



中国地质大学（北京）

土壤生态环境定量化实例分析研究

报告人：周萍

中国地质大学（北京）
2020年3月



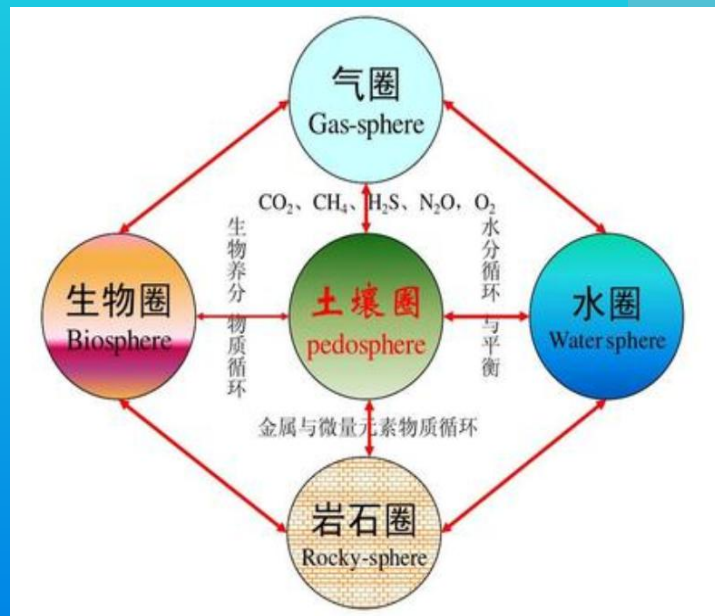
提 纲

- 一、背景意义
- 二、实例分析1
- 三、实例分析2
- 四、几点思考



❖ 背景意义

土壤是满足人类需求的基本资源，如食物和饲料生产、纤维、清洁空气和水。然而，土壤不是一种无限的资源。人类活动和土地利用的变化导致土壤流失的增加，从而降低了自然的循环系统，降低了土地生产力，从而降低了世界范围内的人类福祉！所以研究他刻不容缓！



土壤各圈层中间介质



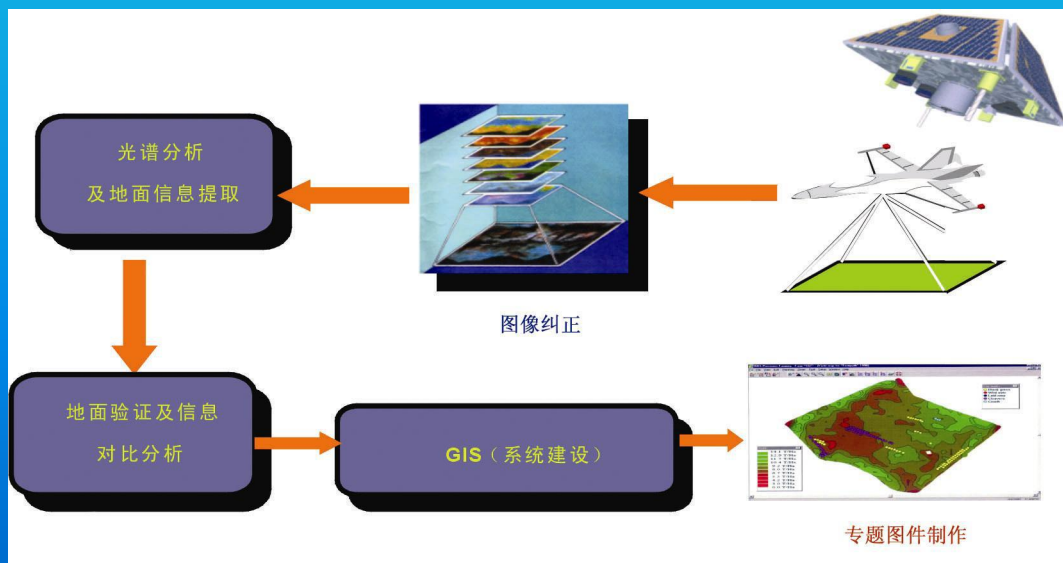
中国地质大学（北京）

土地退化形势严峻





随着土地资源调查的深入,土地资源的管理已从数量管理向质量维护生态保护方向发展,同时向着自动化、定量化方向迈进。在先进的技术理论指导下,利用各种高新技术手段(如:各种量化的数理模型、土壤光谱分析技术等)可高效、快速提高土地质量评价,为土地资源调查、评价和管理提供更丰富、精确的数据源。这对土地资源的管理具有极其重要的指导意义。



土地资源遥感的工作流程



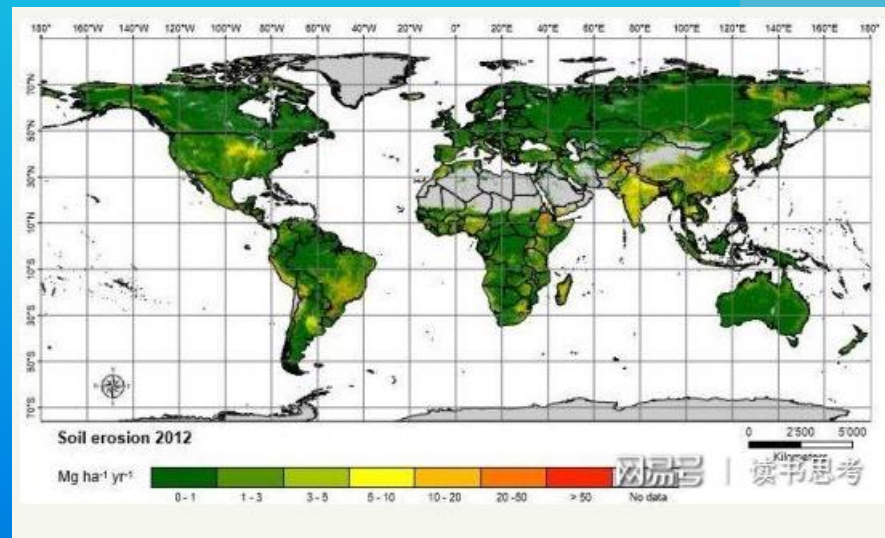
提 纲

- 一、背景意义
- 二、实例分析1
- 三、实例分析2
- 四、几点思考



实例分析1、基于RUSLE模型吉林省土壤水力侵蚀定量评价

土壤侵蚀是指土壤及其母质在水力、风力、冻融或重力等外营力作用下，被破坏、剥蚀、搬运和沉积的过程。土壤侵蚀已成为当今人类面临的一种最普遍、持续性很强的环境灾害，被称为“蠕动的灾难”（倪九派，2005）。从全球范围来看，土壤侵蚀是威胁当今人类生存与发展的全球性环境灾害之一，严重制约着全球社会经济持续发展。据估算全球水土流失面积约 $16.43 \times 10^6 \text{ km}^2$ ，占地表总面积的10.95%。





中国地质大学（北京）

一、国内外研究现状

土壤水土流失是全球问题，我国情形也很严峻！从东到西从南道北到土壤侵蚀面广量大，危害严重，影响深远（近些年我国东北黑土地土壤流失严重）！

土壤侵蚀模型是水土流失危险性评价和水土保持规划的有效工具，是土壤侵蚀研究成果的集中体现。在国际上，土壤侵蚀研究经历了由定性、半定量到定量三个阶段，研究尺度主要有坡度、小流域、区域三个方面。





模型方面研究先后出现了通用土壤流失方程（USLE-Universal Soil-Loss Equation）、非点源农业污染模型（AGNPS、ANSWERS）、修正的通用流失方程（RUSLE-Revised Universal Soil-Loss Equation）、水土资源评价模型（SWAT）、水蚀预报模型（WEPP）、欧洲土壤侵蚀模型（EROSEM）、荷兰土壤侵蚀预报模型（LISEM）、若干风蚀预报模型（WEQ、RWEQ、WEPP、WEPS。郑粉莉等，2010）等，为治理地区水土流失、评价水土保持效益提供了强有力的手段。



通用土壤流失方程USLE是美国农业部专家Wischmeier, W和Smith, D (1968)在对美国东部地区30个州10000多个径流小区近30年观测资料，经系统分析所提出，已被众多国家广泛采用。

美国土壤保持局的Renard在USLE模型基础上修订为RUSLE模型（Renard KG et al, 1994）。RUSLE模型使用数据广泛、数据可获取性高、资料依据充分、涉及区域广泛，推出后迅速在世界得到了广泛的认可和应用。

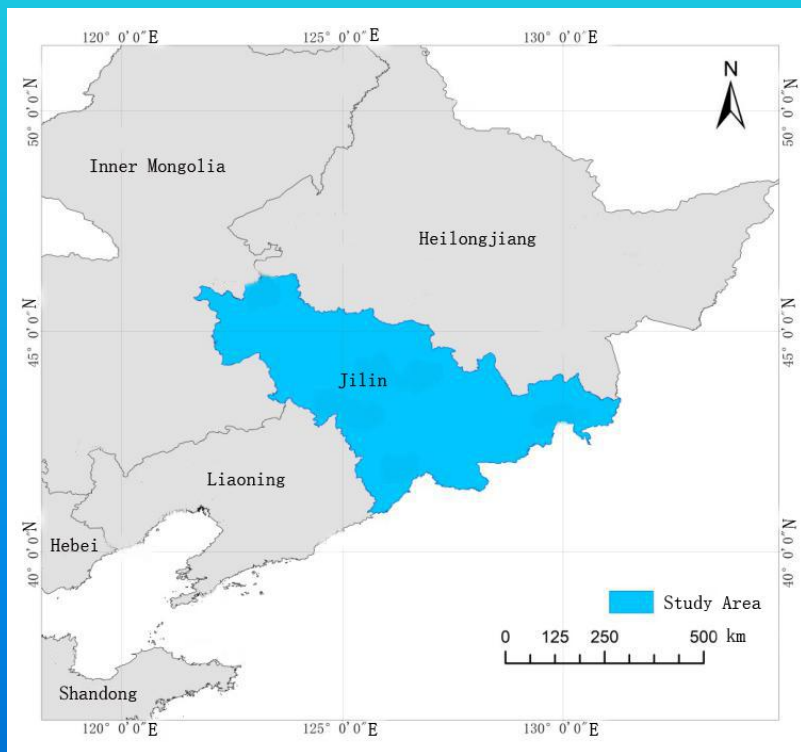


我国学者也普遍以此为基础，在各个试验区进行了示范性研究。张宪奎（1992）在黑龙江设置径流试验区并建立了黑龙江土壤流失方程；江忠善（2005）考虑浅沟侵蚀对坡面土壤侵蚀具有重要贡献的特点，构建了坡面土壤流失预报模型；刘宝元（2002）通过云贵高原、黄土高原等地不同小区试验，建立了中国的水土流失方程CSLE。



三、研究区概况及数据简介

二、研究区概况及数据简介



以吉林省为研究区，地跨东经 $121^{\circ} 38' \sim 131^{\circ} 19'$ ，北纬 $40^{\circ} 50' \sim 46^{\circ} 19'$ ，总面积**18.74万**平方公里，约占国土总面积的2%。

南邻辽宁省，西接内蒙古自治区，北与黑龙江省为邻，东和俄罗斯联邦接壤，东南同朝鲜相望。管辖长春1个副省级城市、吉林、四平、通化、白山、辽源、白城、松原7个地级城市和延边朝鲜族自治州。



- ◆ 降雨数据：CMORPH卫星降雨数据，空间分辨率为 $8\text{km} \times 8\text{km}$
- ◆ 植被指数数据：MODIS NDVI数据，空间分辨率 $250\text{m} \times 250\text{m}$
- ◆ 地形数据：数字高程模型SRTM，空间分辨率 $90\text{m} \times 90\text{m}$
- ◆ 土壤数据：HWSD中国土壤数据集，空间分辨率 $1\text{km} \times 1\text{km}$
- ◆ 土地利用类型数据：LUCC全国土地利用数据产品，空间分辨率 $1\text{km} \times 1\text{km}$



RUSLE（修正后的通用土壤侵蚀方程）模型：

$$A=R*K*LS*C*P$$

- A为单位面积的年土壤流失量($t/hm^2 \cdot a$);
- R 为降雨侵蚀力($MJ \cdot mm/hm^2 \cdot h \cdot a$);
- K为土壤可蚀性因子($t \cdot hm^2 \cdot h/hm^2 \cdot MJ \cdot mm$);
- LS为地形因子，无量纲，其中L为坡长因子，S为坡度因子；
- C为植被覆盖与作物管理因子，无量纲；
- P为水土保持措施因子，无量纲。



RAINFALL EROSIVITY FACTOR

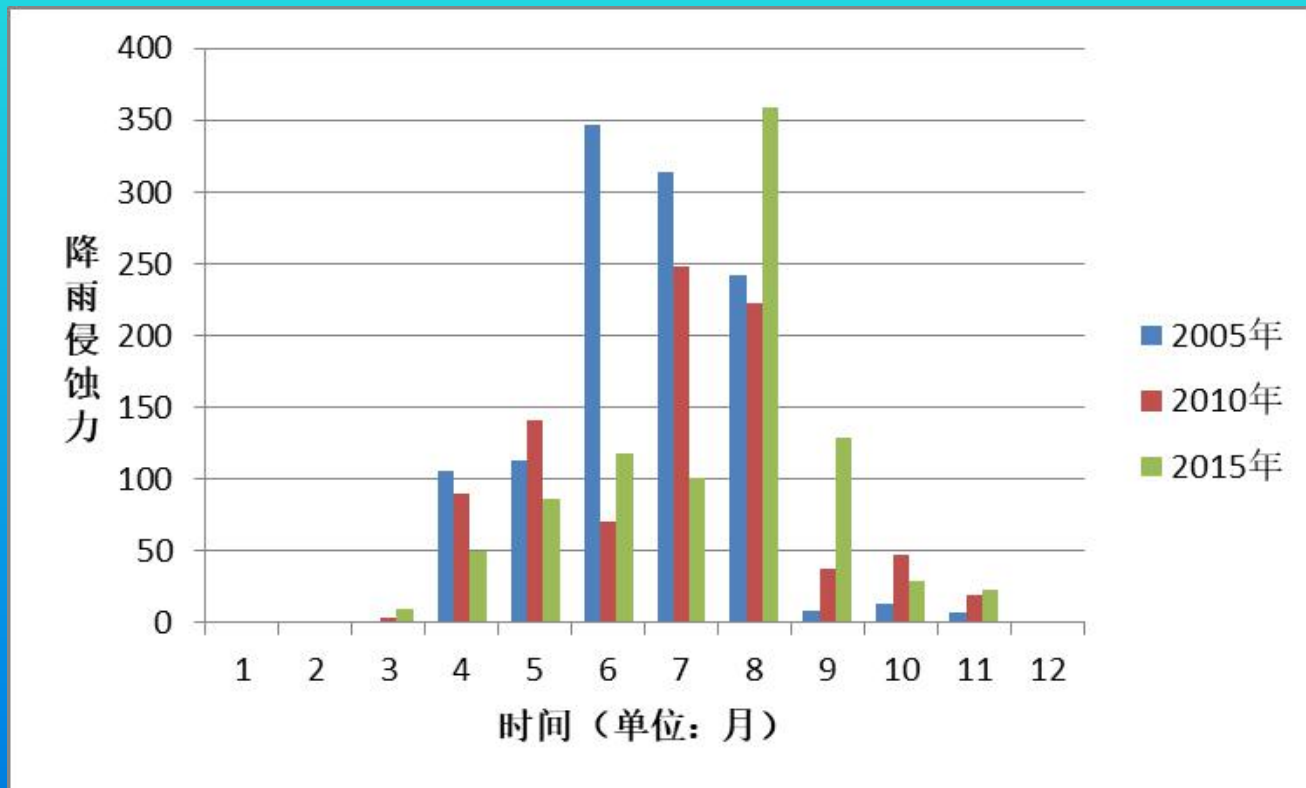
降雨侵蚀力因子模型

$$R_i = E_{60} I_{30}$$

R_i 单位面积次降雨侵蚀力 MJ.mm/hm² .h

E_{60} 单位面积次降雨60分钟最大降雨动能 MJ/hm²

I_{30} 次降雨30分钟最大降雨强度 mm/h





土壤可蚀性因子K

K因子是衡量土壤抗侵蚀性的指标，用于反映土壤对侵蚀的敏感性。在此采用Williams等在EPIC（Erosion-Productivity Impact Calculator）模型中的定量统计模型法($t \cdot \text{hm}^2 \cdot \text{h} / \text{hm}^2 \cdot \text{MJ} \cdot \text{mm}$)。

土壤可蚀性K因子表示在其他土壤侵蚀因子不变时，是一项评价土壤侵蚀力分离冲蚀和搬运难易程度的指标，既土壤的抗侵蚀特性，抗侵蚀能力越强，土壤可蚀性因子K值越低，反之K值越高。

影响K因子的因素是多方面的，通常质地越粗或越细的土壤有较低的K值，而质地适中的反而有较高的K值。土壤可蚀性因子的研究方法主要有：野外观测法、室内实验法、数学分析法、定量统计模型法。

是与土壤质地有关



土壤可蚀性因子K

$$K = \left\{ 0.2 + 0.3 \exp \left[-0.0256 S_a \left(\frac{S_i}{100} \right) \right] \right\} \left[\frac{S_i}{C_i + S_i} \right]^{0.3} \left[-\frac{0.25C}{C + \exp(3.72 - 2.95C)} \right] \left[-\frac{0.7S_n}{S_n + \exp(-5.51 + 22.9S_n)} \right]$$

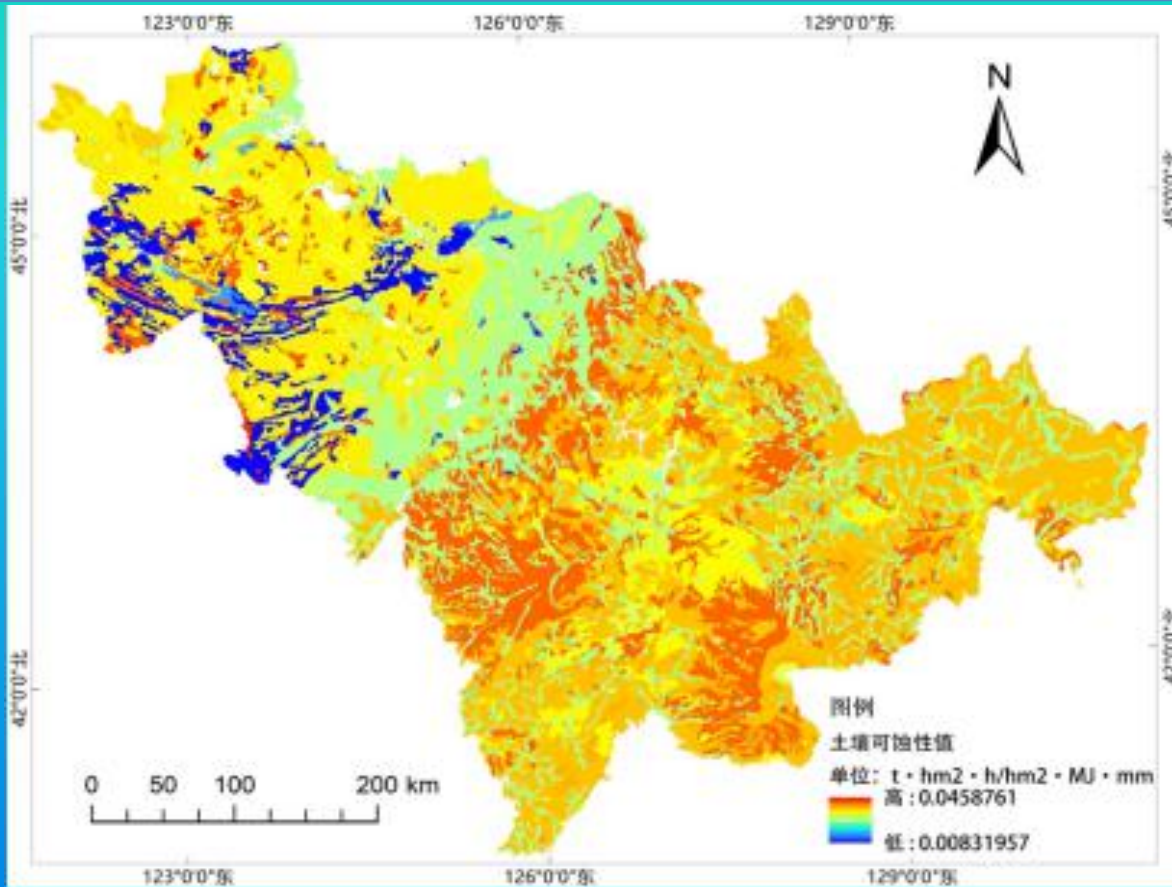
式中：

- $S_n = 1 - S_a / 100$;
- S_a 为砂粒含量%;
- S_i 为粉粒含量%;
- C_i 为粘粒含量%;
- C 为有机碳含量%;

根据土壤有机质和颗粒组成资料推算研究区不同土壤类型的可蚀性因子



土壤可蚀性因子



环形带上的第四纪沉积物主要以冲积、风积、湖积形式出现（属于草原风沙土），成分以细沙、粘土及泥炭为主。其北部和南部分布的第四纪的沉积类型以冲积、堆积的中等砂砾等沙化沉积为主（属于沙化淡黑钙土）。也正是由于该区域独特的地质现象，造成了土地水土流失差异性分化的出现。



VEGETATION FACTOR

C 因子可以衡量植被覆盖和作物管理对土壤侵蚀的抑制作用。C值与植被覆盖度有关，植被覆盖度估算法是目前国内外专家应用最普遍的C因子估算方法。本研究采用蔡崇法提出的C因子与植被覆盖度关系方程

植被覆盖与作物管理因子C因子表示植被覆盖和管理措施对土壤侵蚀的作用，是人为控制土壤侵蚀的重要因子。RUSLE中C因子定义为在一定地表覆盖和管理措施下，土壤流失量与同等条件下无翻耕、连续休闲对照地上土壤流失量之比，无量纲，其值变化范围在0-1之间，评价时长可以是一次降雨历时、一个农作周期、一个轮作期、一个作物年或是一年。它的值的大小取决于研究区具体的作物覆盖或植被覆盖、轮作顺序、植物生长期中侵蚀性降雨的分布、作物残余物管理、人为管理措施等的综合作用，当地面完全裸露时，C值为1.0，当地面得到良好保护时，C值为0.001（张岩等，2002）



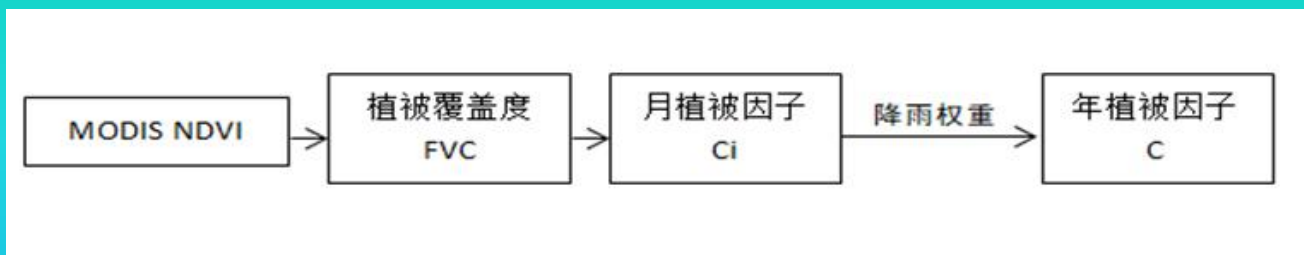
植被覆盖与作物管理因子模型:

$$\begin{cases} C = 1, & FVC = 0 \\ C = 0.6508 - 0.34361 \cdot \lg V, & 0 \leq FVC < 78.3\% \\ C = 1, & FVC > 78.3\% \end{cases}$$

- 式中: FVC为植被覆盖度;



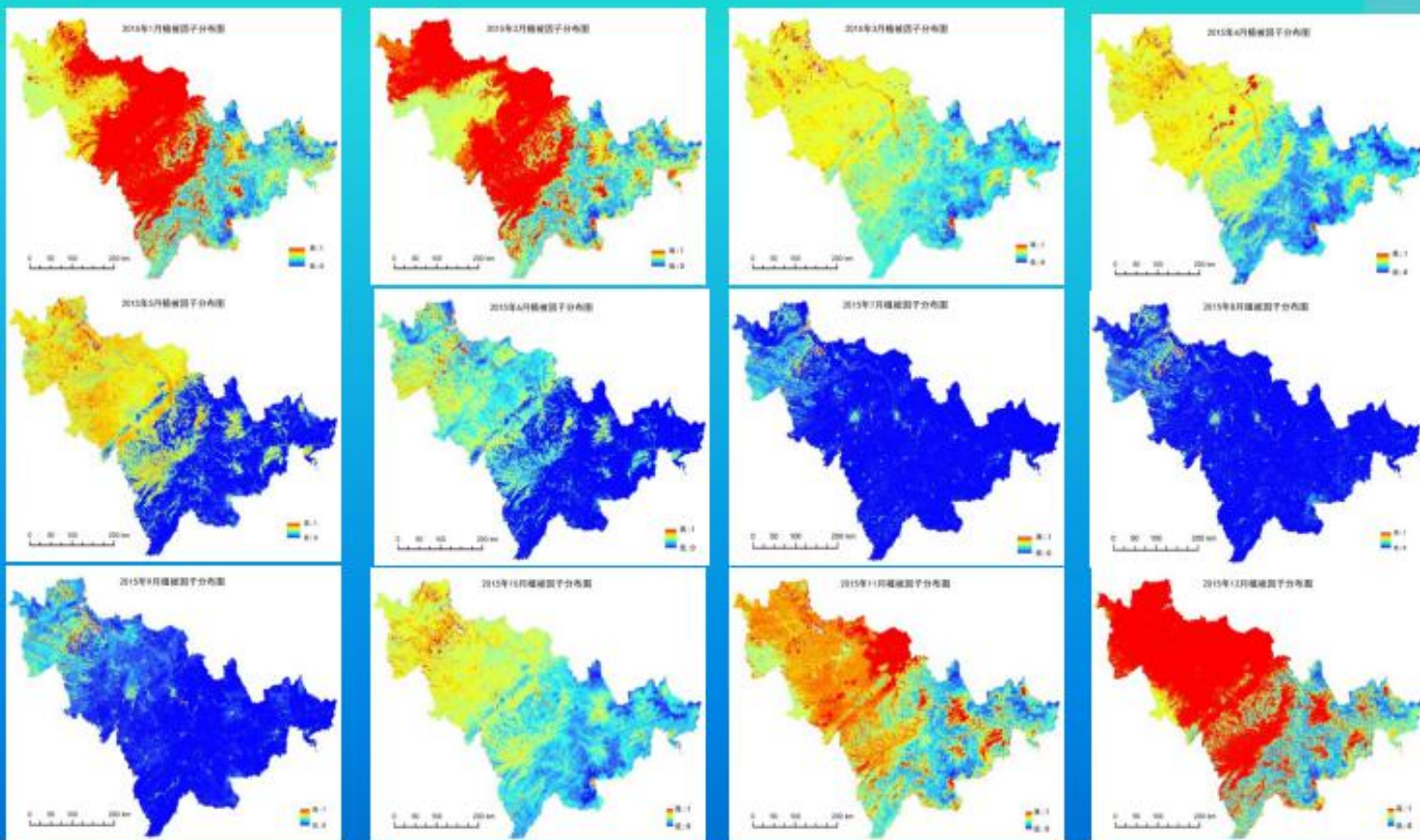
VEGETATION FACTOR

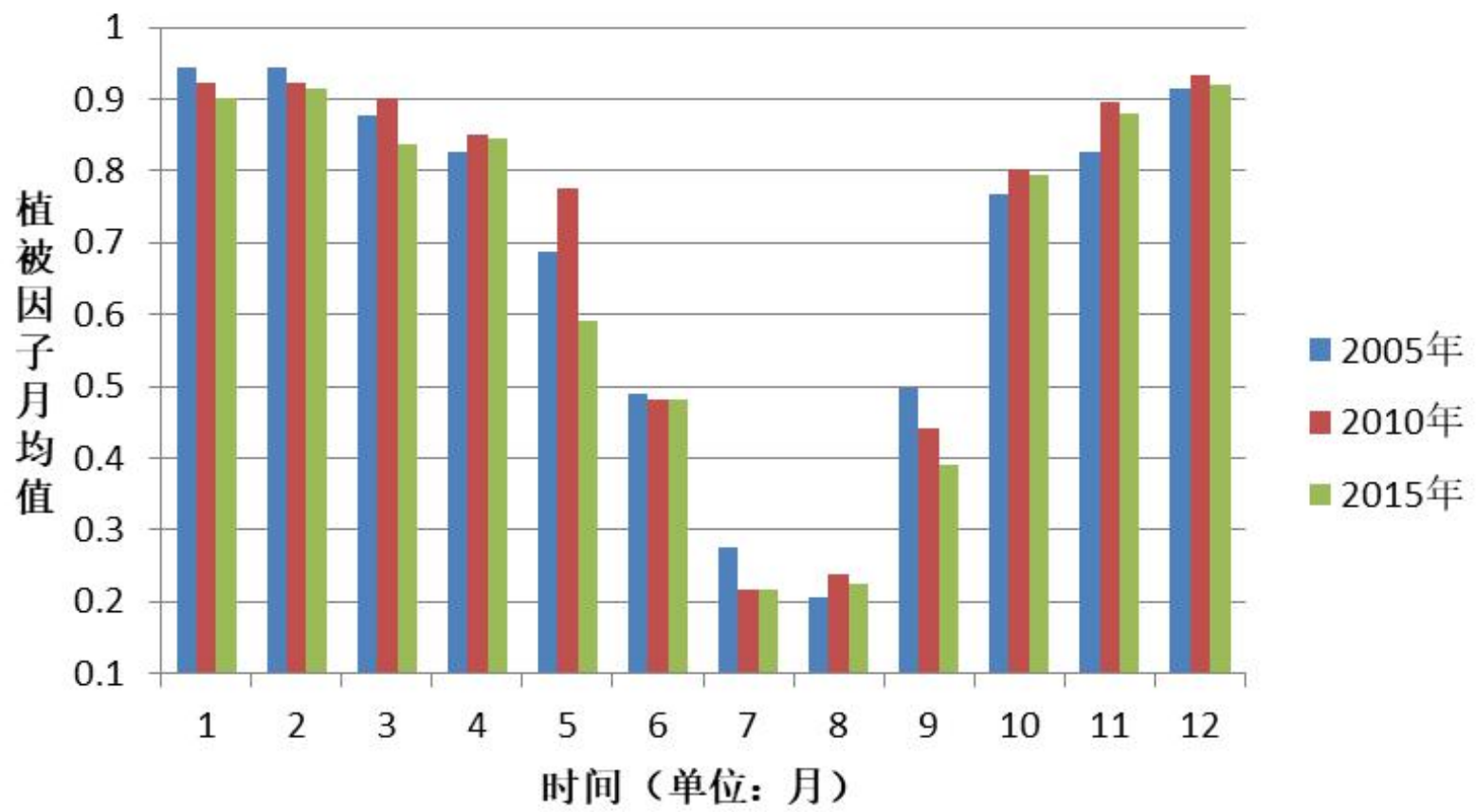


在侵蚀性降雨高发季节，相应的植物覆盖状况或差或好，在地区之间差异很大。因此，推算一个地方较为合适的 C 值需要知道侵蚀性降雨在全年 12 个月中的分布，根据CMORPH降雨数据计算得到的吉林省每个月的降雨量，对相应月份的C值赋权重，计算得到全年总C因子。



2015年每月植被覆盖与作物管理因子分布







SLOPE LENGTH AND SLOPE FACTOR

LS因子主要是通过坡长和坡度直接影响地表径流速率，坡长越长，坡度越大，对土壤侵蚀影响越大

坡度因子模型：

$$\begin{cases} S = 10.8 \sin \theta + 0.03, & \theta < 5^\circ \\ S = 16.8 \sin \theta - 0.5, & 5^\circ \leq \theta < 10^\circ \\ S = 21.9 \sin \theta - 0.96, & \theta \geq 10^\circ \end{cases}$$

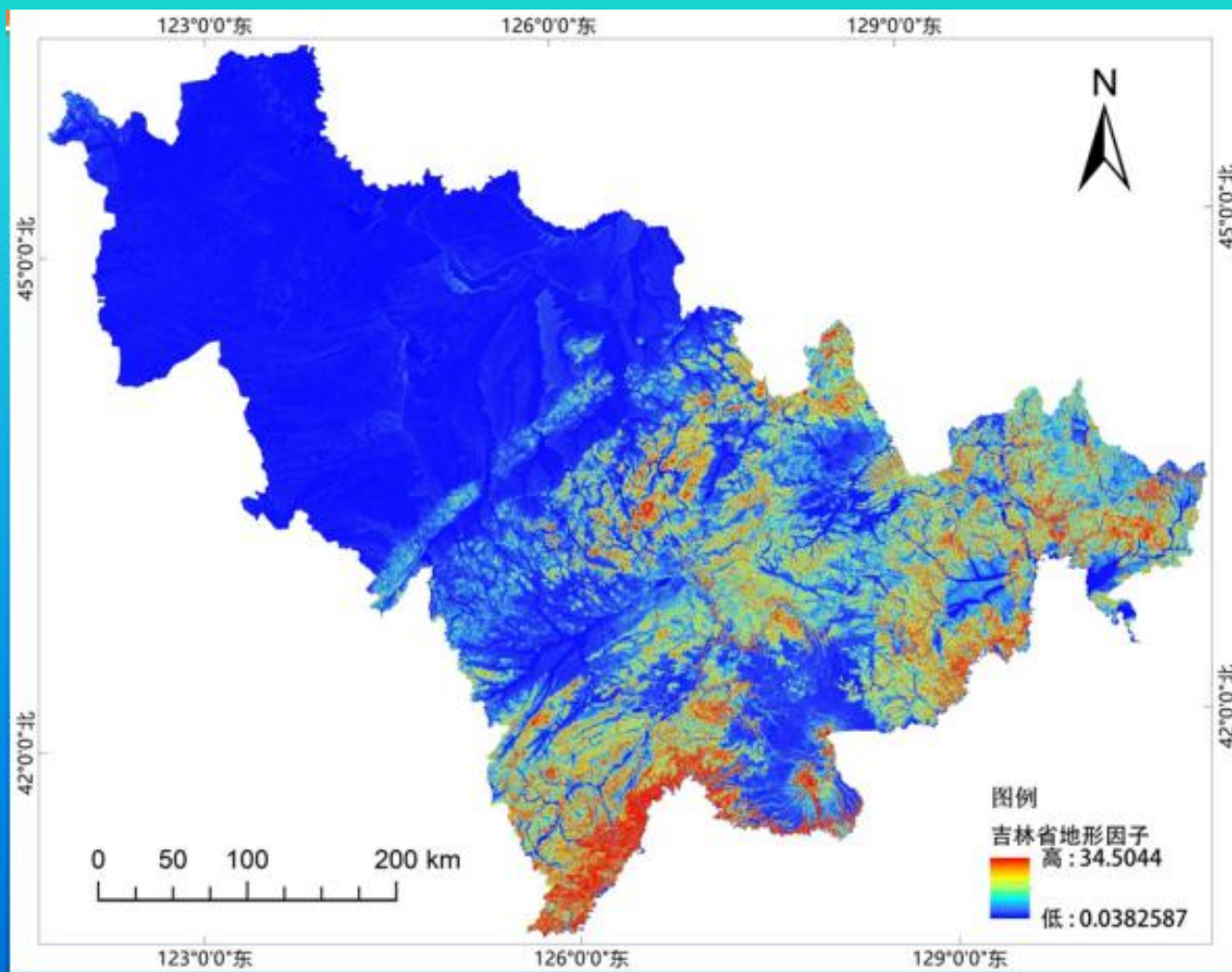
坡长因子模型：

$$L = (\lambda / 22.1)^m$$

- θ 为坡度；
- λ 为坡长（m）；
- m 为随坡度 θ 变化的因子，当 $\theta \leq 1^\circ$ 时， $m=0.2$ ；当 $1^\circ < \theta \leq 3^\circ$ 时， $m=0.3$ ；当 $3^\circ < \theta \leq 5^\circ$ 时， $m=0.4$ ；当 $\theta > 5^\circ$ 时， $m=0.5$ ；



2015年





水土保持防治措施P因子

P定义：采用专门水土保持措施后的土壤流失量与顺坡种植时的土壤流失量的比值，其范围在0-1 之间，0 值代表根本不发生侵蚀的地区，而1值代表未采取任何控制措施的发生侵蚀地区。

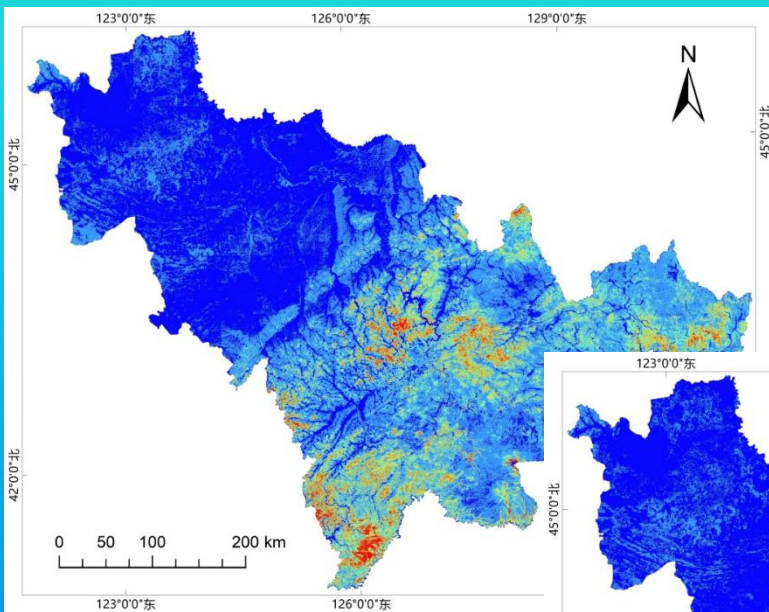
水土保持措施对土壤流失量的减少具有重要作用，目前主要的水土保持措施有等高耕种、等高带状种植、垄作、覆盖耕作、少耕免耕等措施，主要用于减少农用地土壤侵蚀，提高农业生产，原理是通过增加地面粗糙度、减少地表径流量、改变地形及汇流方式、增加土壤入渗能力、减小地表水流流速等来阻止地表水土流失。



水土保持防治措施P因子

参考美国农业部《通用土壤流失方程说明书》中的P值赋值，以及张宪奎通过收集黑龙江全省试验研究资料得到的P值，结合课题组前往吉林省的野外踏勘调研情况，整合得到吉林省不同土地利用类型的P值。

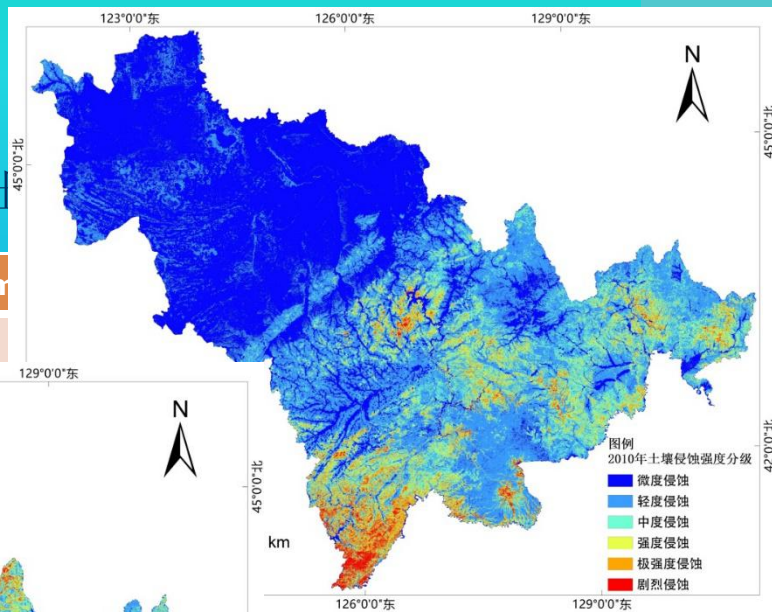
土地利用类型	水田	旱地	林地	草地	水域	建筑用地	未利用土地
P值	0.03	0.35	0.9	0.55	0	0.25	1



黑土区及北方土壤侵蚀现状

平均侵蚀模数 $t/(km^2 \cdot a)$

<200

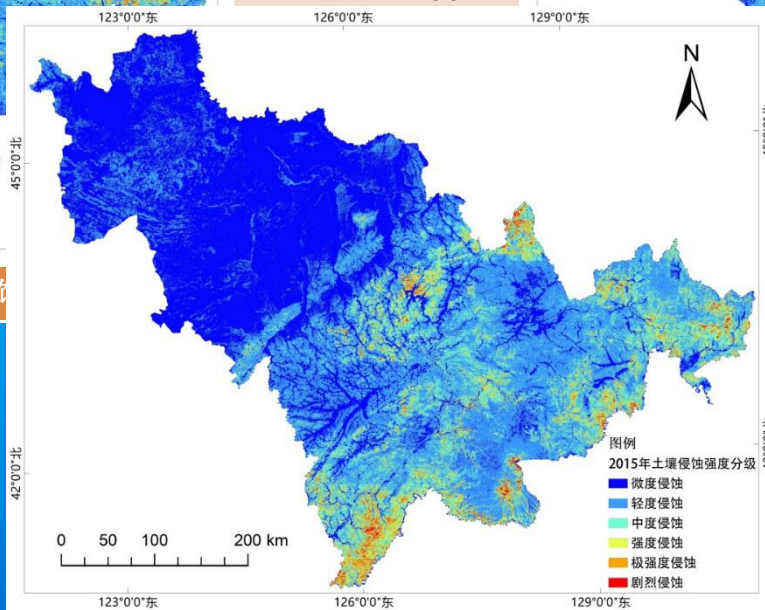


图例
2010年土壤侵蚀强度分级

- 微度侵蚀
- 轻度侵蚀
- 中度侵蚀
- 强度侵蚀
- 极强度侵蚀
- 剧烈侵蚀

剧烈侵蚀

>11.1



图例
2015年土壤侵蚀强度分级

- 微度侵蚀
- 轻度侵蚀
- 中度侵蚀
- 强度侵蚀
- 极强度侵蚀
- 剧烈侵蚀



四、RUSLE模型应用结果分析

模型检验

◆ 定量验证： ^{137}Cs 同位素示踪法

主要是通过检测核爆炸高峰年的峰值 ^{137}Cs 活性即背景值 ^{137}Cs 活性与现阶段活性的对比变化量，从而建立 ^{137}Cs 活性变化量与土壤侵蚀量之间的定量转换关系，以此来计算土壤侵蚀量。

◆ 定性验证：野外实地踏勘验证

通过实地观察土地初始状态、周围状况、耕作层性质等，综合对各个侵蚀等级的野外真实状态进行评定。



野外验证路线图（5条路线、35个验证点）



野外实地验证照片



微度侵蚀



中度侵蚀



极强度侵蚀



验证结果

^{137}Cs 同位素示踪法

- 通过对比分析 ^{137}Cs 示踪法的实测计算数据与RUSLE的估算数据，相关系数达到0.6119；

实地踏勘验证

- 根据野外实际状况进行定性的判断，匹配度较高，共查验验证点35个，较为符合的有32个，准确率为91%。

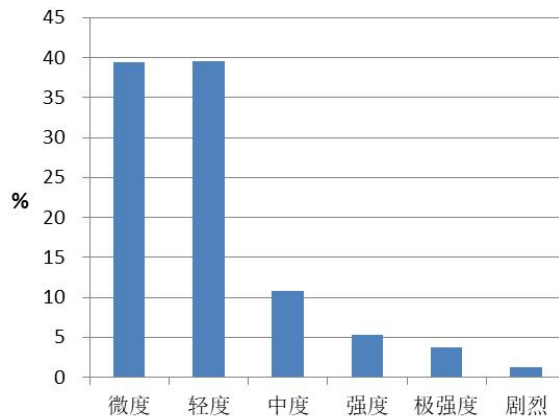


五、土壤侵蚀特征分析

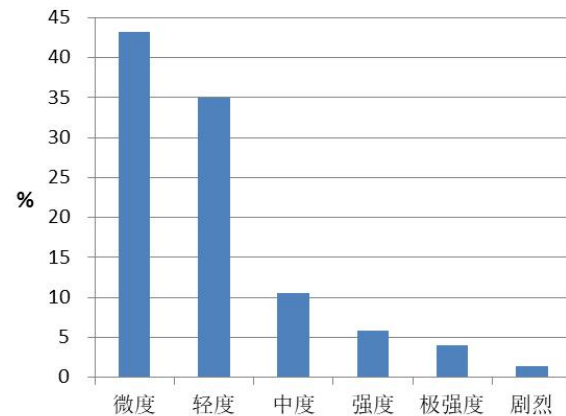
◆ 从土壤侵蚀量来看，2005-2015年期间，先增加后减少。

◆ 从土壤侵蚀面积来看，侵蚀等级呈现由强度以上侵蚀等级向低级侵蚀转移的迹象。

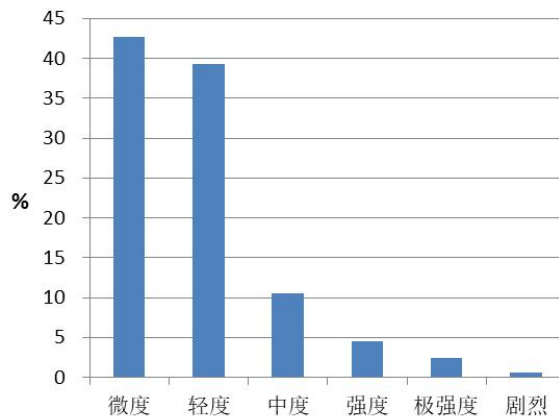
2005年土壤侵蚀强度分布



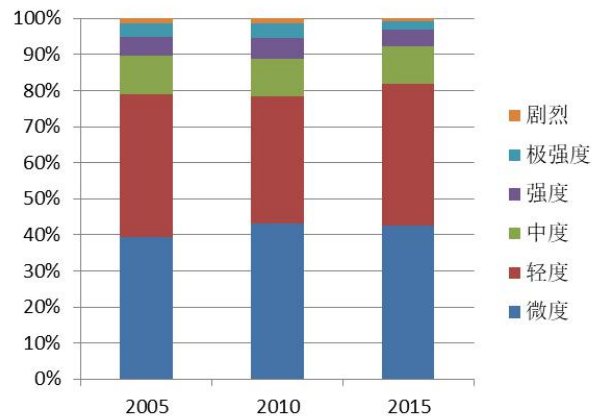
2010年土壤侵蚀强度分布

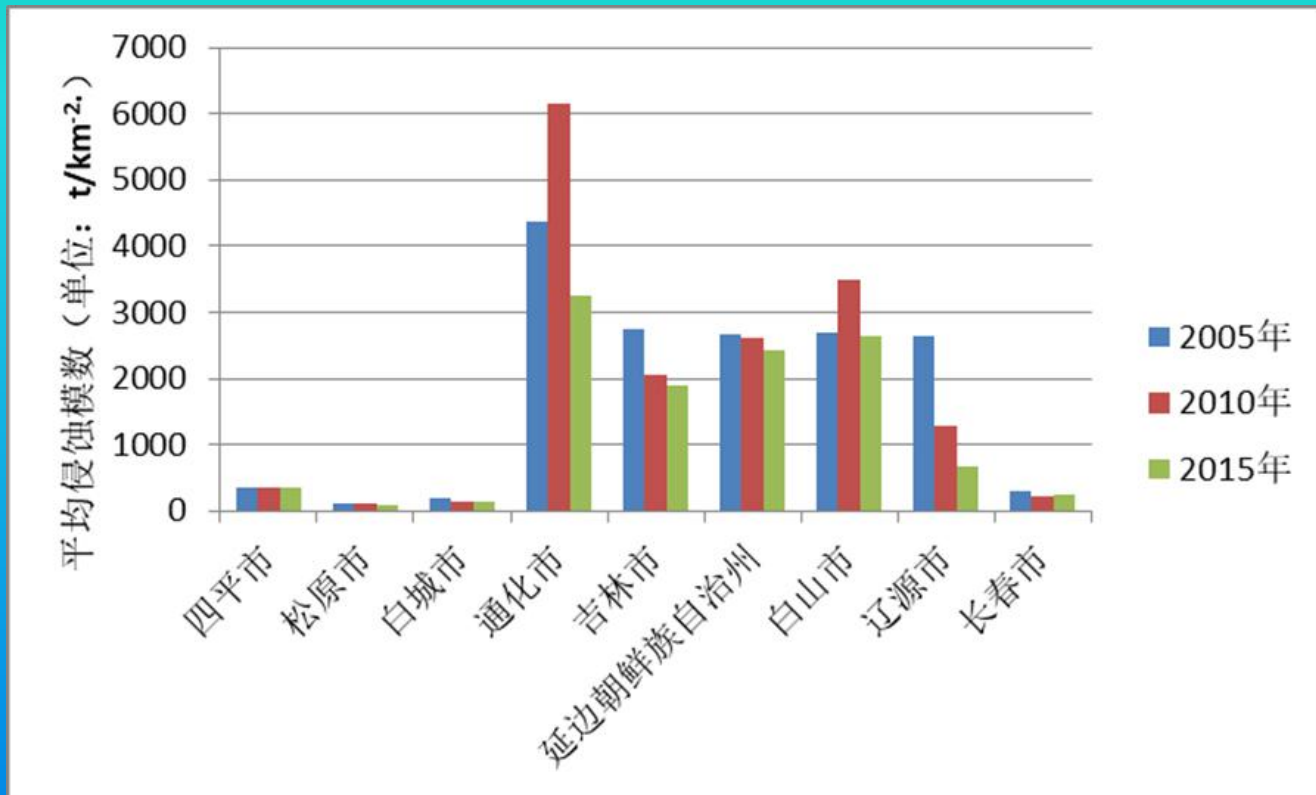


2015年土壤侵蚀强度分布



2005、2010、2015土壤侵蚀强度分布对比



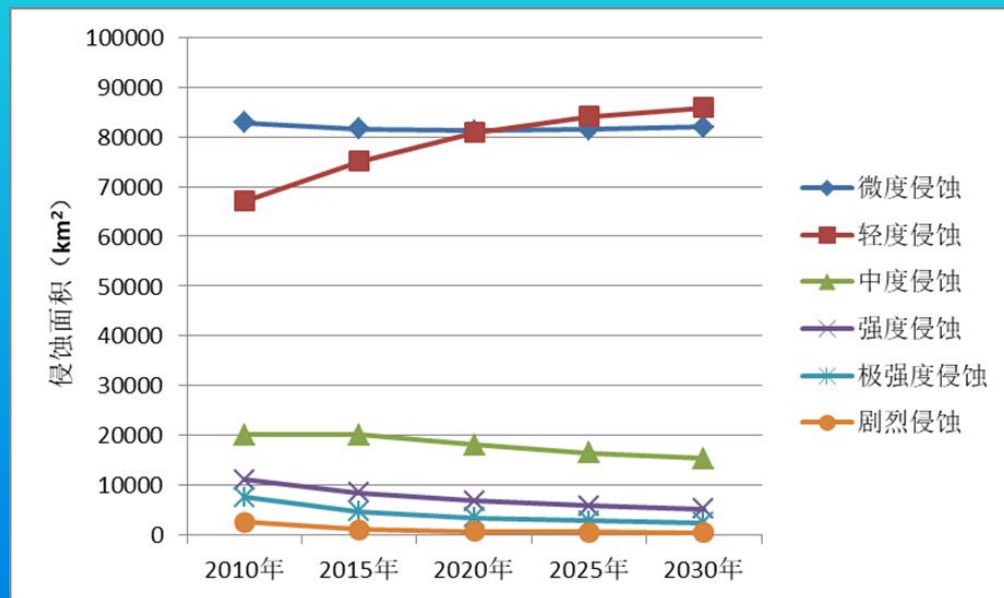


- ◆ 通化市的土壤侵蚀模数三年中一直是最高
- ◆ 松原市的土壤侵蚀模数三年中一直是最低
- ◆ 辽源市土壤侵蚀量下降幅度最大



六、马尔科夫模型预测2020年以后

马尔科夫（Markov）预测模型是关于事件发生的概率预测方法，由俄国数学家Markov A A在1907年用数学方法研究布朗运动过程时发现的一种随机运动规律，是它利用某一系统的现在状况及其发展动向预测该系统未来状况。



由2020年以后预测趋势是，微度侵蚀保持平稳，轻度侵蚀大幅度增加，中度侵蚀及以上侵蚀都在逐年减少，而且其减少速度比较快。这种重度侵蚀逐渐向低级侵蚀能转化的趋势将保持很长一段时间，最终达到一个相对稳定的状态。



提 纲

- 一、背景意义
- 二、实例分析1
- 三、实例分析2
- 四、几点思考



实例分析2、高光谱技术土地质量信息提取

高光谱提取土地质量信息通常包括：

- 1、提取土壤有机质、全氮、速效磷、速效钾、重金属等信息。
- 2、提取土壤含水量、土壤质地等信息。

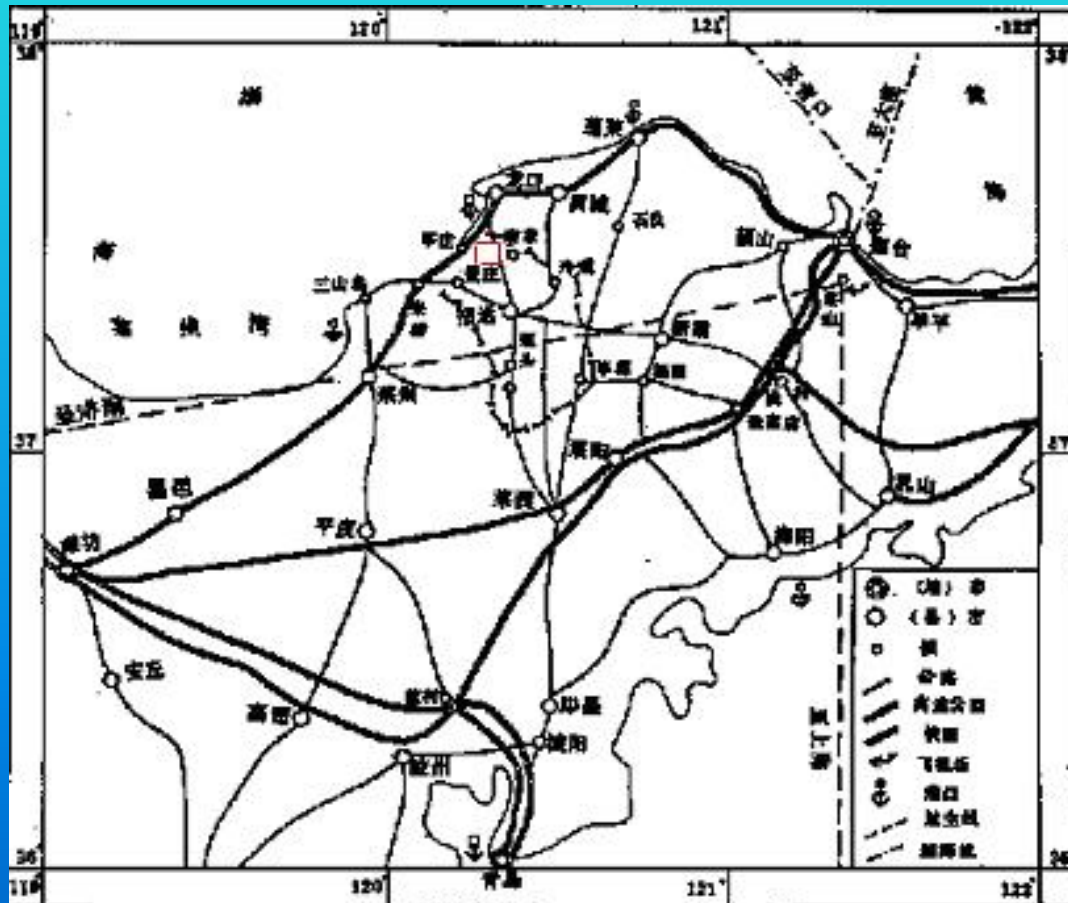
使得土地信息——定量化！



土壤光谱特征组分定量分析

农用地

山东省招远市位于山东半岛西北部，地处胶东低山丘陵地带。地理坐标是：东经 $120^{\circ} 08' \sim 120^{\circ} 38'$ ，北纬 $37^{\circ} 05' \sim 37^{\circ} 33'$ 。



土壤光谱特征组分定量分析



招远土壤状况：

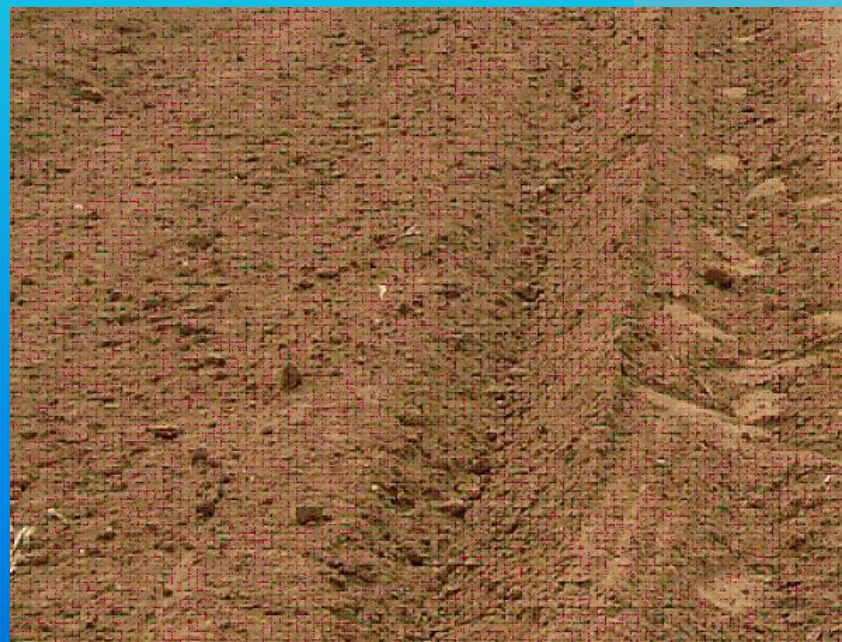
《中国土壤分类系统》将全国土壤从上至下共分7级，分别为土纲、亚纲、土类、亚类、土属、土种和亚种。招远全县境内土壤80%左右为农业用土包括棕壤土、潮土、褐土和水稻土4个土类。

山东招远土壤分类

土 类	棕壤土类			潮土类		褐 土			水稻土
亚 类	棕壤性土	棕壤	潮棕壤	河潮土	滨海潮土	褐土	淋溶褐土	褐土性土	水稻土



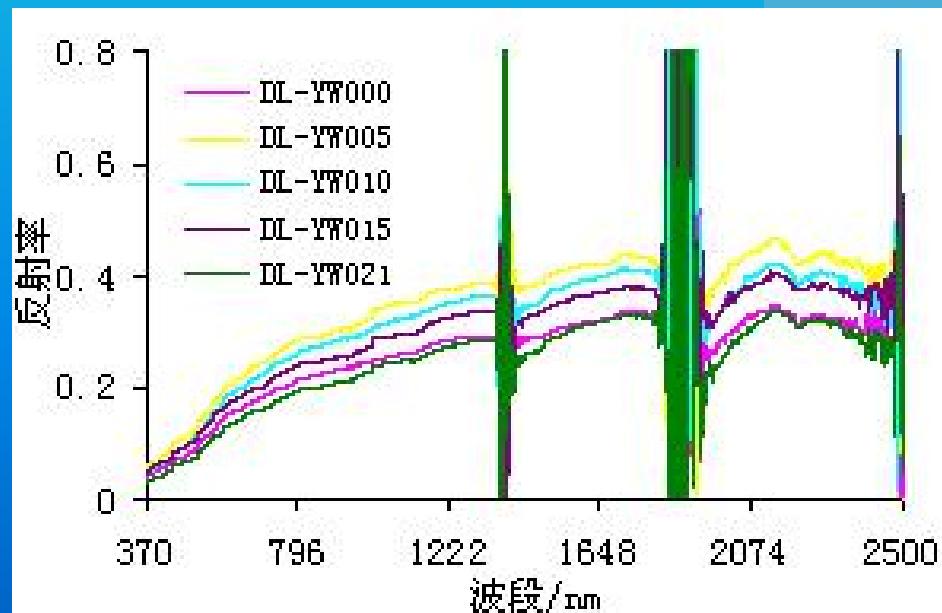
东良乡位于招远西北端，东经 $120^{\circ} 23' \sim 120^{\circ} 27'$ ，北纬 $37^{\circ} 51' \sim 37^{\circ} 54'$ ，面积约 5.7 km^2 是招远主要的农业产粮区。东良乡农用土为潮棕壤亚类俗称“黑油土”是棕壤性土长期受地下水影响发育而成的。颜色由棕色至褐色，上部颜色较暗下部出现铁锈斑，土体深厚（在1米左右），水分状况良好，生产性能好，适种性广，是招远粮食作物的高产农用土。





地面光谱获取

地面土壤光谱采集用美国ASD (Analytical Spectral Device) 公司生产的ASD FieldSpec Pro FRTM型光谱仪。该光谱仪的采样间隔为1.4 nm (350—1000nm区间) 和2nm (在1000—2500nm区间)，光谱分辨率为3.5nm (350—1000nm区间)、10nm (1000—2500nm区间)。



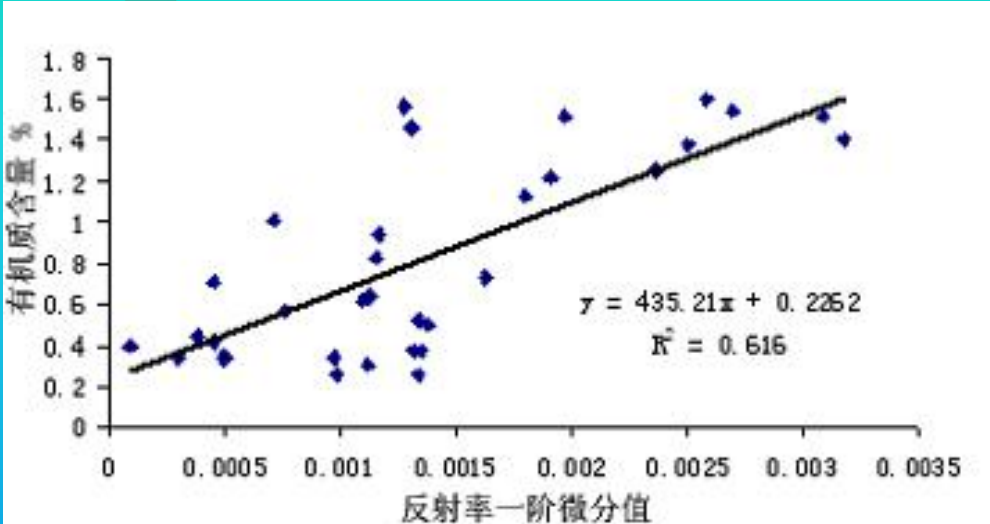


单相关分析——找波段

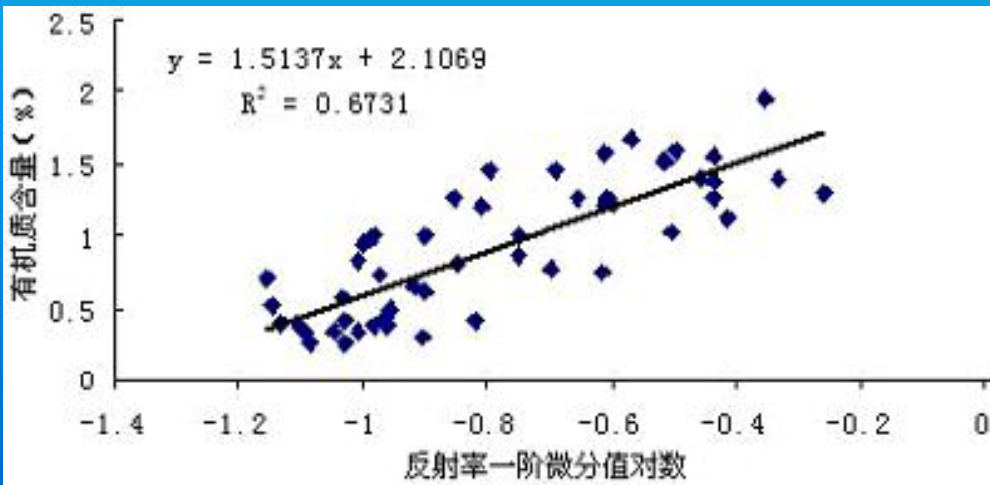
$$\gamma_i = \frac{\text{Cov}(R, \text{OM})}{\sqrt{D(R)} \sqrt{D(\text{OM})}} = \frac{\sum_{n=1}^N (R_{ni} - \bar{R}_i)(\text{OM}_n - \overline{\text{OM}})}{\sqrt{\sum_{n=1}^N (R_{ni} - \bar{R}_i)^2 \sum_{n=1}^N (\text{OM}_n - \overline{\text{OM}})^2}}$$

通过几种反射率变换值与有机质含量的相关性分析认为：反射率一阶微分和反射率对数一阶微分的这两种相关性分析结果均较佳，其反射率变换形式所体现的吸收峰位置具有很好的对应性（762nm、874nm、1667nm），而且反射率对数一阶微分相关效果最好，可用于建立相应的回归方程。

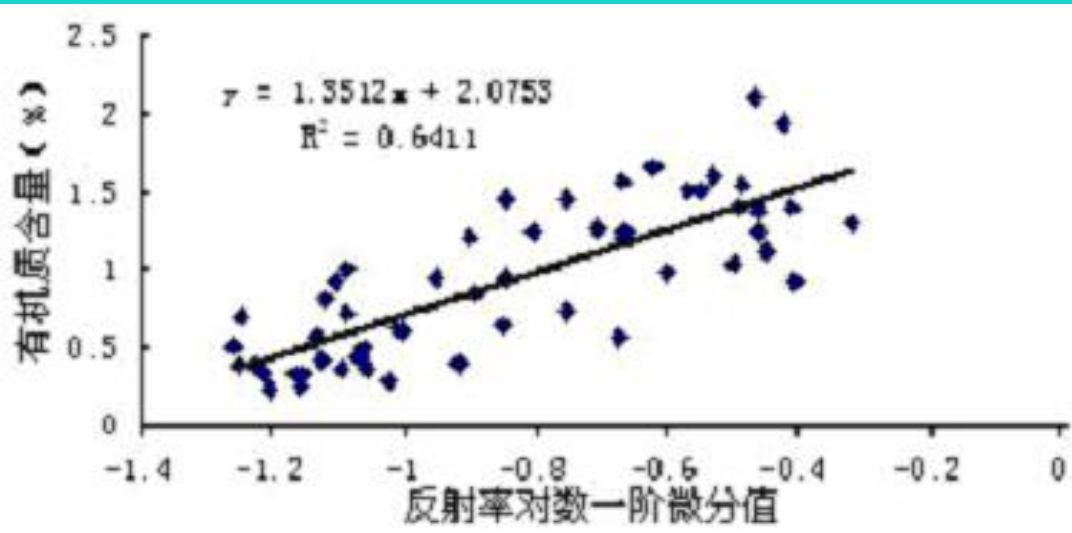
单相关分析



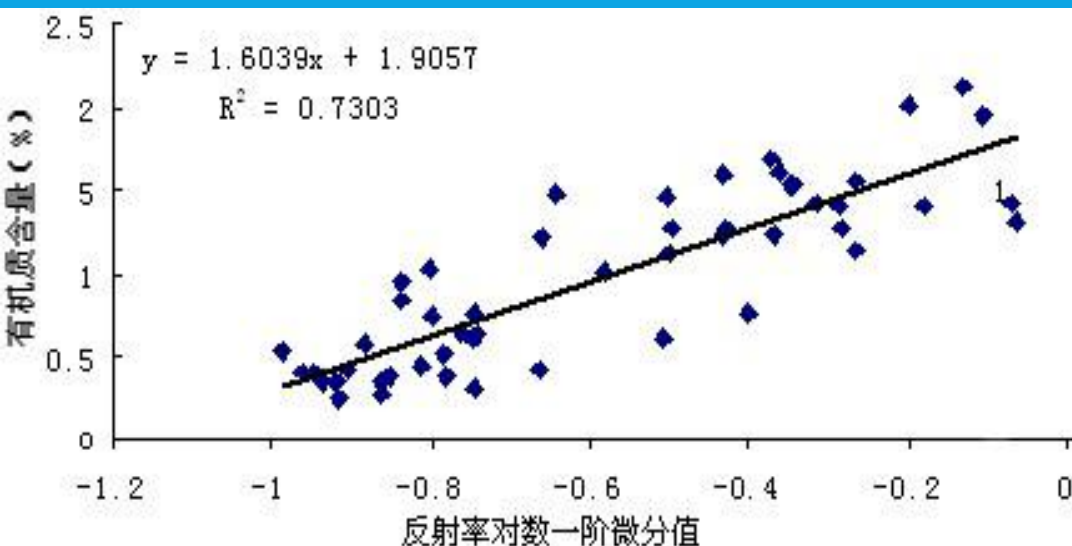
SOM与反射率一阶微分值：
band=874nm



SOM与反射率对数一阶微分值，
band=762nm



SOM与反射率对数一阶微分值,
band= 874nm



SOM与反射率对数一阶微分
值, band=1667nm



将46个样本随机地分成两组：一组用于建立回归预测模型（34个、占总数的76%），另一组用于验证已建立的回归模型（12个、占总数的24%）。

$$Y = b_0 + \sum_{i=1}^k b_i x_i$$

$$Y_p = 0.812 + 16.82 x_{762} - 24.26 x_{874} + 7.04 x_{1667}$$

为检验方程的预测精度，我们引入国际上常用的检验参数 SEC 与 SEP 来判断估测效果。SEC 是评价建模过程估算效果；SEP是评价检验过程估算效果，SEC、SEP分别表示：



建模样本的总均方根差 RMSE :

$$SEC = RMSE - PRMSE = \sqrt{\frac{1}{N_c - k - 1} \sum_{i=1}^{N_c} (Y_i - \hat{Y}_i)^2}$$

检验样本的总均方根差 RMSE :

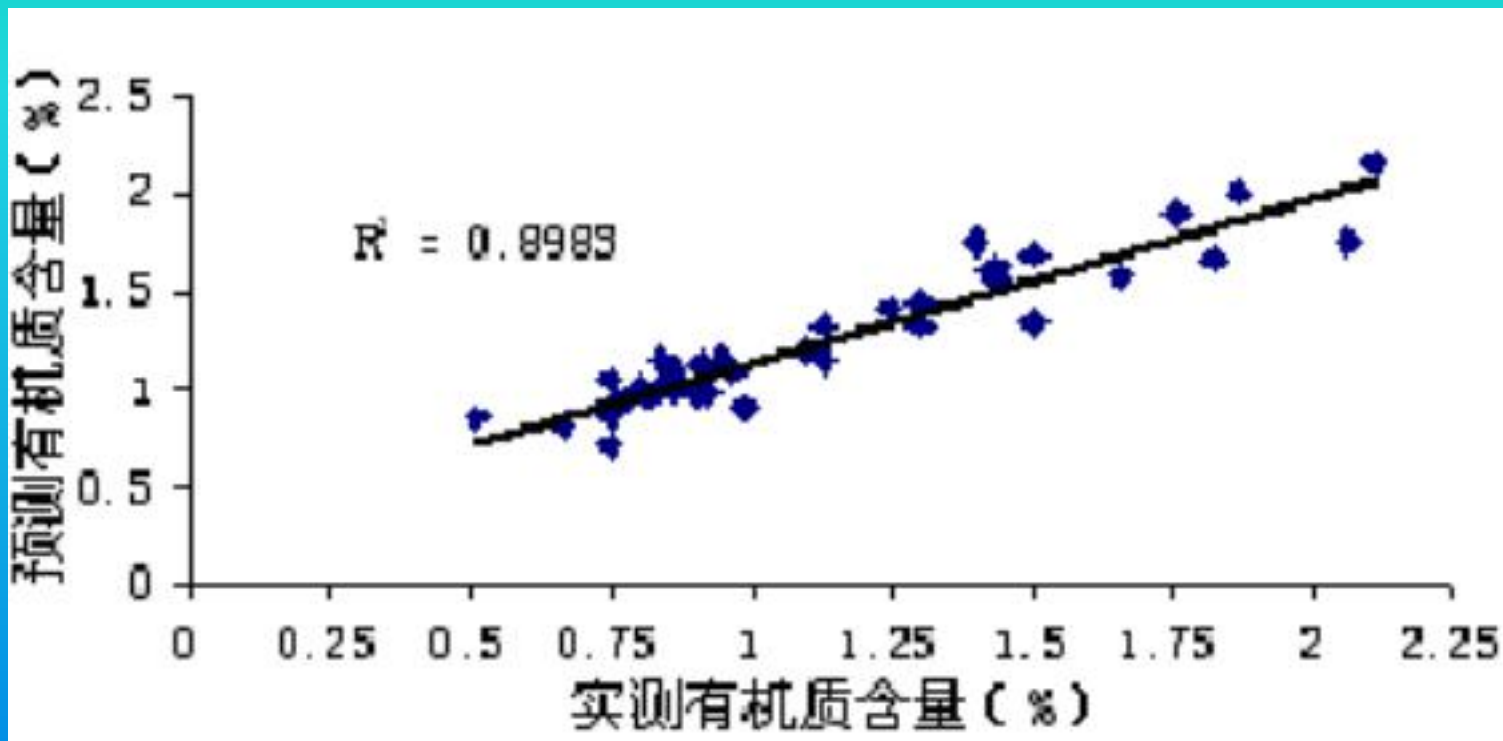
$$SEP = \sqrt{\frac{1}{N_p - 1} \sum_i (Y_i - \hat{Y}_i)^2}$$

反演有机质的上述多元线性回归方程的:

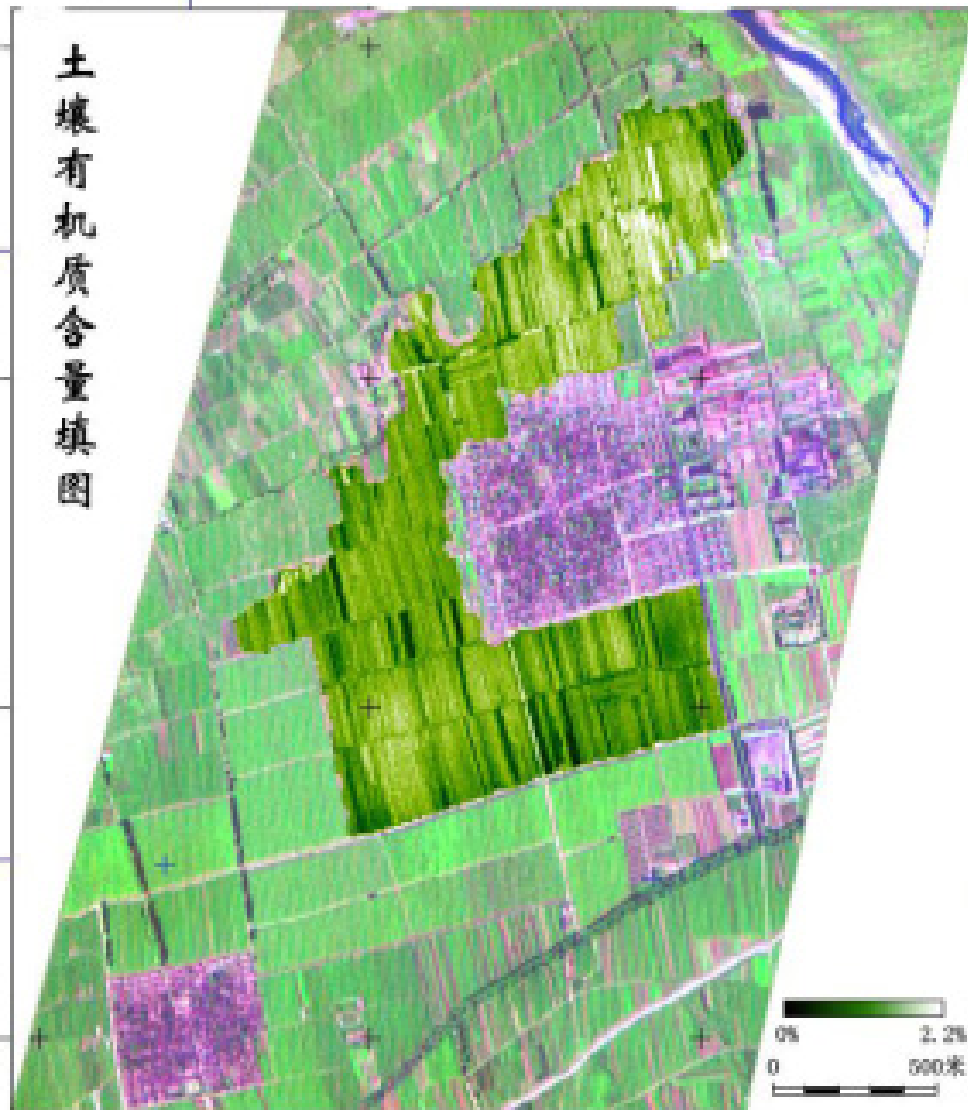
$$SEC = 0.5474、SEP = 0.5426$$



实测值与预测值的拟合分析



土壤有机质含量填图





土壤光谱特征组分定量分析

(土壤有机质含量多模型反演)

多元统计分析、模糊数学分析、主成分分析





德兴铜矿尾矿坝土壤复垦状况

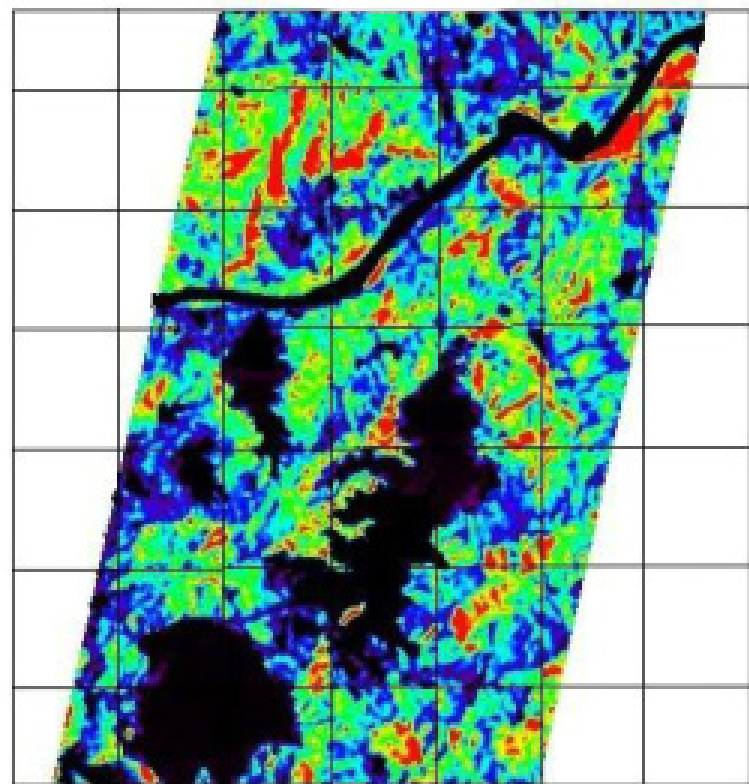
(土壤有机质含量定量反演模型)

德兴铜矿区 google 影像图



研究区位于江西省德兴市境内，是我国最大的斑岩型铜矿之一，为我国大型的露天开采铜矿山。矿区范围大致覆盖北纬 $29^{\circ}02'$ - $29^{\circ}05'$ ，东经 $117^{\circ}43'$ - $117^{\circ}47'$ ，全矿面积约 100 km^2 。自1958年露天开采，其中大规模的开采历史已有50余年（杨修，2011）。

左图为采样点的Google影像图



比例尺 1:100 000



SOM 0

2.53%

图3 基于主成分分析模型的有机质含量填图

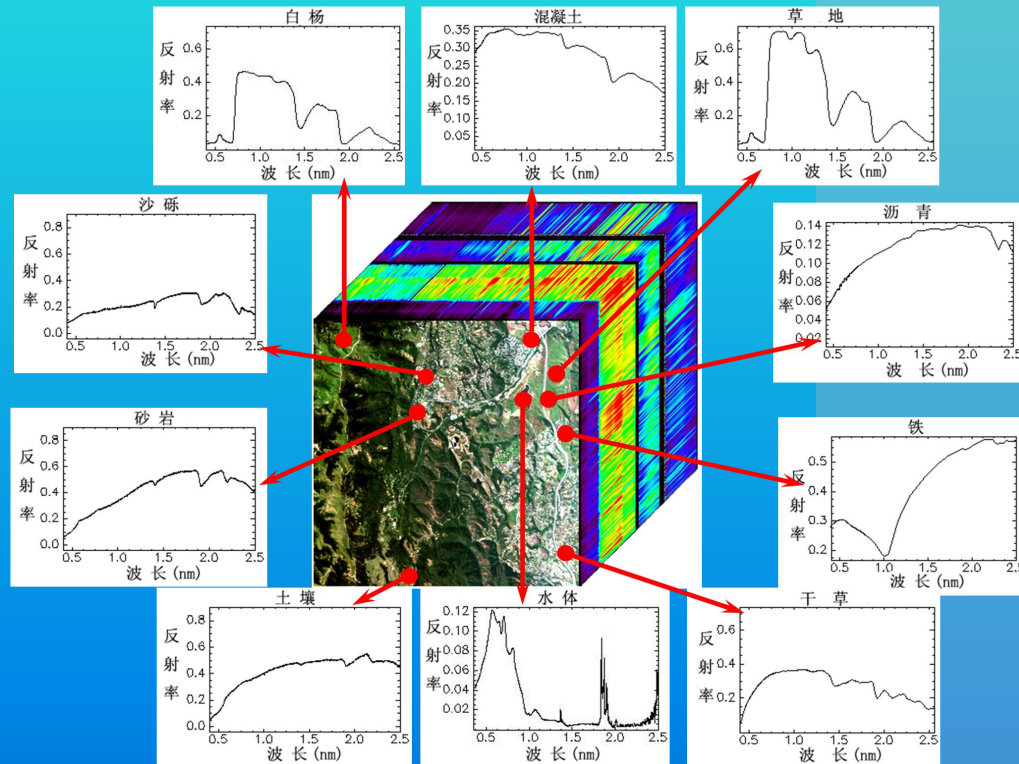
在填图效果上，不同的模型表现出了不同的填图效果，由于模型的建立是基于实地采样数据进行的，并且矿区土壤样品的有机质含量平均值为0.7%，从有机质预测图件上看，模糊数学更接近研究区内有机质含量的平均水平；而纵观整个德兴地区，根据德兴志记载不难得到，整个德兴市的农田有机质含量中等，一般在3%左右，并且研究区内位于德兴铜矿坝附近，原有机质含量较低，因而三幅填图结果在结果上，都是合理的，没有太大的偏差。



土壤光谱特征组分定量分析

基于Hymap数据的土壤重金属As含量遥感调查

作为遥感领域前沿的高光谱数据除了拥有遥感技术的一切优势外，还有“图谱合一”定量特点。而高光谱遥感数据含有的丰富光谱信息，使得研究地球表层物质的类型和成分、丰度时，做到“定性--定量”，使得大范围监测土壤重金属成为可能。





中国地质大学（北京）

研究现状

国外研究现状

利用近红外反射光谱预测了土壤中重金属Fe、Al、Mg、Cu、Zn、Mn等含量，研究表明，近红外光谱反演土壤中重金属含量是可行的。

国内研究现状

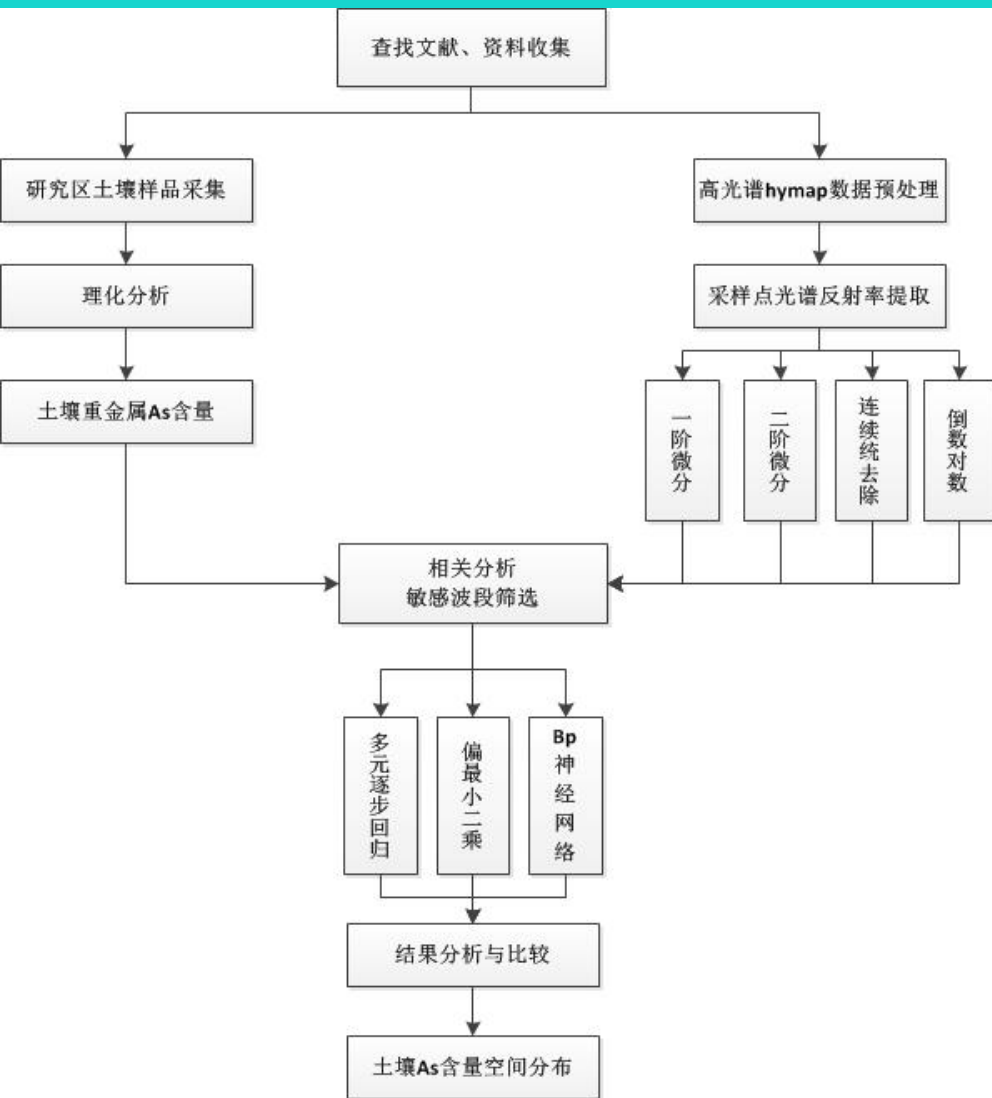
实测土壤ASD高光谱遥感数据和多光谱遥感影像数据预测土壤重金属Cu、Fe、Zn等，找到重金属元素的敏感波段，建模方法单元回归、多元回归、偏最小二乘等。

土壤As研究现状

目前，利用土壤反射光谱预测土壤As元素的研究相对较少。研究数据均为ASD测量的光谱数据。



中国地质大学（北京）



初步探索利用高光谱Hymap影像数据进行土壤重金属As含量的填图



二、研究方法

光谱预处理

连续统去除
一阶微分
二阶微分
倒数对数

建模方法

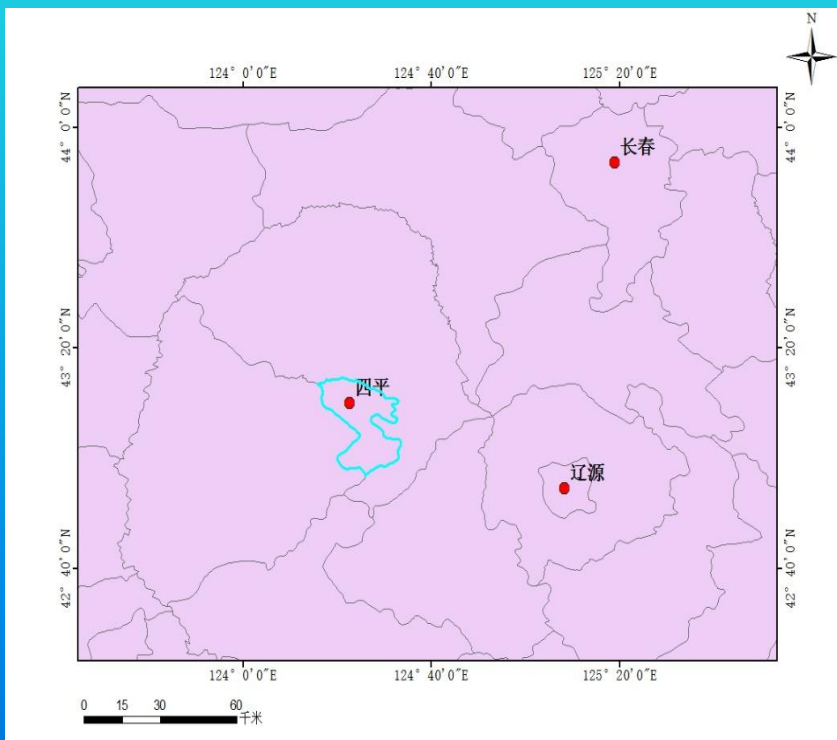
多元逐步回归
偏最小二乘回归
Bp神经网络

精度评价

决定系数(R^2)
均方根误差(RMSE)



三、数据处理



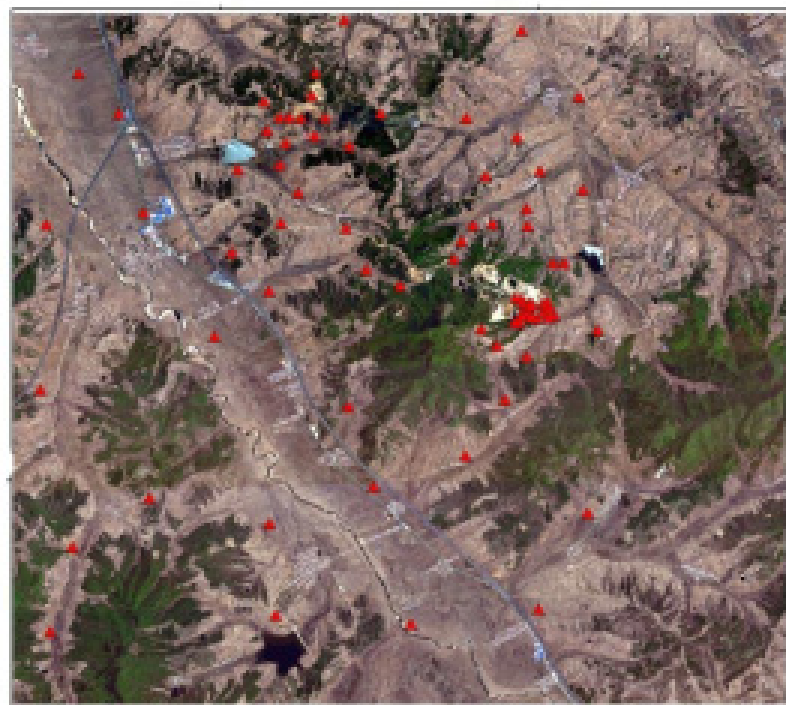
研究区地理位置图

研究区位于吉林省四平市境内的伊通满族自治县。本文研究区处于北纬 $125^{\circ}21'125^{\circ}30'$ ，东经 $43^{\circ}18' \sim 43^{\circ}13'$ 之间，占地面积约为106.5046平方公里，土壤类型为黑甸土，主要种植玉米。



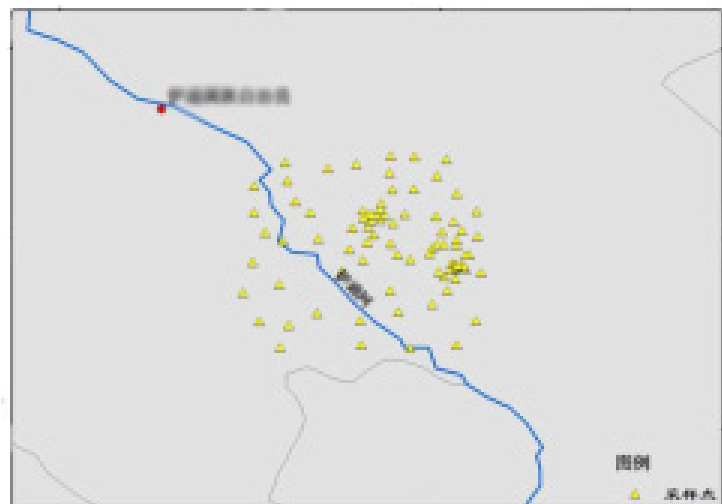
中国地质大学（北京）

土壤砷（As）含量背景值的国家标准：一级土壤环境质量标准规定 ≤ 15 mg/kg；二级标准是 ≤ 30 mg/kg；三级标准是 ≤ 40 mg/kg。研究区土壤砷含量为6.58-419.96mg / kg，其平均值约为土壤背景值的3.23倍，表明研究区的土壤达到了一定污染程度。



图例

▲ 采样点



图例

▲ 采样点



研究区采样过程图

采样地点：四平市伊通满族自治县测区。

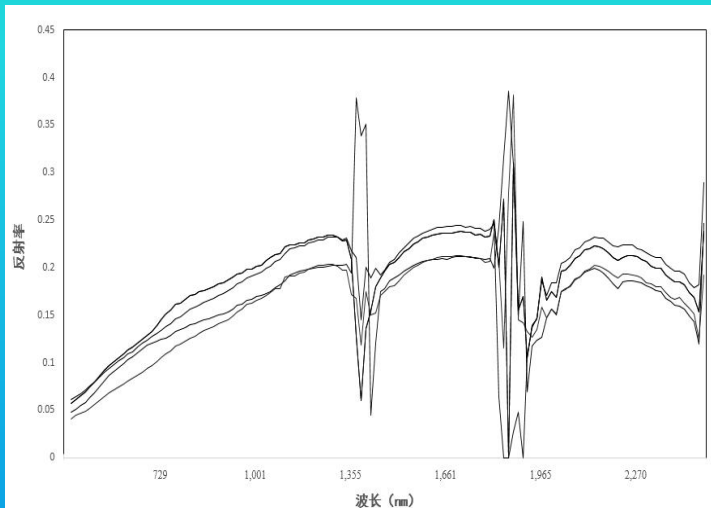
采样时间：2017年4月，阳光充足上午9点到下午4点之间完成采样工作。

采样方法：混合方法进行取样。

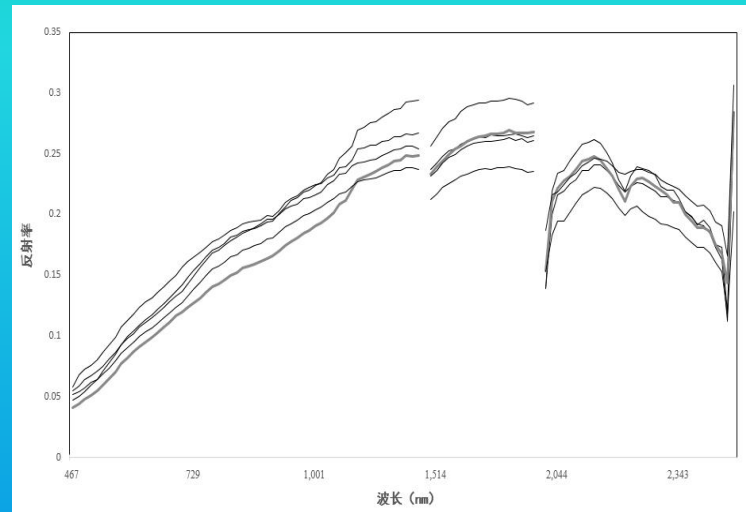
采样工具：铲子、透明样品密实袋、马克笔。



三、数据处理



原始的光谱曲线

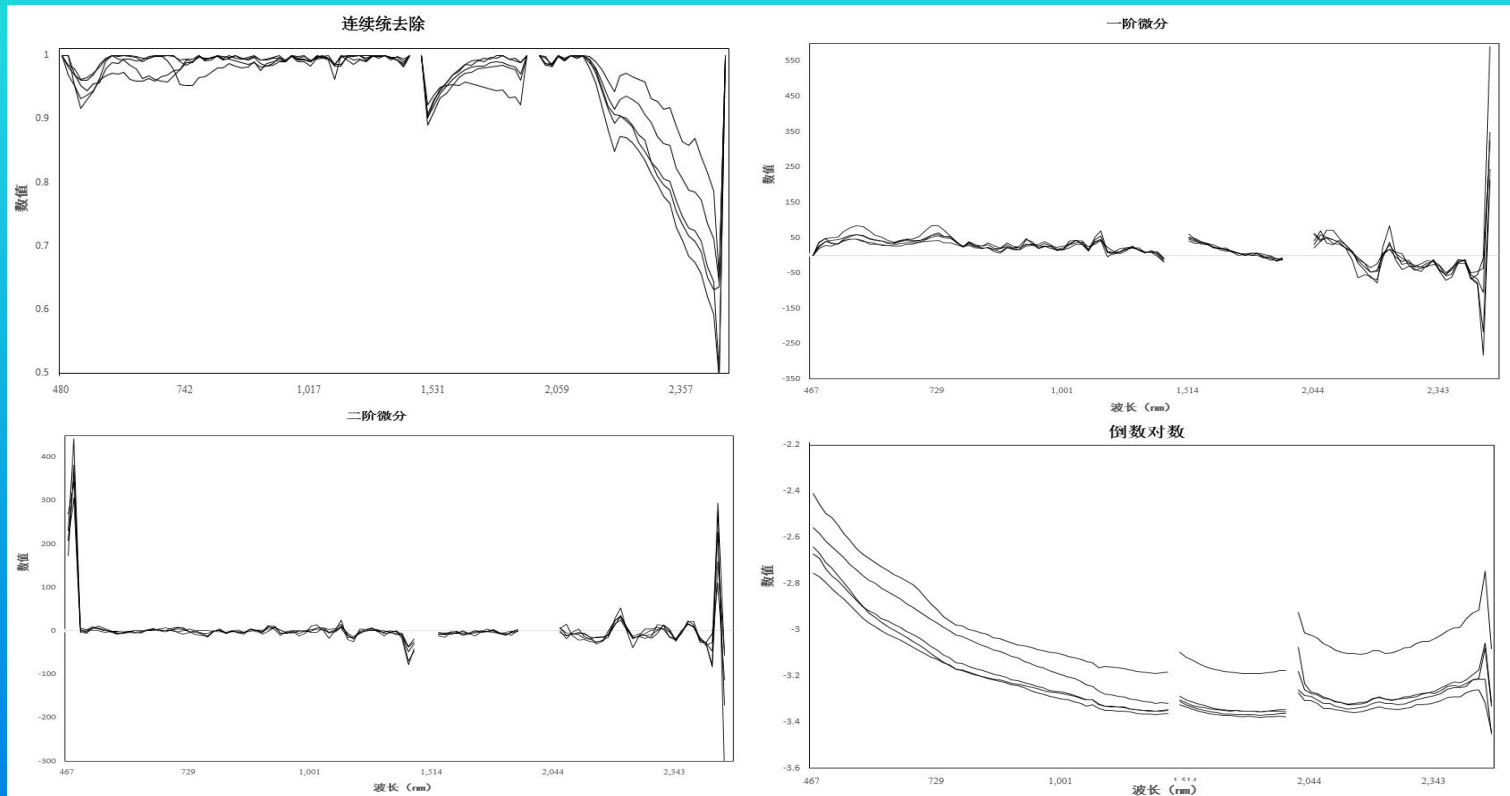


去除水汽波段后的光谱曲线

去除波段范围 $1.32-1.48\mu\text{m}$ 、 $1.78-1.99\mu\text{m}$ ，余下108个波段，土壤反射率的值整体都很低，在0.3以下，在2200nm附近存在一个波谷值，此外很难直观的从图像上找出特征波段，对于后续的研究造成一定程度的困难。



三、数据处理

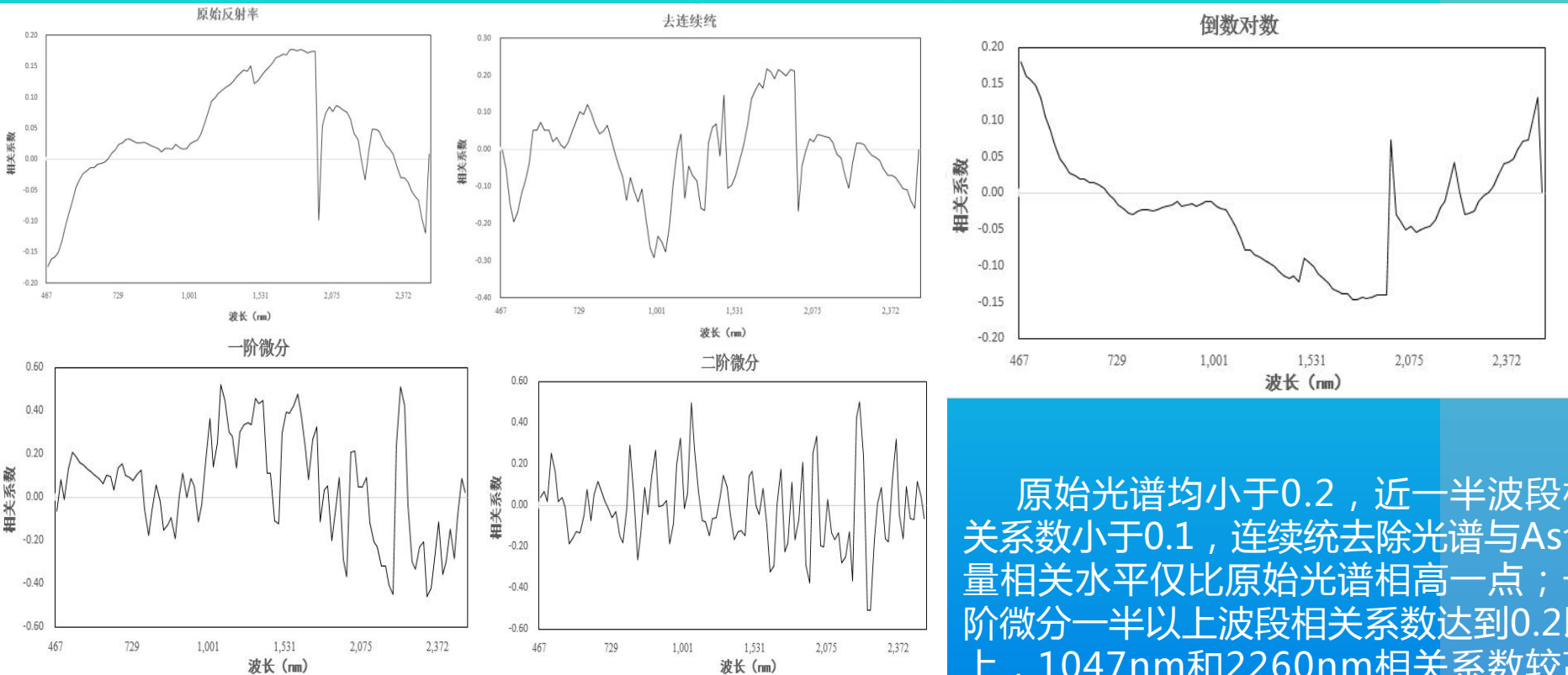


经过预处理后的光谱曲线特征波段也并不明显，大致看出吸收波段的位置，较为明显的吸收带位于520nm、1100nm、2150nm、2210nm、2385nm附近。



中国地质大学（北京）

三、数据处理



原始光谱均小于0.2，近一半波段相关系数小于0.1，连续统去除光谱与As含量相关水平仅比原始光谱相高一点；一阶微分一半以上波段相关系数达到0.2以上，1047nm和2260nm相关系数较高0.5，还有几处超过0.3都属于中等相关，二阶微分的相关性和一阶微分的差不多；倒数对数，整体上与土壤As含量的相关性很低，大部分波段呈现负相关。

土壤As含量与光谱的相关分析



三、数据处理

土壤As含量与土壤特征光谱的相关系数

特征波段 (nm)	反射率	连续统去除	一阶微分	二阶微分	倒数对数
521.85	-0.132	-0.121	0.21*	0.165	0.129
986.40	0.017	-0.233*	0.177	0.324**	-0.012
1047.04	0.041	-0.090	0.52**	0.273**	-0.032
1107.02	0.075	0.042	0.300**	-0.074	-0.062
1660.88	0.177*	0.210*	0.325*	-0.225*	-0.146*
1709.71	0.169	0.209**	0.051	-0.165	-0.145
2150.55	0.065	-0.013	-0.318**	-0.250*	-0.037
2210.67	-0.034	-0.035	0.237*	0.499**	0.042
2240.30	0.048	0.017	0.424**	-0.506**	-0.029
2385.89	-0.037	-0.092	-0.355**	-0.161	0.048

注：*、**分别表示显著相关($p < 0.05$)和极显著相关($p < 0.01$)



四、模型建立

根据不同的光谱数据源，分为以下两种进行建模：

1. 基于挑选的特征波段建模：

根据前面分析可以得出521.85nm、986.40nm、1047.04 nm、1107.02nm、1660.88nm、1709.71nm、2150.55nm、2210.67nm、2240.30nm、2385.89nm这10个波段的反射率以及相应的变换光谱作为模型的输入参数

2. 基于相关性高的波段建模：

Hymap影像提取的土壤反射率光谱曲线不存在明显的特征，不像铜、铁元素有明显的吸收特征带，因此结合相关系数进行筛选，剔除一部分相关性很低并且极其不显著的波段挑选后的结果：原始反射率68个波段；连续统去除73个波段；一阶微分数据剩余76个波段；二阶微分删除73个波段；倒数对数剩余70个波段。



四、模型建立

1.多元逐步回归分析：

基本思路：考虑所有变量对回归方程贡献值的大小，按照其重要性对每一次进入方程中的变量进行显著性检验，然后剔除不显著的变量。新变量的加入可能改变之前变量的显著性，剔除的变量也有可能因为新变量的加入而变得显著，此时把它再加入方程，直到没有可以剔除和再次加入的变量时结束计算。

将78个土壤样本按照As含量高低排序，等间隔抽取18个样本作为检验样本，其余样本用来建模。原始的光谱反射率和As含量实测值这组数据没有变量进入回归方程，说明自变量对因变量数据没有显著影响，而经过变换后的光谱数据均可进行建模。

1.多元逐步回归分析：

土壤As含量的多元逐步回归光谱模型结果

	光谱模型	建模R ²	建模RMSE	检验R ²	检验RMSE	F统计量
特征波段	连续统去除	0.135	71.629	0.030	63.523	9.042
	一阶微分	0.479	54.702	0.295	50.538	14.380
	二阶微分	0.416	58.852	0.237	51.927	13.296
	倒数对数	0.300	64.433	0.221	56.691	12.213
相关性高的 波段	连续统去除	0.474	55.849	0.033	73.268	7.091
	一阶微分	0.741	39.203	0.480	50.519	18.226
	二阶微分	0.700	42.174	0.434	48.272	14.881
	倒数对数	0.538	52.329	0.390	48.916	10.606

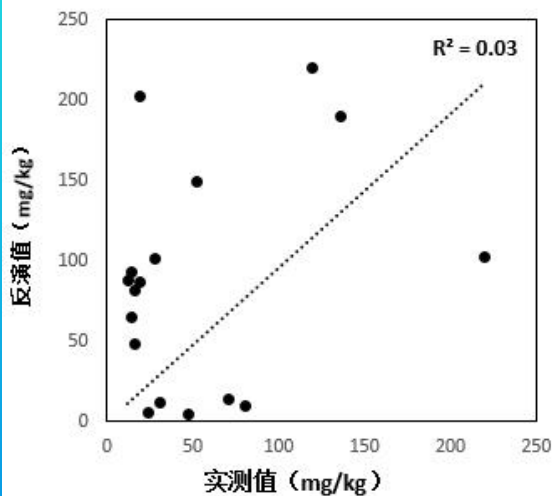
R²最大、RMSE最小和F统计量最高是最佳回归模型



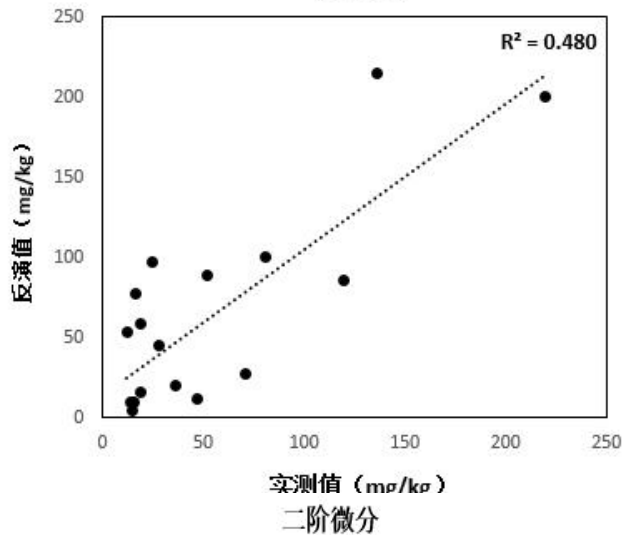
四、模型建立

1.多元逐步回归分析：

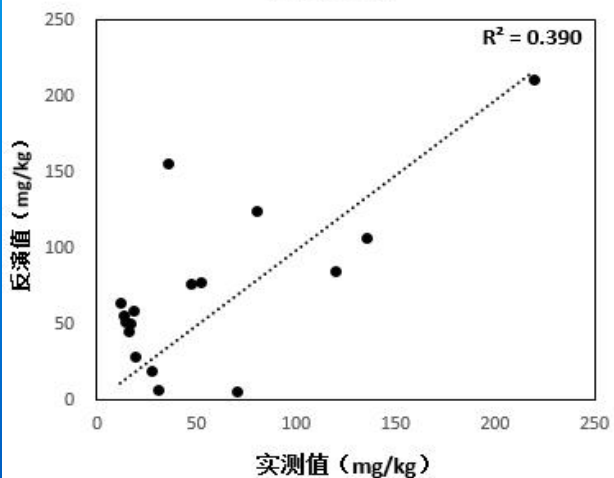
连续统去除



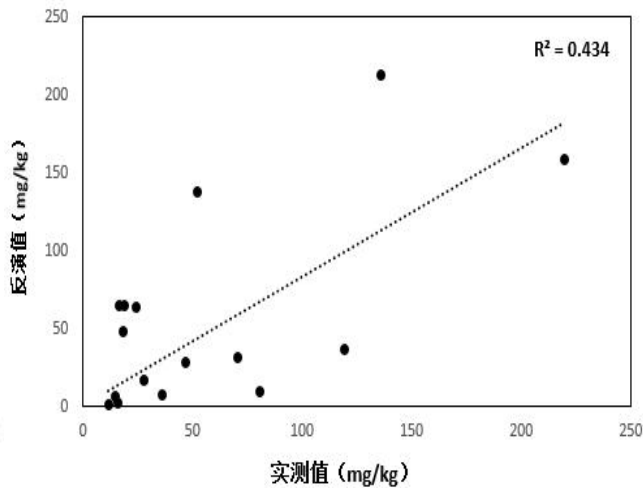
一阶微分



倒数对数



二阶微分



As含量多元逐步回归
反演值与实测值的散
点图



四、模型建立

2. 偏最小二乘分析：

偏最小二乘回归（Partial least squares regression, PLSR）它结合了多元线性回归、典型相关分析以及主成分分析的特点，而被称之为第二代回归分析方法。当建模数据变量个数很多且相互之间存在多重相关性时，偏最小二乘法就能够解决这种存在多重相关性且建模样本数量较少的问题。偏最小二乘法具有的特点能够简化高光谱数据的结构，也能够改善高光谱数据的信息冗余问题，对于多元逐步回归无法解决的光谱波段多重相关性问题也可以解决，利用这种方法建立的高光谱模型，可以包含更多的光谱信息并且增强对重金属的解释能力。



四、模型建立

2. 偏最小二乘分析：

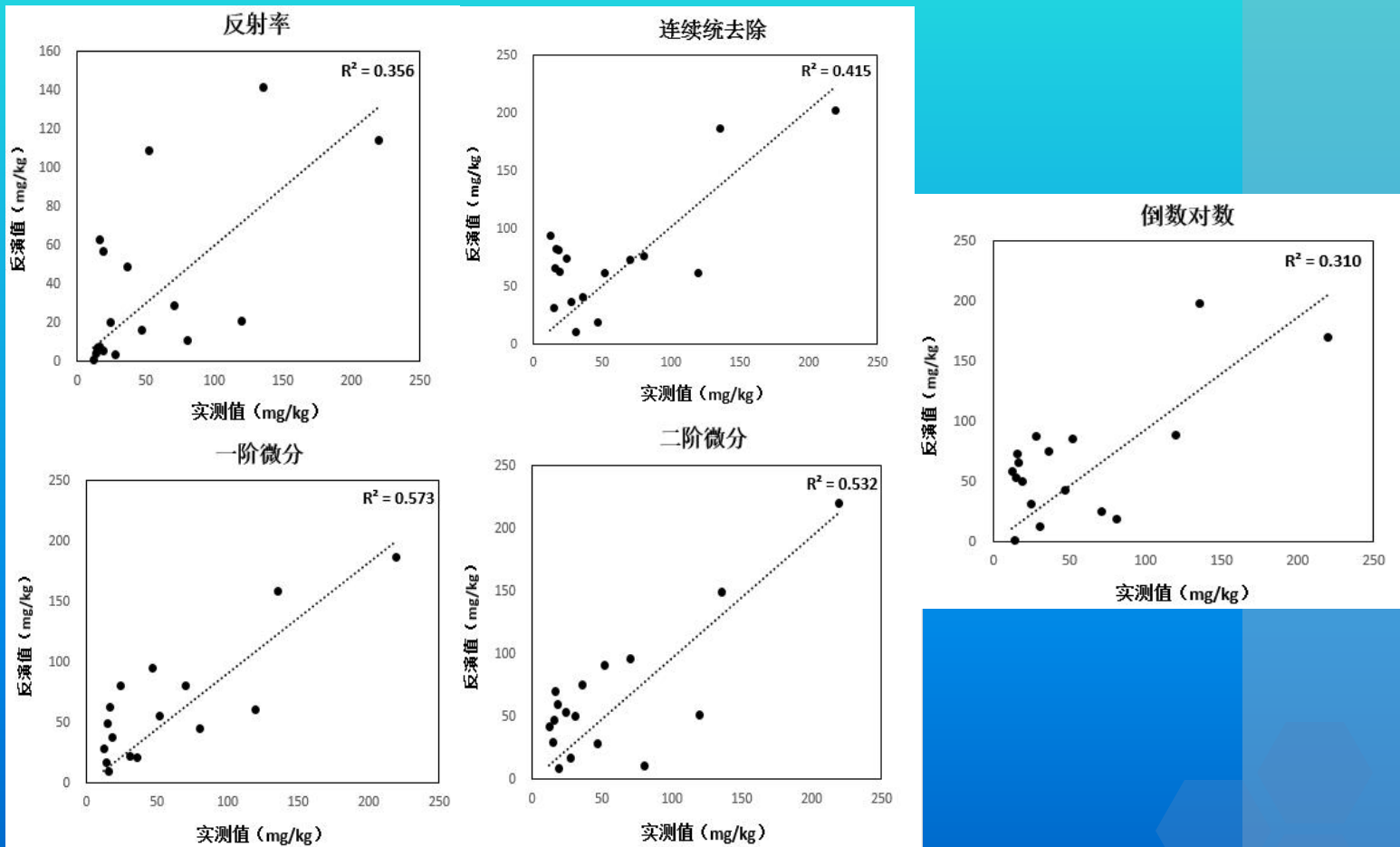
土壤As含量的偏最小二乘光谱模型结果

光谱变量	主成分个数	建模R ²	建模RMSE	验证R ²	检验RMSE
反射率	8	0.606	48.349	0.356	58.893
连续统去除	4	0.683	43.367	0.415	47.005
一阶微分	6	0.702	42.047	0.573	41.097
二阶微分	3	0.685	43.250	0.532	38.082
倒数对数	5	0.681	43.518	0.310	64.067



四、模型建立

2. 偏最小二乘分析：



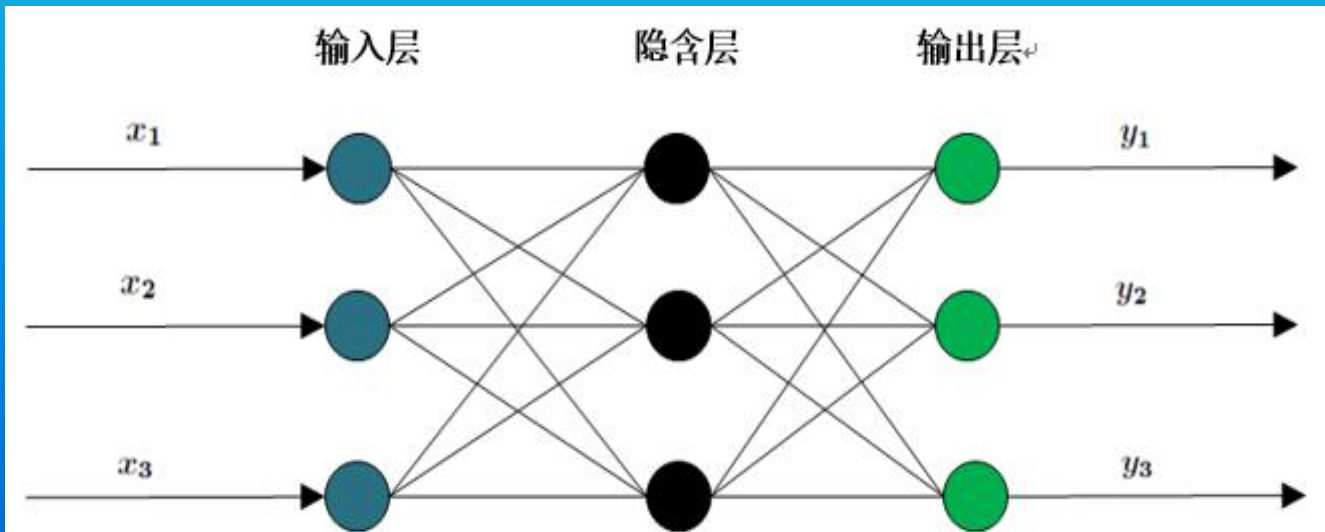
土壤As含量偏最小二乘模型反演值与实测值散点图



四、模型建立

3. Bp神经网络：

Bp神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络。其算法原理是利用模型输出的误差直接返回到前一层误差，再将前一层的误差向更前一层传递，从而获得每一层的误差估计，这一过程会一直循环，直至网络模型的输出误差减小到我们能够接受的范围，这一系列操作称之为训练（学习）的过程。



三层Bp神经网络结构示意图



四、模型建立

3.Bp神经网络：

参数设置

隐含层节点个数根据公式

$h = \sqrt{m + n} + a$ 可以确定：h为隐含层节点数目， m 为输入层节点数目， n 为输出层节点数目， a 为1~10之间的常数，可自行调节。

迭代次数：

`net.trainparam.epochs=2000`

允许误差：

`net.trainparam.goal=0.002`

学习速率：

`net.trainparam.lr=0.05`

训练方案	隐含层传递函数	隐含层节点数
1	logsig	10
2		12
3		14
4		16
5	tansig	10
6		12
7		14
8		16



四、模型建立

3.Bp神经网络：

土壤As含量的Bp神经网络模型检验结果

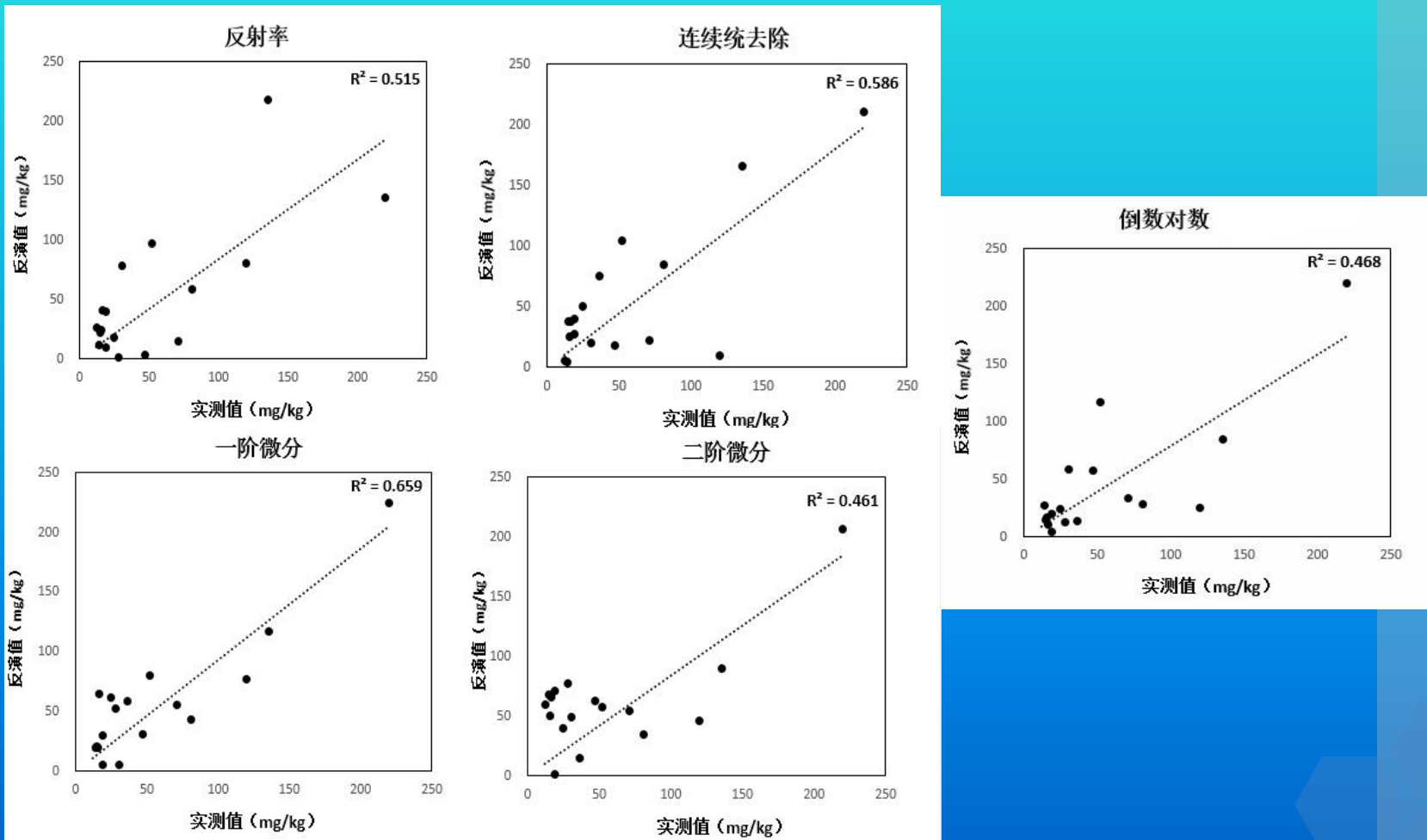
光谱变量	隐含层传递函数	隐含层节点数	建模R ²	建模RMSE	检验R ²	检验RMSE
反射率	logsig	14	0.769	21.295	0.515	41.221
连续统去除	tansig	10	0.779	21.163	0.586	40.143
一阶微分	tansig	10	0.786	19.245	0.659	31.513
二阶微分	tansig	12	0.786	29.241	0.461	42.58
倒数对数	tansig	12	0.747	27.665	0.468	43.635

一阶微分模型的效果最好。



四、模型建立

3. Bp神经网络:



土壤As含量Bp神经网络模型反演值与实测值散点图

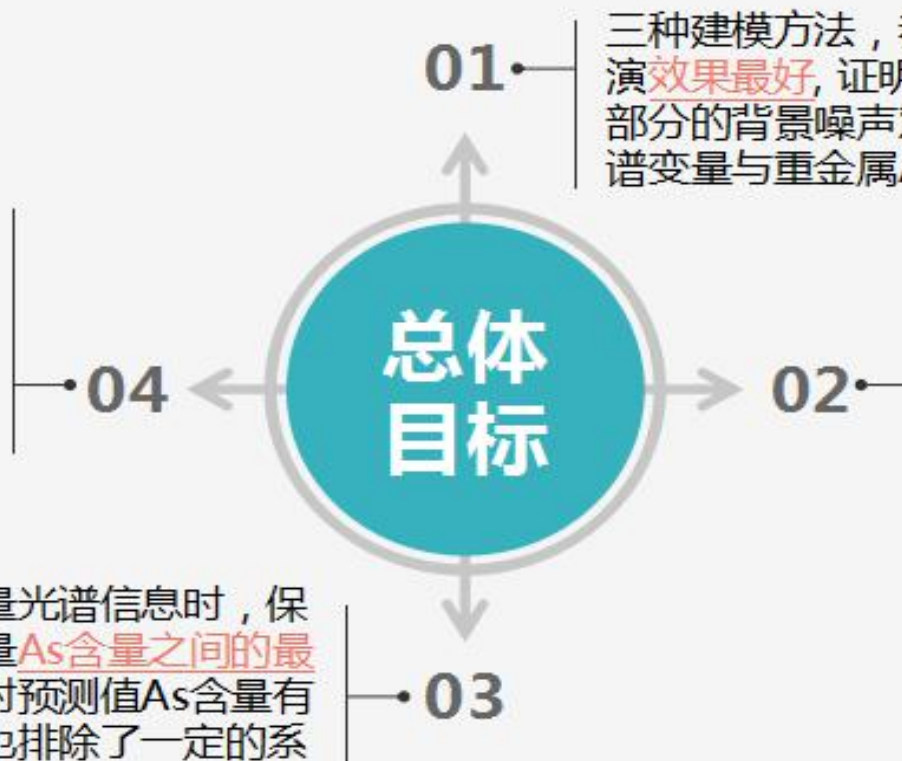


四、模型建立

4. 建模方法对比和分析

Bp神经网络具有很好的自学习、**推广和概括的能力**，且具有一定的容错能力，但其缺点是运算量大，需要不断调整隐含层函数、隐含层节点数，模型的推广能力与训练样本的典型性密切相关。

偏最小二乘提取自变量光谱信息时，保证了光谱变量与因变量**As含量之间的最大相关性**，这些变量对预测值As含量有很好的解释性，同时也排除了一定的系统噪声。



三种建模方法，都是一阶微分建立的模型反演**效果最好**，证明一阶微分变换能较好地去除部分的背景噪声对土壤光谱的影响，增强光谱变量与重金属As的关联，提高建模精度。

多元逐步回归当光谱变量对于建模方程的贡献很低时，如土壤光谱的反射率数据和As含量的**相关性很低，无法建模**。土壤反射率光谱和重金属As含量的关系复杂，不是简单的线性关系。



对问题的思考：

土壤本身成分复杂，土壤中重金属的存在形态并不唯一，而研究中只分析了单一的As的数据，受主、客观影像。如：野外采样状况等。

土壤重金属含量太低，对于土壤光谱几乎没有影响，通过分析土壤的特征波段来估算重金属含量是十分困难的。建模的样本点对于模型的影响很大，存在一定的局限性。

Hymap遥感影像经过预处理后，还是会存在几何校正的误差。搭载Hymap成像光谱仪的飞机飞行速度快，无法保证采样时间和成像时间一致（同步）。

实际的田间情况，土壤反射光谱还会受到太阳光照、大气状况、土壤质地等很多因素的干扰，还需要进行大量的田间土壤光谱处理和分析，才能够建立精确的土壤As元素的高光谱反演模型。



提 纲

- 一、背景意义
- 二、实例分析1
- 三、实例分析2
- 四、几点思考



四、几点思考

土壤生态环境定量化评价几点认知：

一、土壤生态环境指标定量化评价前景光明，但道路曲折

例如：改进的通用**RUSLE**模型不是放之四海而皆准！在用时还要因地制宜。如：中国的南北方差异性，5个因子（植被C）就要有所不同。

我国还应在此做大量实验总结出适于自己的一套经验模型！



二、高光谱技术在土壤质量评价中作用

1. 土壤是复杂的，就其种类、质地结构、成分、环境背景、光谱属性目前很多问题没搞清楚。（光谱机理问题）
2. 普适性模型没有。目前经验和统计模型有局限性。
3. 传感器（光谱分辨率-航天数据）、精度的问题（数据处理、模型建立、野外采样等等）。
4. 与传统的地球化学方法相比精度上还是差些。但高光谱技术用于该领域方向没错，其优势也是显而易见的（大范围区域性的对比）。



中国地质大学（北京）



谢谢