举例说明？

参 考 文 献

边金虎, 李爱农, 宋孟强, 等. 2010. MODIS 植被指数时间序列 Savitzky-Golay 滤波算法重构[J]. 遥感学报, 14(4): 725-741.

陈正昌, 程炳林, 陈新丰, 等. 2005. 多变量分析方法[M]. 北京: 中国税务出版社.

杜志渊. 2006. 常用统计分析方法——SPSS 应用[M]. 济南: 山东人民出版社.

冯琦胜, 高新华, 黄晓东, 等. 2011. 2001～2010 年青藏高原草地生长状况遥感动态监测[J]. 兰州大学学报(自然科学版), 47(4): 75-81.

龚江, 石培春, 李春燕. 2011. 巧用 Excel 解决多元非线性回归分析[J]. 农业网络信息, 01: 46-48.

贾俊平. 2008. 统计学(第三版)[M]. 北京: 中国人民大学出版社.

李传华, 赵军. 2013. 2000～2010 年石羊河流域 NPP 时空变化及驱动因子[J]. 生态学杂志, 32(3): 712-718.

梁天刚, 冯琦胜, 夏文韬, 等. 2011. 甘南牧区草畜平衡优化方案与管理决策[J]. 生态学报, 34(4): 1111-1123.

刘承香, 阮双琛, 伍小芹. 2004. 基于 Kriging 插值的数字地图生成算法研究[J]. 深圳大学学报(理工版), 21(4): 295-300.

罗纳德·科迪, 杰弗里·史密斯. 2011. SAS 应用统计分析(第 5 版)[M]. 辛涛译. 北京: 人民邮电出版社.

毛飞, 侯英雨, 唐世浩, 等. 2007. 基于近 20 年遥感数据的藏北草地分类及其动态变化[J]. 应用生态学报, 18(8): 1745-1750.

任继周. 2004. 草地农业生态系统通论[M]. 合肥: 安徽教育出版社.

施光燕, 钱伟懿, 庞丽萍. 2007. 最优化方法(第 2 版)[M]. 北京: 高等教育出版社.

孙明, 杨洋, 沈渭寿, 等. 2012. 基于 TM 数据的雅鲁藏布江源区草地植被盖度估测[J]. 国土资源遥感, 31(03): 71-77.

汪海波, 罗莉, 吴为, 等. 2013. SAS 统计分析与应用——从入门到精通(第 2 版)[M]. 北京: 人民邮电出版社.

王丽婧, 郭怀成, 王吉华, 等. 2005. 基于 IMOP 的流域环境——经济系统规划[J]. 地理学报, 60(2): 219-228.

王振龙. 2000. 时间序列分析[M]. 北京: 中国统计出版社.

谢中华, 李国栋, 刘焕进, 等. 2012. MATLAB 从零到进阶[M]. 北京: 北京航空航天大学出版社.

徐建华. 2002. 现代地理学中的数学方法: 第二版[M]. 北京: 高等教育出版社: 105-120.

薛微. 2013. SPSS 统计分析方法及应用(第 3 版)[M]. 北京: 电子工业出版社.

薛毅, 陈立萍. 2007. 统计建模与 R 软件[M]. 北京: 清华大学出版社.

严祖梁. 1988. 多目标规划方法综述[J]. 北京农业工程大学学报, 8(2): 112-120.

周鸿年, 周民. 2000. Excel 2000 中文版及其综合应用[M]. 北京: 电子工业出版社.

邹锐, 郭怀成, 刘磊. 1999. 不确定性条件下经济开发区环境规划方法与应用研究(Ⅰ)——不确定性多目标混合整数规划模型及算法研究[J]. 北京大学学报(自然科学版), 35(6): 794-801.

Béla M. 2010. Spatial Analysis 4. Digital elevation modeling[EB/OL]. http://www.tankonyvtar.hu/hu/tartalom/tamop425/0027_SAN4/cho1so2.html.

Chang KT. 2010. 地理信息系统导论(原著第五版)[M]. 陈建飞, 张筱林译. 北京: 科学出版社: 328-360.

Daly C, Nelson RP, Phillips DL. 1994. A statistical-topographic model for mapping climatological precipitation over mountainous terrain[J]. Journal of Applied Meteorology, 33(33), 140-158.

Feng QS, Liang TG, Huang XD, et al. 2013. Characteristics of global potential natural vegetation distribution from 1911 to 2000 based on comprehensive sequential classification system approach[J]. Grassland Science, 59(2): 87-99.

Franke R.1982. Scattered data interpolation: Tests of some methods[J]. Mathematics of Computation, 38(157): 181-200.

Hijmans RJ, Cameron SE, Parra JL, et al. 2005. Very high resolution interpolated climate surfaces for global land areas[J]. International Journal of Climatology, 25(15): 1965-1978.

Hutchinson MF. 1991. ANUDEM and ANUSPLIN User Guides[R]. Canberra, Australian National University, Centre for Resource and Environmental Studies Miscellaneous Report 91/1.

Jolliffe IT. 1986. Principal Component Analysis[M]. New York: Springer-Verlag.

Kuhn, Tucker. 1950. Nonlinear programming[A]. *In*: Neyman, J. The Second Berkeley Symposium on Mathematical Statistics and Probability (held in Berkley 1951)[C]. Berkley: University of California Press: 481-492.

Lam NS. 1983. Spatial interpolation methods-A review[J]. The American Cartographer, 10(2): 129-149.

Liang TG, Feng QS, Cao JJ, et al. 2012a. Changes in global potential vegetation distributions from 1911 to 2000 as simulated by the Comprehensive Sequential Classification System approach[J]. Science Bulletin, 57(11): 1298-1310.

Liang TG, Feng QS, Yu H, et al. 2012b. Dynamics of natural vegetation on the Tibetan Plateau from past to future using a comprehensive and sequential classification system and remote sensing data[J]. Grassland Science, 58(4): 208-220.

Lin HL, Feng QS, Liang TG, et al. 2013b. Modelling global-scale potential grassland changes in spatio-temporal patterns to global climate change[J]. International Journal of Sustainable Development & World Ecology, 20(1): 83-96.

Lin HL, Wang XL, Zhang YJ, et al. 2013a. Spatio-temporal dynamics on the distribution, extent, and net primary productivity of potential grassland in response to climate changes in China[J]. The Rangeland Journal, 35: 409-425.

Lin HL, Zhang YJ. 2013. Evaluation of six methods to predict grassland net primary productivity along an altitudinal gradient in the Alxa Rangeland, Western Inner Mongolia, China[J]. Grassland Science, 59(2): 100-110.

Messina JP, Taylor SM, Meshnick SR, et al. 2011. Population, behavioural and environmental drivers of malaria prevalence in the Democratic Republic of Congo[J]. Malaria Journal, 10: 161

Oliver MA, Webster R. 1990. Kriging: A method of interpolation for geographical information systems[J]. International Journal of Geographic Information Systems, 4(3): 313-332.

Sakawa M, Yano H. 1989. An interactive fuzzy satisficing method for multiobjective nonlinear programming problems with fuzzy para-meters[J]. Fuzzy Sets and Systems, 30(10): 221-238.

Von Neumann J, Morgenstern O. 1947. Theory of games and economic behavior[J]. Princeton University

Press, 21(1): 2-14.
Wabba G. 1981. Spline interpolation and smoothing on the sphere[J]. SIAM Journal on Scientific and Statistical Computing, 2: 5-16.
Watson DF, Philip GM. 1985. A refinement of inverse distance weighted interpolation[J]. Geo-Processing, 2(2): 315-327.
Zadeh L. 1963. Optimality and non-scalar-valued performance criteria[J]. IEEE Transactions on Automatic Control, 8(1): 59-60.
Zadeh LA. 1965. Fuzzy sets[J]. Information and Control, 8(3): 338-353.

第四章　草业信息模拟模型

草业信息模拟模型是信息化背景下草业科学研究的重要内容之一。本章从信息模拟模型分类及模拟模型构建的角度，重点介绍草业信息学中与3S系统（GIS、GPS、RS）和专家系统（ES）密切相关的模型选择及模型构建等内容。

第一节　草业信息模拟模型的概念、特征与功能

草业模拟模型是草业科学研究方法的重要组成部分，也是草业专家系统主要核心功能的实现途径，它的概念的外延与内涵、特征与功能决定了其与数学、统计学、软件工程等学科的紧密联系，同时还具有其自身特点。

一、草业信息模拟模型的概念和类型

（一）模拟、模拟模型和草业信息模拟模型概念

模拟（Simulation）：广义上说，就是将一个实际事物的某些特征提取出来，通过简化的方式模拟出类似的情景与状态，达到模拟真实的效果，同时要具备模拟的易操作性。从科研角度来说，模拟是为了实验的简化或研究实际系统问题的抽象简化，基于相似准则，通过实验模拟实际环境的某些特征，从而简化大型实验和系统特征的需要，同时达到定量分析的效果。经常采用的模拟手段有计算机模拟、实验模拟、数学建模模拟等。

模拟是用物质的或观念的形式对实际物体、系统、过程或情境的仿真。模拟有物理模型、情境模拟、类比模拟和数学模型等形式。物理模型往往与所要再现的物体有某种相似性。情境模拟可以是对环境中的物理特征的模拟，也可以是对态度、气氛等观念性东西的模拟。类比模拟是用更容易观念化、更容易使用或操作的特征代表所要模拟的特征，它往往具有抽象化的特点。数学模型是由数量化的描述所要研究特征的公式组成。在工程心理学中，模拟可用于对人机系统的研究、测试、评价及人员训练。采用模拟技术，可以降低研究与训练的成本，提高其效率；可以开展在现实中无法进行的研究、测试、评价和训练；可以确保危险技能训练的安全性等。采用模拟技术应考虑模拟的逼真度。逼真的模拟是获得有价值的研究结果和产生训练迁移的必要条件。模拟逼真度包括物理模拟逼真度和心理模拟逼真度。前者指模拟物与实际东西的物理特征的相似程度，后者指人在模拟条件下的感受、行为与在实际情境中的相似程度。在工程心理学的研究中，心理模拟逼真度更为重要。提高物理模拟的逼真度一般能提高心理模拟的逼真度，但两者之间并不完全相关。有时有限的物理模拟逼真度也能产生较高的心理模拟逼真度。为了达到所要求的心理模拟逼真度，设计模拟物时要特别注意仿制那些对人的作业测量和训练迁移起作用的特征。