



# 基于地统计学方法的土壤采样设计研究——以有机碳和全磷为例

报告人 傅伟

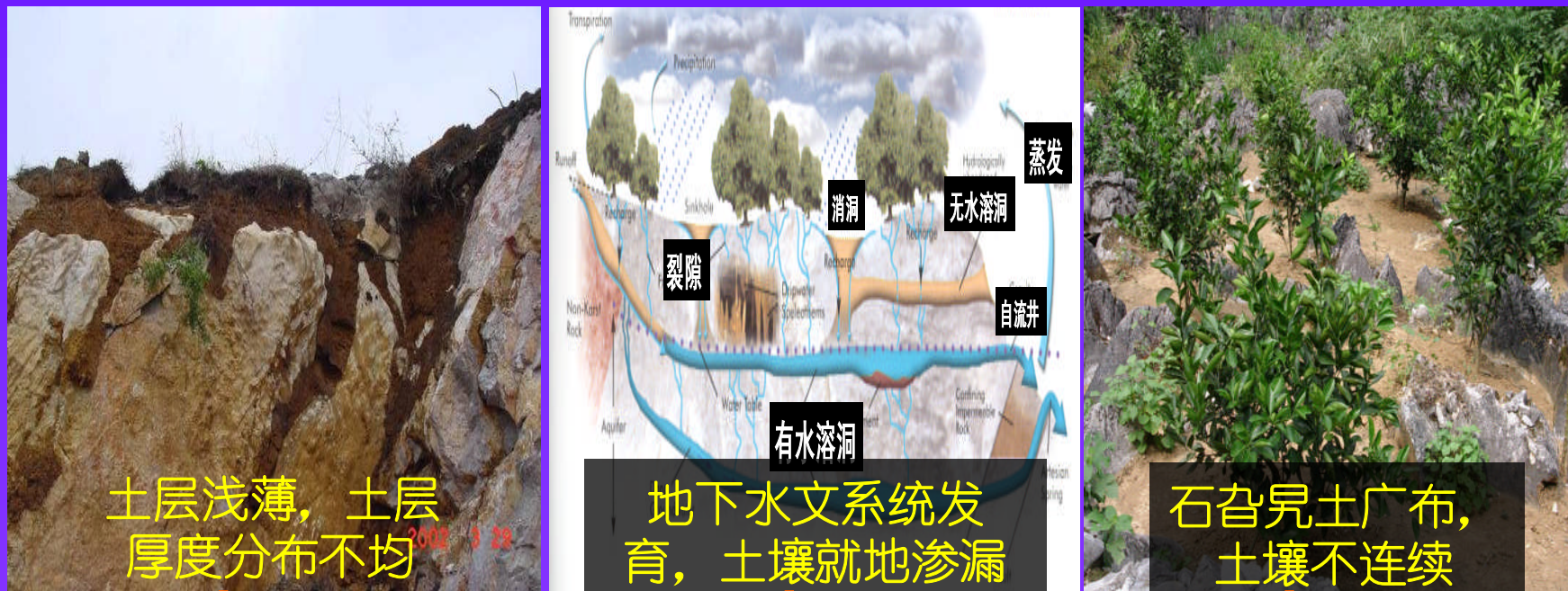
2007年11月

# 内容提要

- 引言
- 研究内容与研究方法
- 结果与分析
- 主要结论

# 引言

喀斯特土壤生态环境的特殊性增加了土壤采样操作的复杂性



代表性土壤样点的选取困难

# 引言

因此，在喀斯特地区的土壤养分研究中，采样样本数的确定以及样点在采样区内如何布设，成为重要的讨论内容。

采样数的确定

经典统计学样本估算法

样点布设

随机布点

规则网格布点

四叉树布点

# 引言

本研究在较为详尽的前期土壤采样调查的基础上，计算了土壤全磷和有机碳在不同期望精度及不同置信水平下的合理采样数目。最后，比较了四叉树布点、规则网格布点和随机布点的插值精度，评价了三种样点布设方案在固定样本数目下采样效果。

# 研究内容与研究方法

## 研究内容

- 土壤养分空间变异特征
- 土壤养分采样设计
  - 1、土壤样本容量
  - 2、样点的布设研究

# 研究内容与研究方法

## 研究区概况

海拔：376m~816m

平均气温：16.5~20.5℃

平均降雨量：1389.1mm

平均土层深度：80~100 cm(洼地)

20~30 cm(坡地)

森林覆盖率：6%

# 研究内容与研究方法

## 研究区概况

### 土地利用方式:

96年前: 玉米、大豆

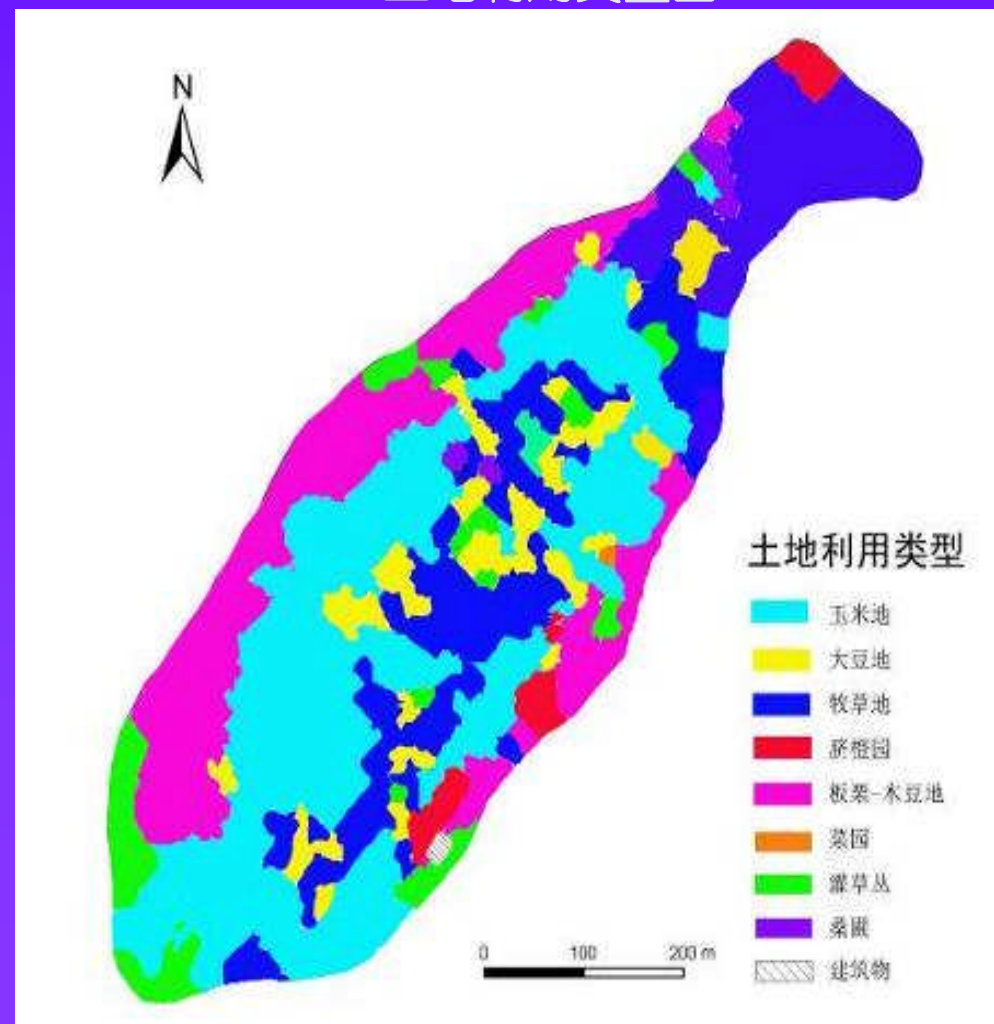
96年后: 玉米、牧草、大豆、木豆-板栗、脐橙、桑苗等

### 施肥管理:

玉米、大豆: 翻耕, 农家肥、人粪尿、草木灰、尿素、钙镁磷

牧草: 30~40天收割, 施肥与玉米地相似, 但不施人粪尿和草木灰

土地利用类型图





# 研究内容与方法

## 采样方法

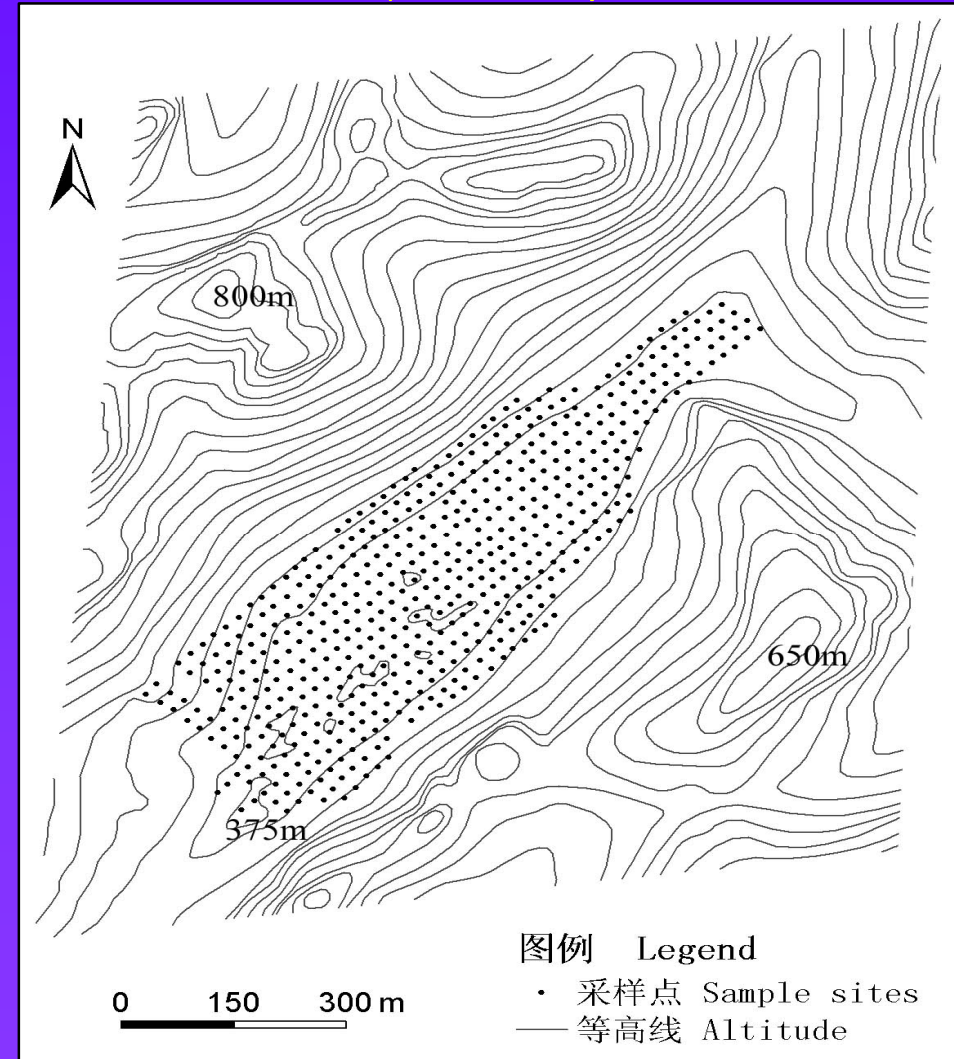
洼地：网格取样 ( $20 \times 20\text{m}$ )

528个样点

定点方法：GPS+测绳

样本采集：随机采取5个样本 ( $2 \times 2\text{m}$ )

样点分布图



# 研究内容与研究方法

## 分析指标

- 全磷：NaOH熔融——钼锑抗比色法
- 有机碳：重铬酸钾氧化—外加热法测定

## 数据处理

- 地统计半变异函数，克里格插值
- 经典统计学确定采样样本数
- 比较随机布点、规则网格布点、四叉树布点的采样效果

# 研究内容与研究方法

## 地统计学基本原理

### 半变异函数

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

半变异函数是地统计学的基本计算公式，也是描述区域化变量空间变异特征的主要工具之一。

半变异函数是距离的函数，描述土壤性质在空间上的差异程度。当区域化变量在采样尺度内存在空间自相关时，半变异函数可以表征空间变量变异尺度和程度。

# 研究内容与研究方法

## 地统计学基本原理

### 半变异函数模型的拟合及基本参数的意义

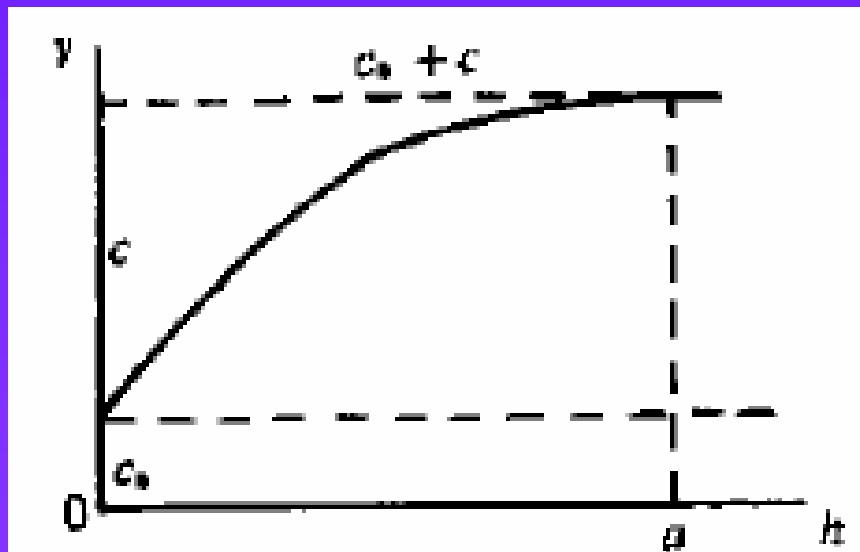


图1 半方差图的结构示意图

➤基台值( $C_0+C$ ): 表示系统属性最大的变异;

➤块金值( $C_0$ ): 表示系统属性随机的变异, 由采样尺度及系统属性本身变异特征控制;

➤变程(Range): 表示系统属性空间自相关的范围。

# 研究内容与研究方法

## 地统计学基本原理

### 空间自相关分析

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left( \sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (x_i - \bar{x})^2}$$

$x_i$ 和 $x_j$ 分别为变量 $x$ 在相邻配对空间点 $i$ 和 $j$ 上的取值。 $w_{ij}$ 是相邻权重。在进行空间自相关分析时可以计算不同滞后距( $h$ )上的自相关系数 $I$ 并绘制其 $h$ 展点图,从而反映变量的空间结构特征。

# 研究内容与研究方法

## 地统计学基本原理

### 异常值剔除

比较特定点与邻域内其它点的差异（相关）程度，来判别该点是否异常点。例如，Moran's I系数检测法：

$$I_i = \frac{Z_i - \bar{Z}}{\sigma^2} \sum_{j=1, j \neq i}^n [w_{ij} (Z_j - \bar{Z})]$$

$$w_{ij} = \frac{1}{d_{ij}^2}$$

$Z_i$ 、 $Z_j$ 表示系统属性在*i*点和*j*点的取值

$\sigma^2$ 表示区域化变量*Z*的方差

$w_{ij}$ 是*i*点和*j*点的距离权重

$d_{ij}$ 表示空间上*i*点和*j*点的距离

# 研究内容与研究方法

## 地统计学基本原理

### 异常值剔除

Moran's I系数可以转化为z值，基于正态分布的Moran's I系数的显著水平可进行检验。z值得计算公式如下：

$$z_i = \frac{I_i}{\sigma} \sqrt{n}$$

$z_i$ 为i点的z值

$\sigma$ 为研究区I系数的标准差

n为样本点数

之后，根据5%置信水平的z取值，判断取舍相应的样点值。

# 研究内容与研究方法

## 地统计学基本原理

### 异常值剔除

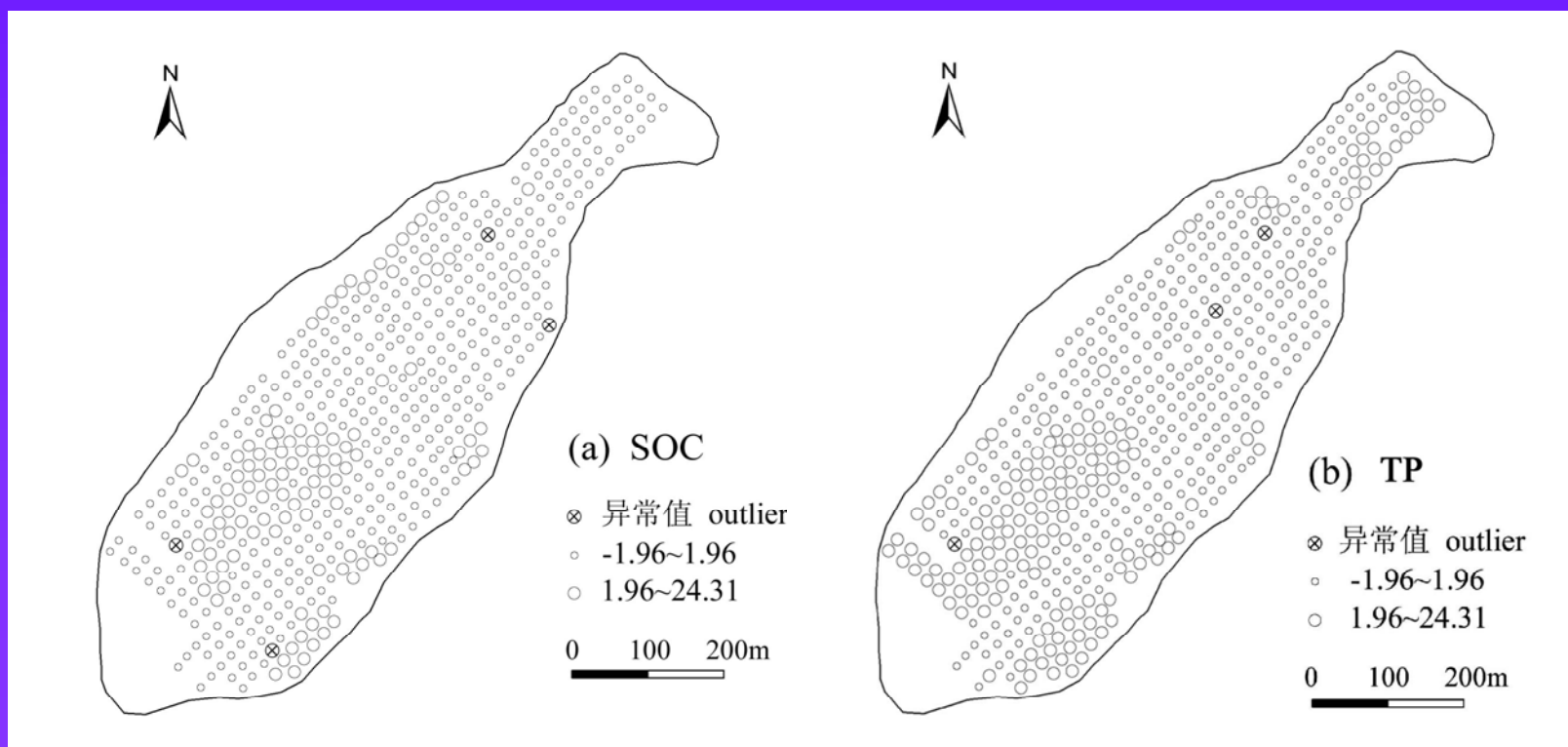


图2 SOC和AP异常值空间分布图

Fig. 2 Spatial symbol maps of SOC and AP outliers



# 研究内容与研究方法

## 地统计学基本原理

### 克立格 (Kriging) 插值

$$Z^{\#}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

在普通克立格插值中，对于任意待估点（块段） $x_0$  的实际值  $Z(x_0)$ ，其估计值是通过该待估点（块段）周围的  $n$  个有效样点  $Z(x_i)$  的线性组合得到。

# 研究内容与研究方法

## 合理采样数的确定

### 合理采样数：

在总体中抽出一定量的样本，用所抽样本的均值与方差能较好地估计总体的均值与方差，即抽取样本的均值具有足够的精度和较大概率近似于总体均值。

当所研究的土壤属性在统计学上是正态分布特征时，可构造统计量公式，并推导出采样需要置信限下的合理采样数：

$$\eta = \frac{\bar{x}_n - \mu}{\sqrt{\sigma^2/n}}$$



$$n = \frac{\eta_a^2 \sigma^2}{(\bar{x}_n - \mu)^2}$$

# 研究内容与研究方法

## 几种土壤采样布点法的比较

- **随机布点法：**

将监测单元分成网格并编号，决定采样点样品数后，随机抽取规定的样品数的样品，其样本号码对应的网格号，即为采样点。

- **规则网格布点法：**

把研究区分成大小相等的方格，线交点为采样点，每个采样点由1m范围内的8-10个小样构成混合样。

- **四叉树布点法：**

将研究区四等分并对各区层土壤属性进行方差评价，方差较大的区层不断地进行四等分，直到方差较大值小于指定的阈值为止，每区层内随机选取一个采样点。

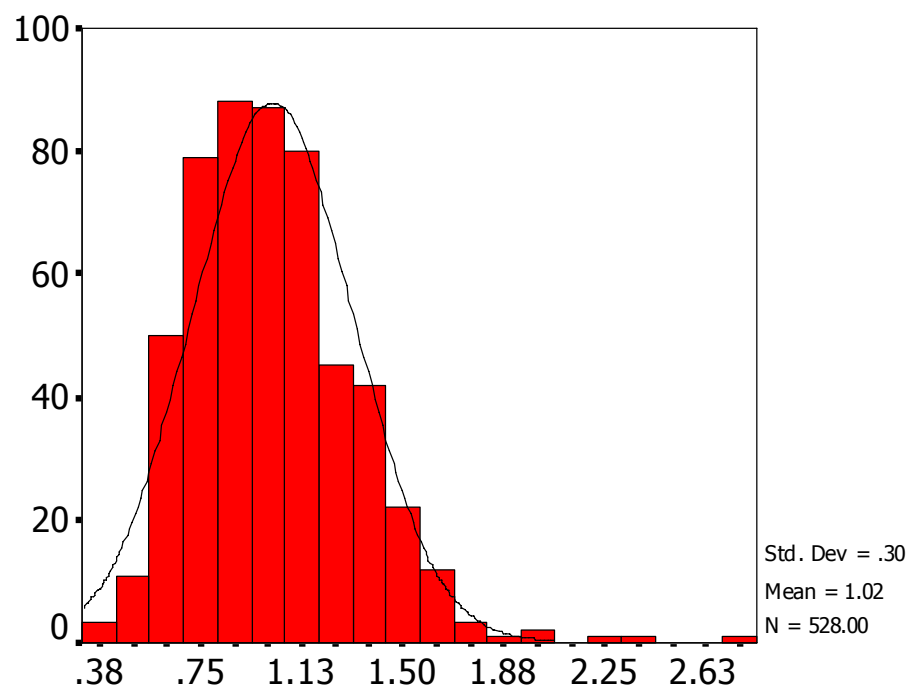
# 研究内容与研究方法

## 几种土壤采样布点法的比较

	优点	缺点
随机布点法	土壤采样操作简便、不带主观限制条件	没有考虑土壤的空间自相关性，当样本总体呈偏态分布时，增加错误估计的概率
规则网格布点法	定量描述区域化土壤属性变异的 空间尺度和模式	采样精度受样本数量的限制
四叉树布点法	采样设计依据土壤变异程度或密 或疏布点，提高采样效率	需要有研究区土壤特性的先 验空间分布信息，对空间分布 特征未知的田块的首次采样 设计不适合

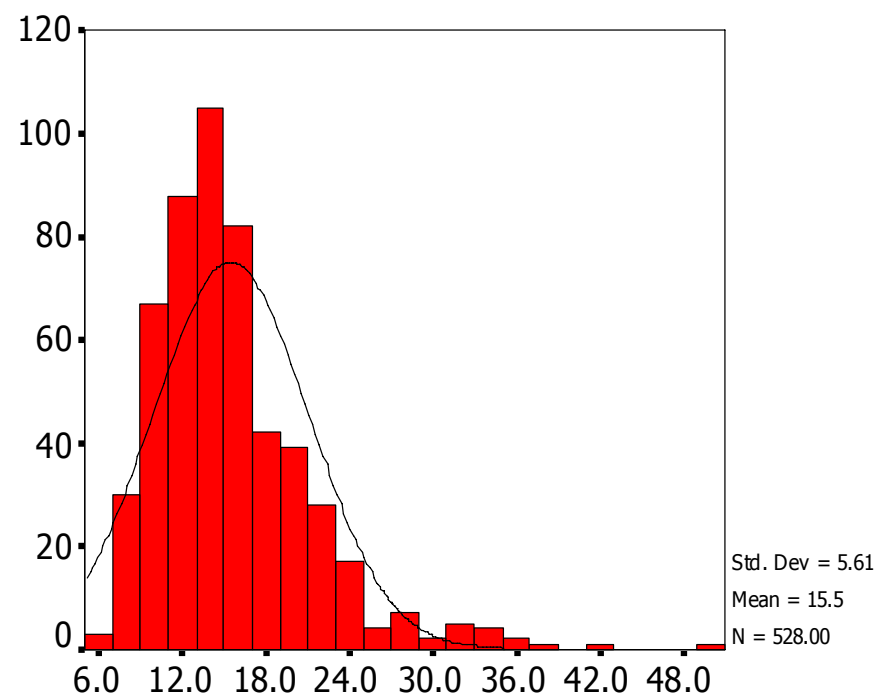
# 结果与分析

## 土壤全磷、有机碳的统计分析



TP

近似正态分布



SOC

非正态分布

# 结果与分析

## 土壤全磷、有机碳的统计分析 with 合理采样数的确定

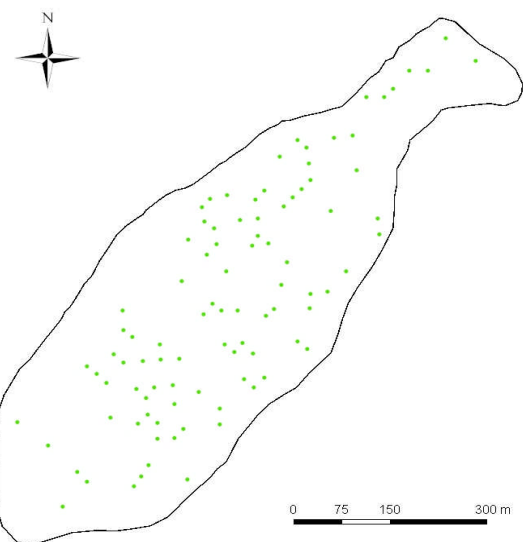
	均值	方差	变异系数	95%置信水平下的 合理样本数		90%置信水平下的 合理样本数	
				允许误差(5%)	允许误差(10%)	允许误差(5%)	允许误差(10%)
SOC	15.50	31.42	0.36	<b>201</b>	<b>51</b>	143	36
TP	1.02	0.09	0.29	<b>133</b>	<b>34</b>	95	24

# 结果与分析

## 样点布设比较

置信水平90%、允许误差5% 情况下全磷样点布设图(95)

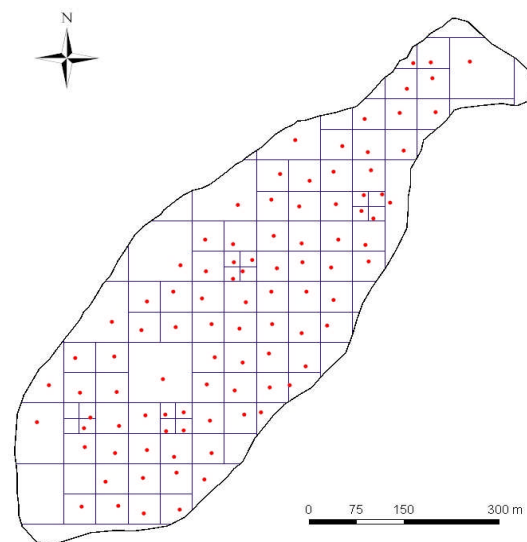
随机采样样点设计图



网格法采样样点设计图



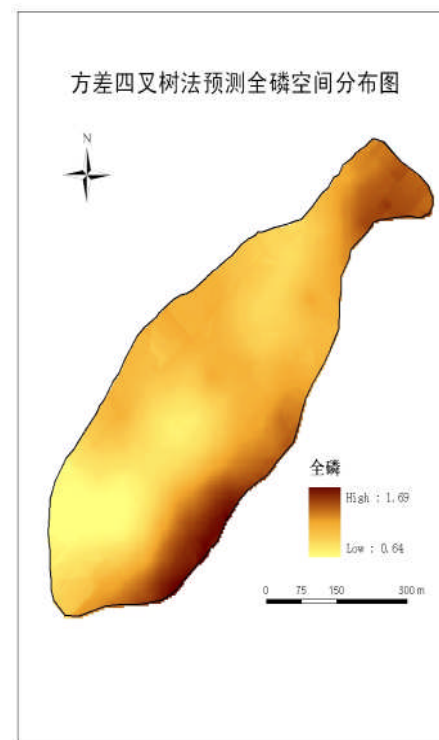
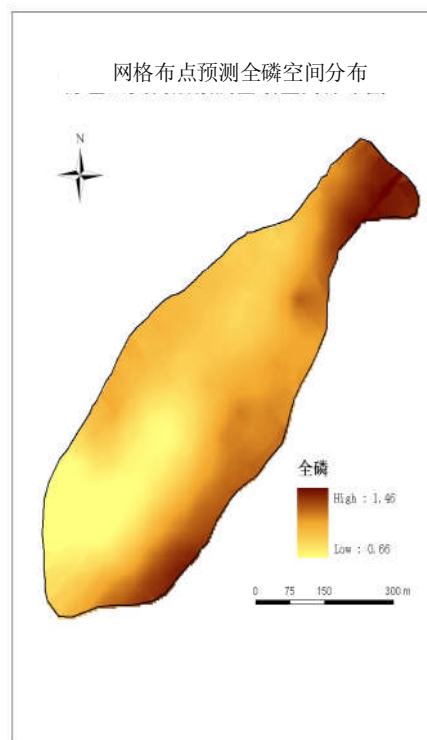
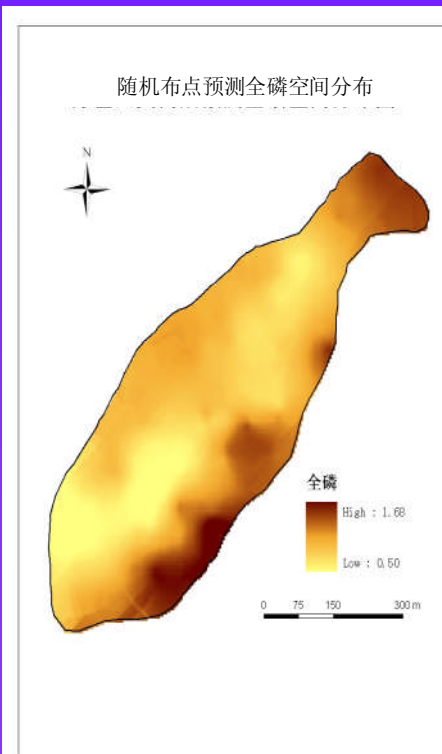
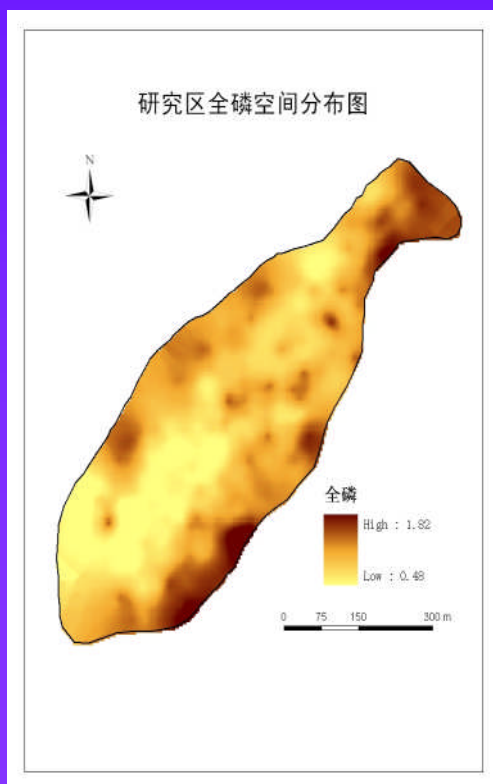
四叉树法采样样点设计图



# 结果与分析

## 样点布设比较

置信水平90%、允许误差5% 情况下全磷插值比较





# 结果与分析

## 样点布设比较

置信水平90%、允许误差10%情况下全磷样点布设图(24)

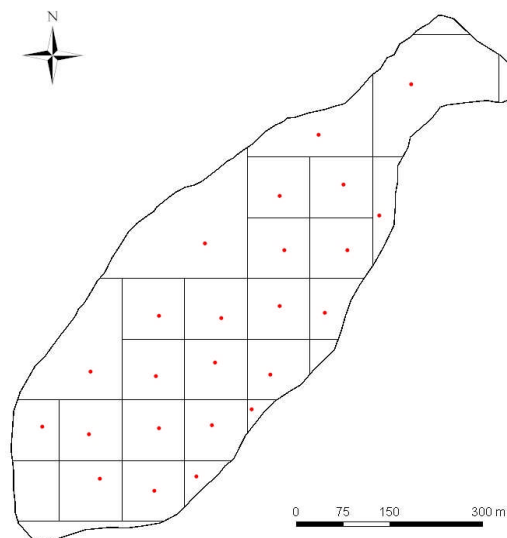
随机采样样点设计图



网格法采样样点设计图



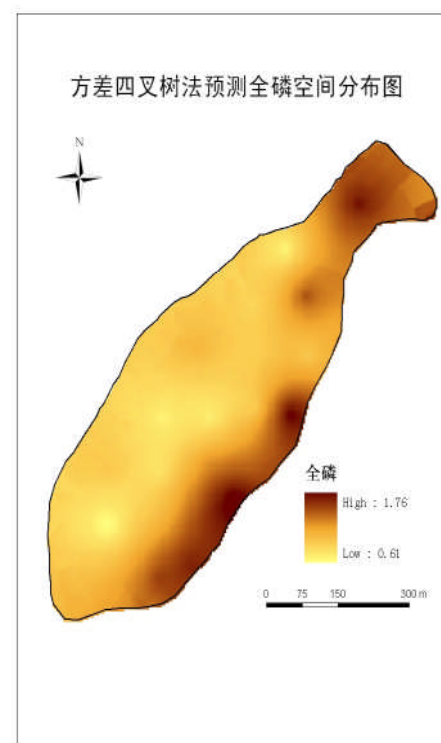
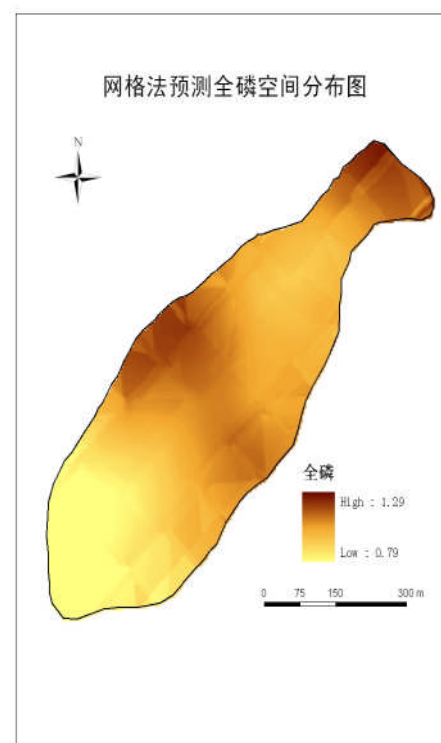
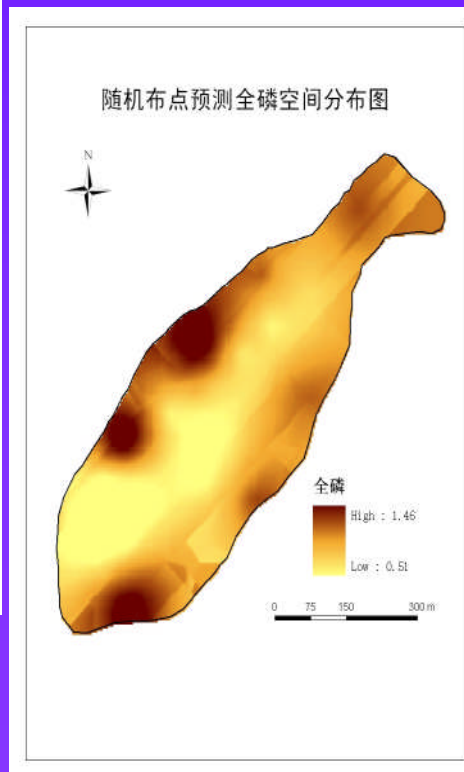
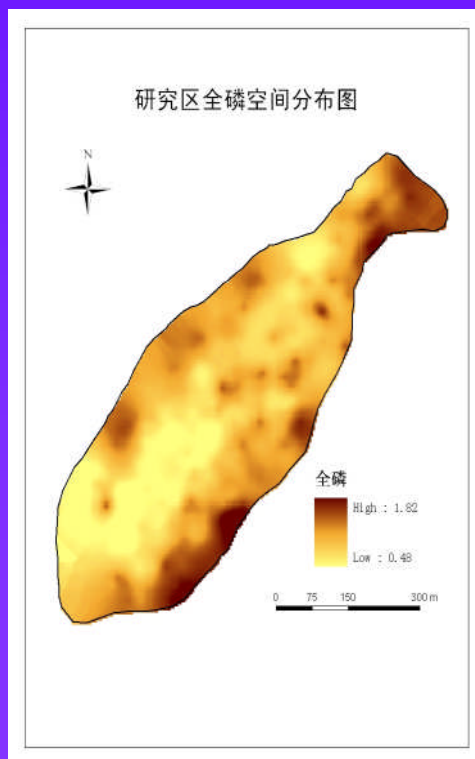
四叉树方法采样样点设计图



# 结果与分析

## 样点布设比较

置信水平90%、允许误差10%情况下全磷插值比较

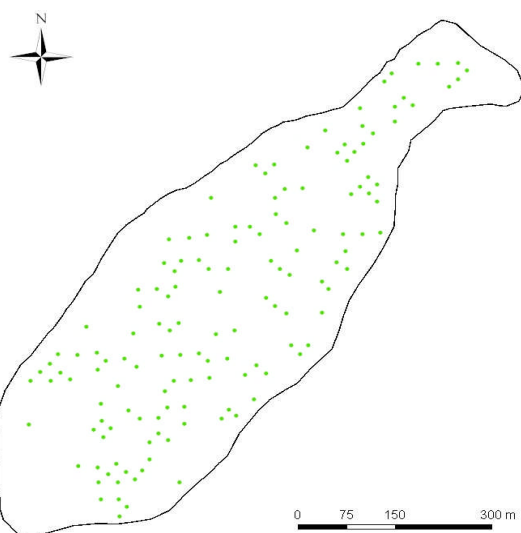


# 结果与分析

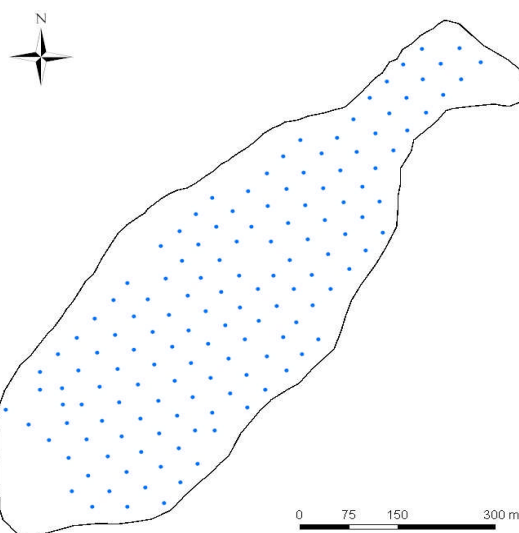
## 样点布设比较

置信水平90%、允许误差5%情况下有机碳样点布设图(143)

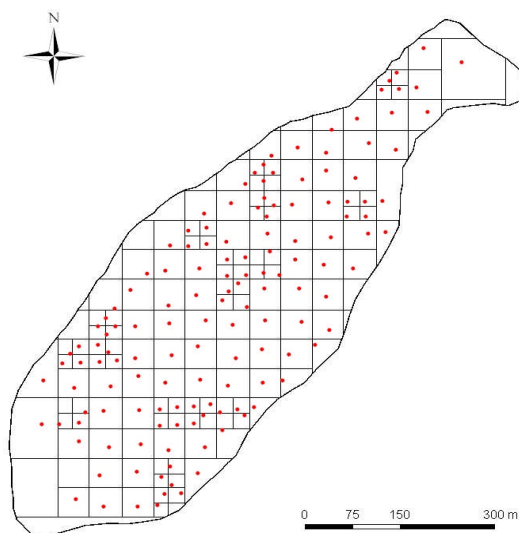
随机采样样点设计图



网格采样样点设计图



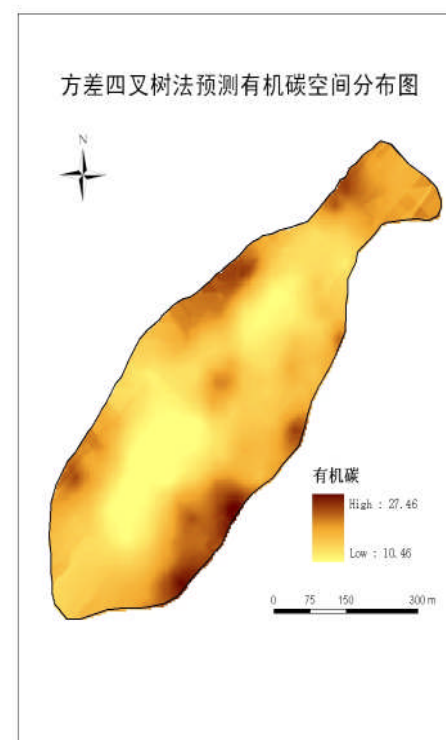
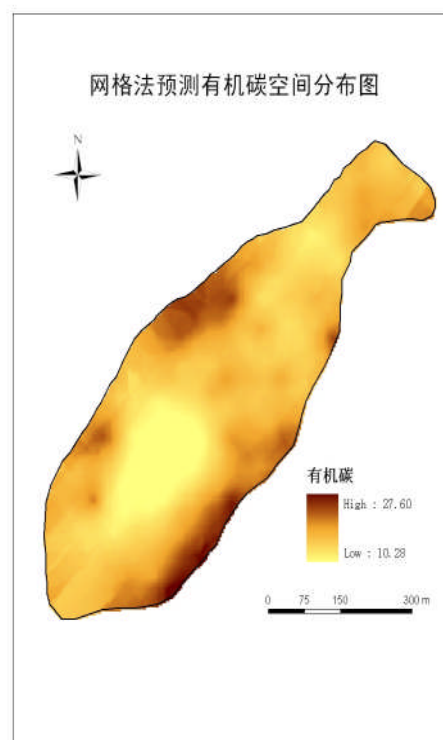
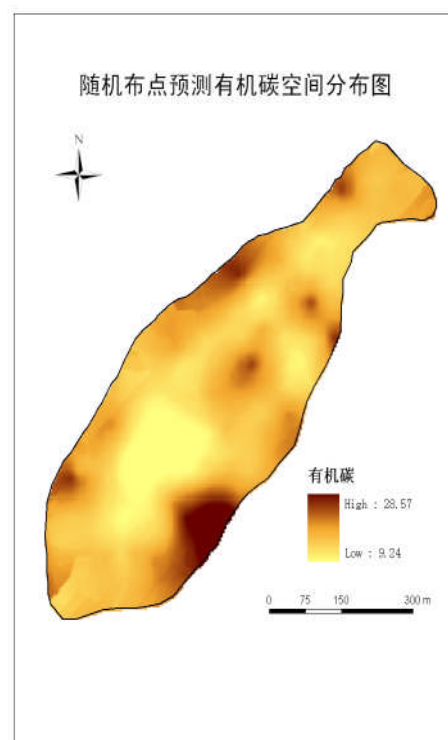
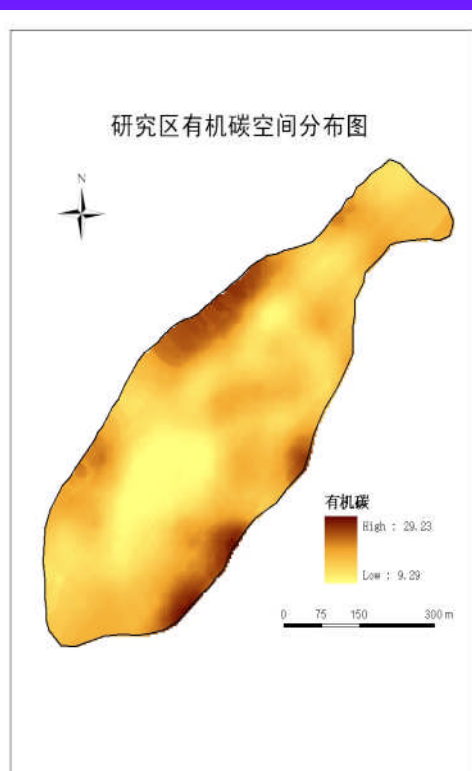
四叉树方法采样样点设计图



# 结果与分析

## 样点布设比较

置信水平90%、允许误差5% 情况下有机碳插值比较

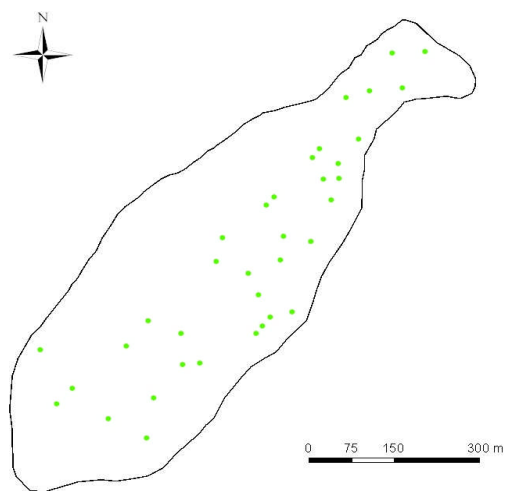


# 结果与分析

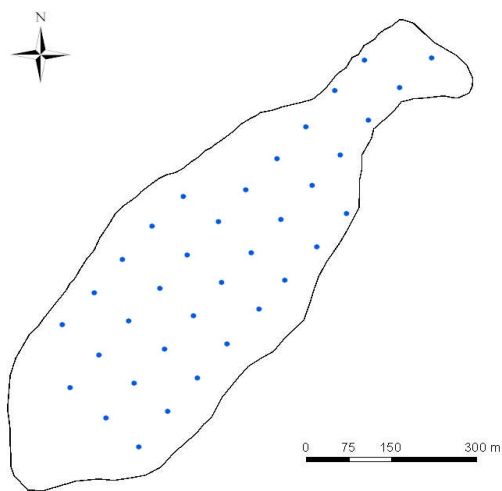
## 样点布设比较

置信水平90%、允许误差10%情况下有机碳样点布设图(36)

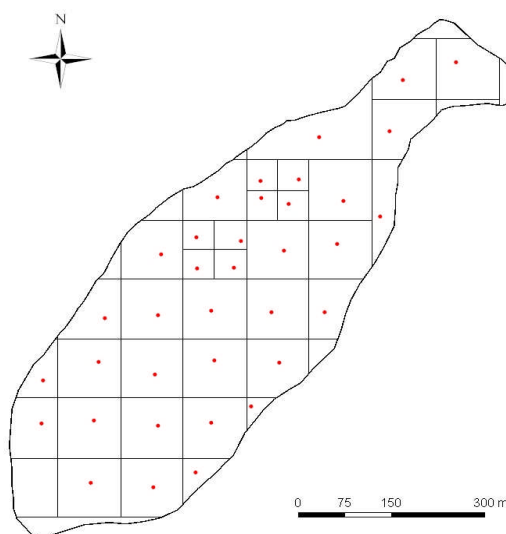
随机法采样样点设计图



网格法采样样点设计图



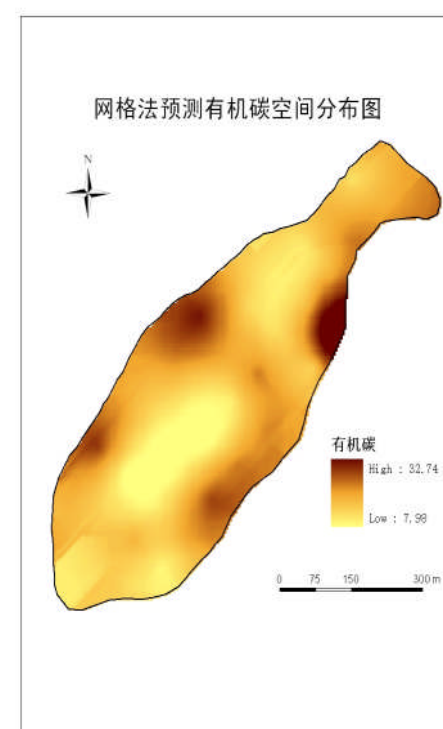
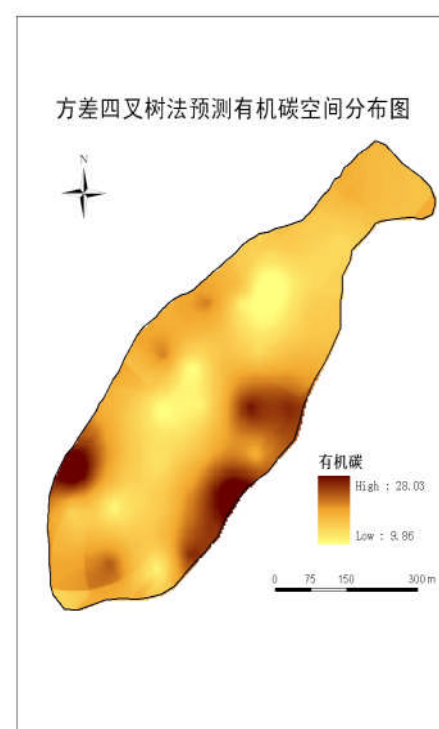
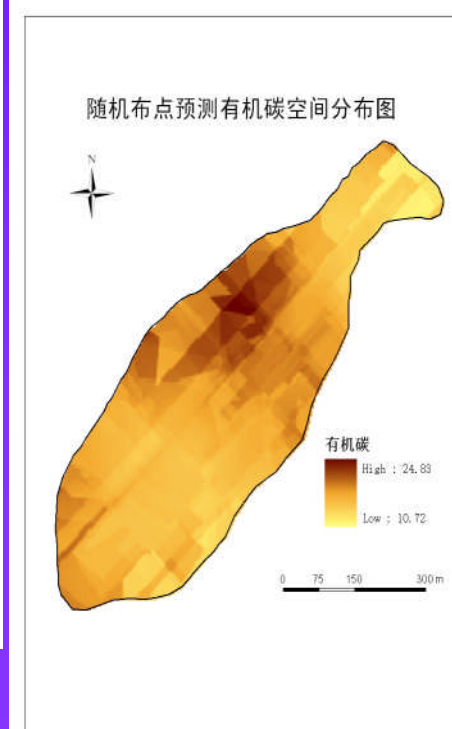
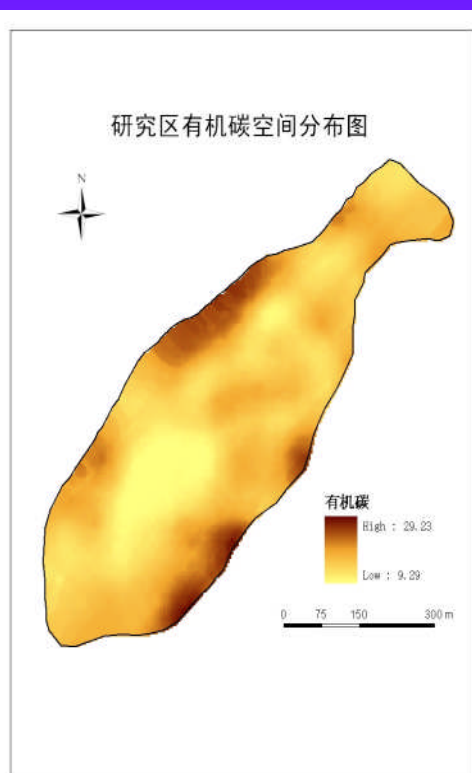
四叉树法采样样点设计图



# 结果与分析

## 样点布设比较

置信水平90%、允许误差10%情况下有机碳插值比较



# 结果与分析

## 不同样点布设方法的效率评价

研究目的不同，对土壤样本采集的要求也不同。土壤采样设计主要有两个目的：研究区土壤性质的含量水平和方差特征；研究区土壤性质的空间分布。

基于以上两种研究目的，从两方面对土壤采样方案进行评价：

- 1、目标样本与样本总体之间的差异（方差分析）
- 2、目标样本空间插值的预测精度（**Kriging**方法进行插值，**RMSE**对插值精度进行评价）

# 结果与分析

## 目标样本与样本总体比较分析

有机碳不同样点布设方法样本方差齐次性检验和均值比较

样点布 设方法	SOC (36 样点)		SOC (143 样点)	
	Levene' s F	Mean difference	Levene' s F	Mean difference
四叉树法	0.00	1.02	8.72**	1.44**
网格法	0.84	-0.72	0.01	-0.03
随机布点	0.15	0.13	0.01	0.18

\*\*表示该样点布设方法确定的样本统计量与样本总体存在极显著差异

全磷不同样点布设方法样本方差齐次性检验和均值比较

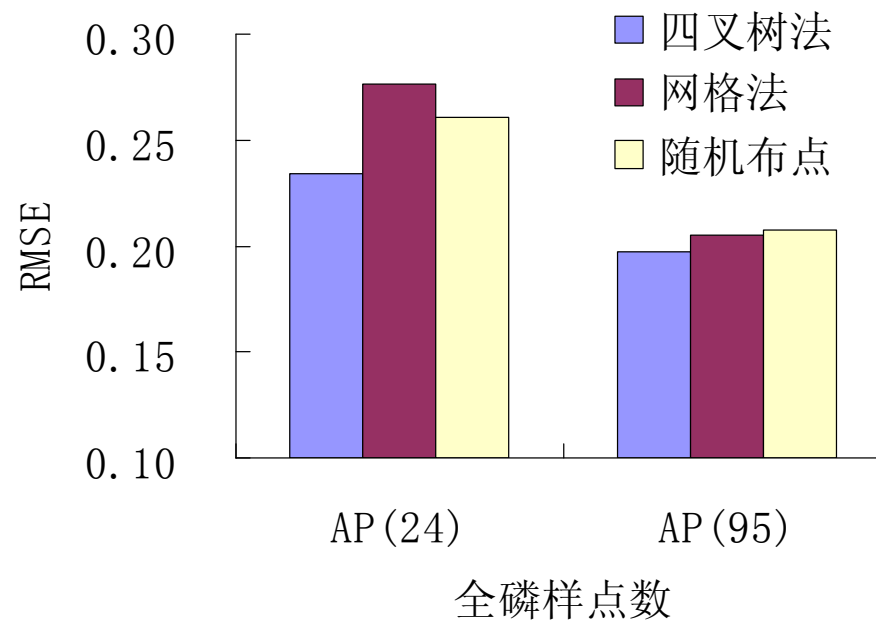
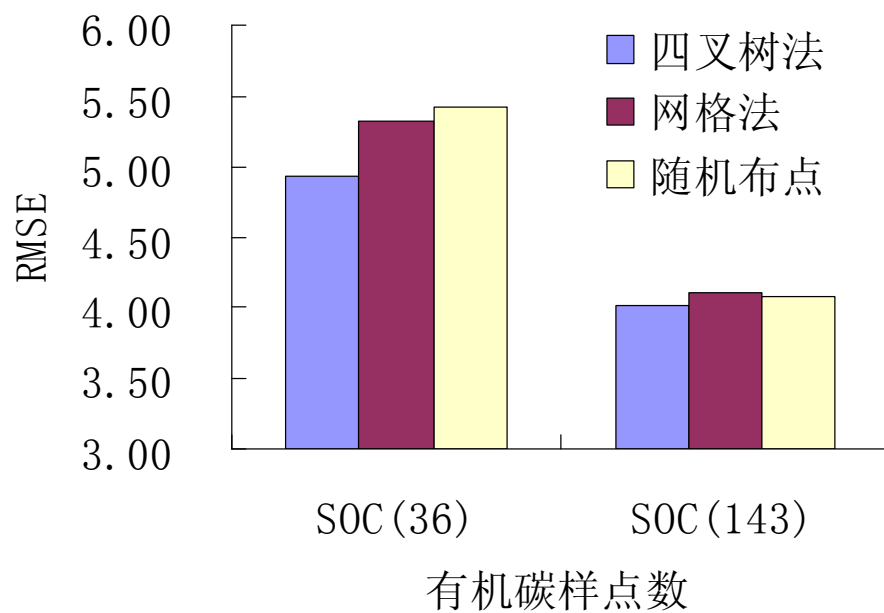
样点布 设方法	AP (24 样点)		AP (95 样点)	
	Levene' s F	Mean difference	Levene' s F	Mean difference
四叉树法	2.85*	0.07	1.30	0.01
网格法	0.12	0.00	0.10	-0.01
随机布点	0.07	-0.11*	0.90	-0.03

\*表示该样点布设方法确定的样本统计量与样本总体存在显著差异



# 结果与分析

## 不同样点布设方法插值精度的比较



四叉树法确定的样点布设方案的插值精度明显高于网格法和随机布点方法。

# 主要结论

- 网格取样能够更加稳健地反映研究区土壤性质的含量水平、方差特征
- 四叉树方法与其它两种方法比，可以更加精确的预测研究区土壤性质的空间分布特征。



谢谢！请批评指正！