

信息技术在海岸带遥感中的应用例说

马 毅、张 杰

任广波、王常颖、王小龙

国家海洋局第一海洋研究所

2014年12月23日

二〇一二年 IDM 组本部全体合影






国家海洋局第一海洋研究所

海洋物理与遥感研究室

开展海洋物理基础与应用研究
发展海洋遥感及其与多学科交叉





海洋遥测工程技术研究中心

(国家海洋局与航天科技集团共建)

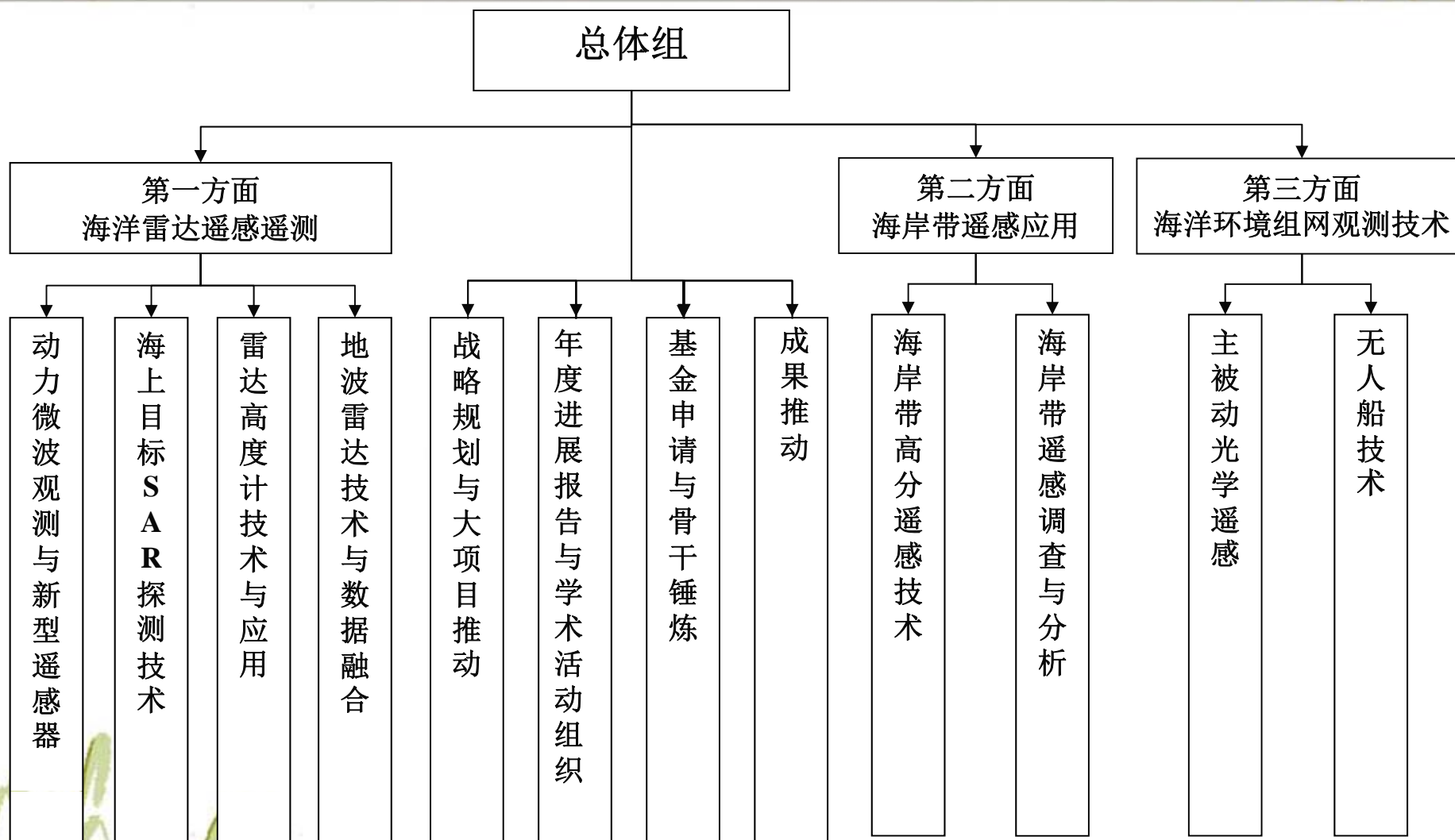
国家海洋局第一海洋研究所

中国航天704所

中国海监总队







报告内容



(一) 半监督方法与海岸带遥感分类

(二) D-S理论与海岸带遥感分类

(三) 最优化方法与赤潮生物量高光谱遥感反演

(四) De Morgan法则与海岛多源风险评价



(一) 半监督方法与海岸带遥感分类

1. 研究背景


1) 遥感影像的小样本分类问题：遥感影像中的地物类型统计分布特征本身具有很高的复杂性和不确定性，分类样本的人为选择又不可避免的会存在精力、经验知识有限和盲目性等缺点，所以人为选择样本常常数量少，且很难保证对待分类影像的代表性。。



2) 遥感影像的分类样本拓展应用问题

分类样本的拓展应用对无法获取有效训练样本的灾情的应急监测、跨区域的遥感监测研究和跨国界的军事监测具有十分重要的意义。

由于地物在不同时期、不同区域的影像中所处的辐射环境不同，而很难保证从一景影像中选择的分类样本可对另一景影像中的同种地物特征也有好的代表性。






2) 遥感影像的自动分类问题

遥感影像的自动分类是今后大规模、周期性、机械性区域遥感监测技术发展的重要方向。长期以来，从遥感影像上高精度自动提取信息一直是遥感应用研究者们的一个愿望。

建立预设样本集是实现自动化分类的最直接的途径，但却要面临预设样本对待分类影像代表性不好的问题。

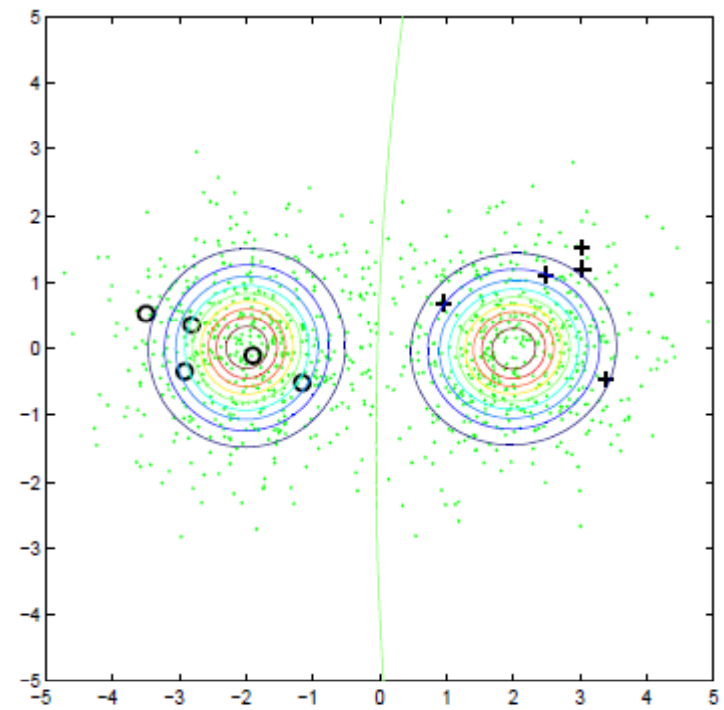
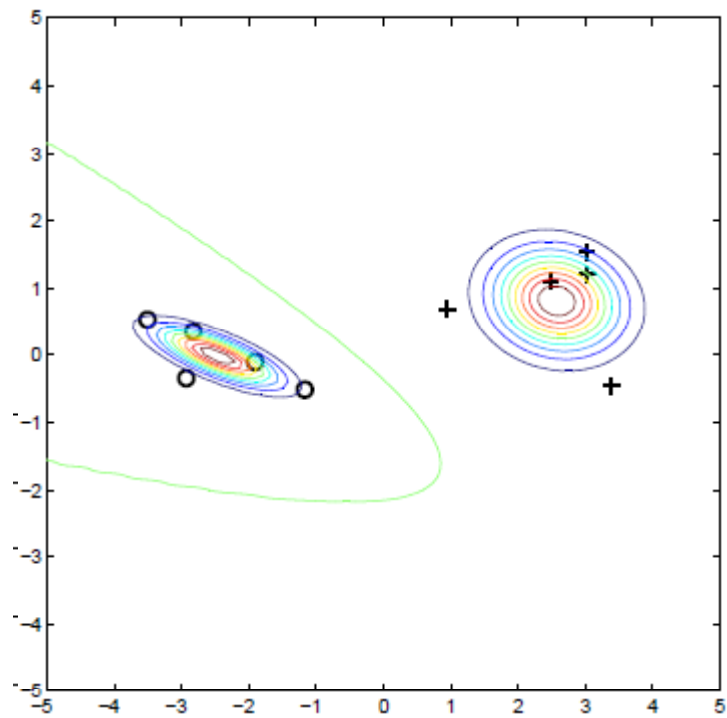


半监督学习分类方法

半监督学习是数据挖掘领域中的一种方法，该方法同时利用已知类别样本和未知类别样本，通过挖掘未知类别样本中所隐含的各待分类类型在特征空间中的固有结构信息，对已知类别样本可能因代表性不好而造成的拟合分类器有偏差情况进行矫正。

半监督学习分类方法

半监督学习是数据挖掘领域中的一种方法，该方法同时利用已知类别样本和未知类别样本，通过挖掘未知类别样本中所隐含的各待分类类型在特征空间中的固有结构信息，对已知类别样本可能因代表性不好而造成的拟合分类器有偏差情况进行矫正。

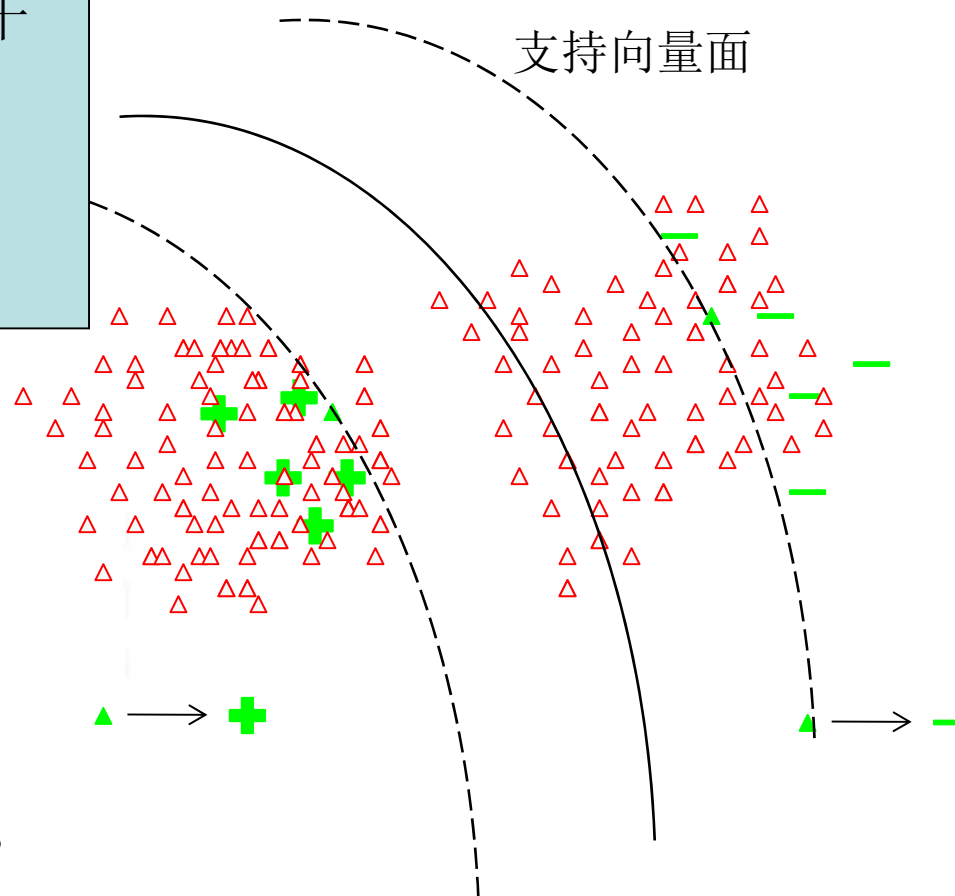


问题

遥感影像中未知样本动辄数千
标注效率低
未结合遥感影像数据的特点

目前常用标注方法：成对
标注

+ - labeled samples
△ unlabeled samples



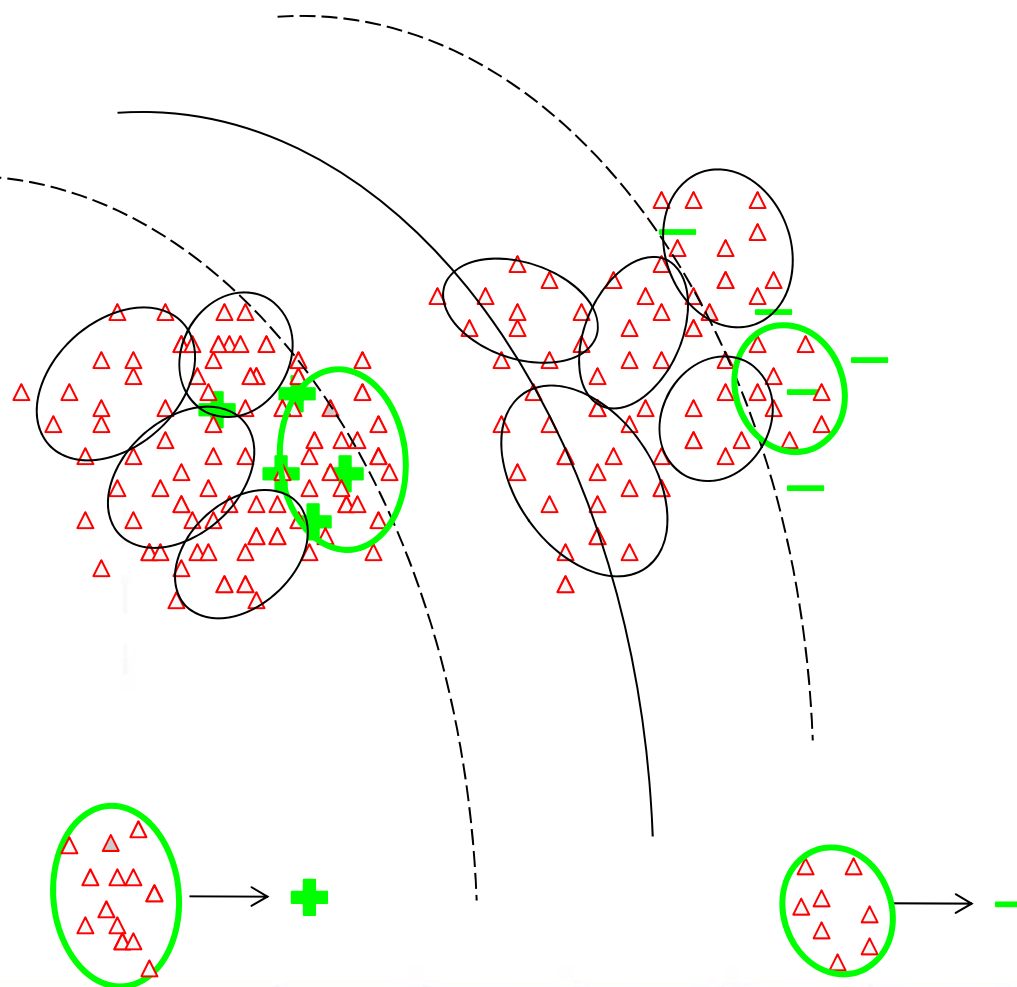
适用于遥感影像的未知样本标注法——聚类团簇

优点:

效率高

结合遥感影像数据的特点

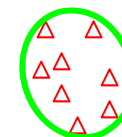
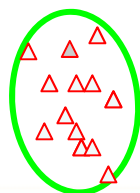
体现了数据降维的思想



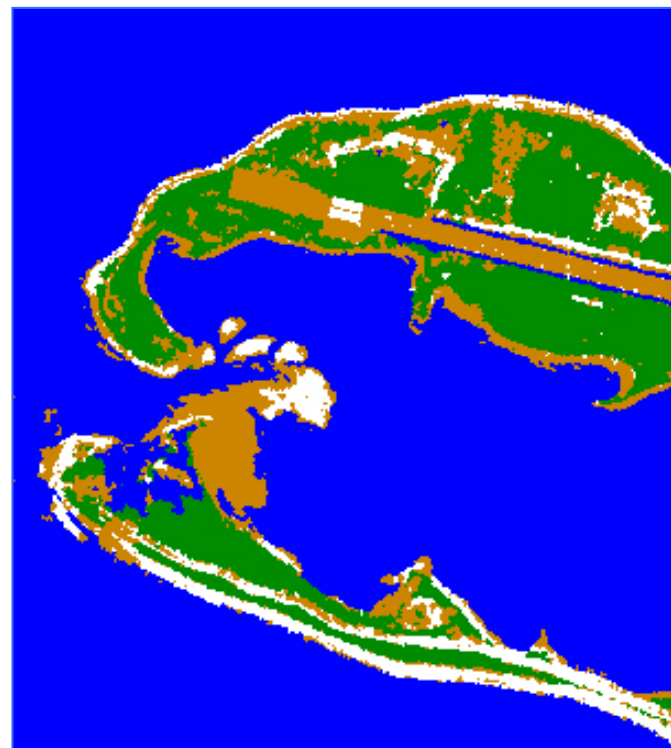
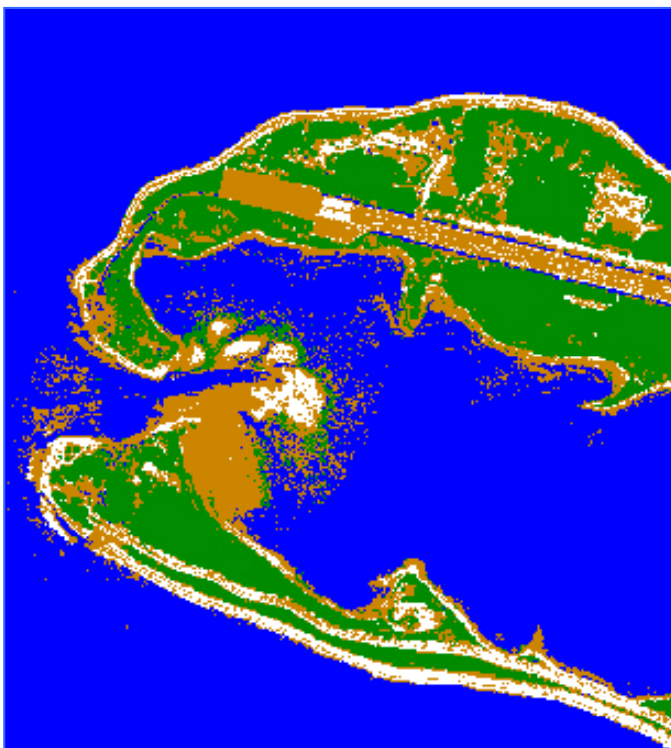
labeled samples



unlabeled samples



基于分割的高分辨率遥感影像聚类团簇



- 将坐标位置相邻的象元聚类——图像分割
- 以分割斑块为最小——聚类团簇

成对标注与聚类团簇标注对比



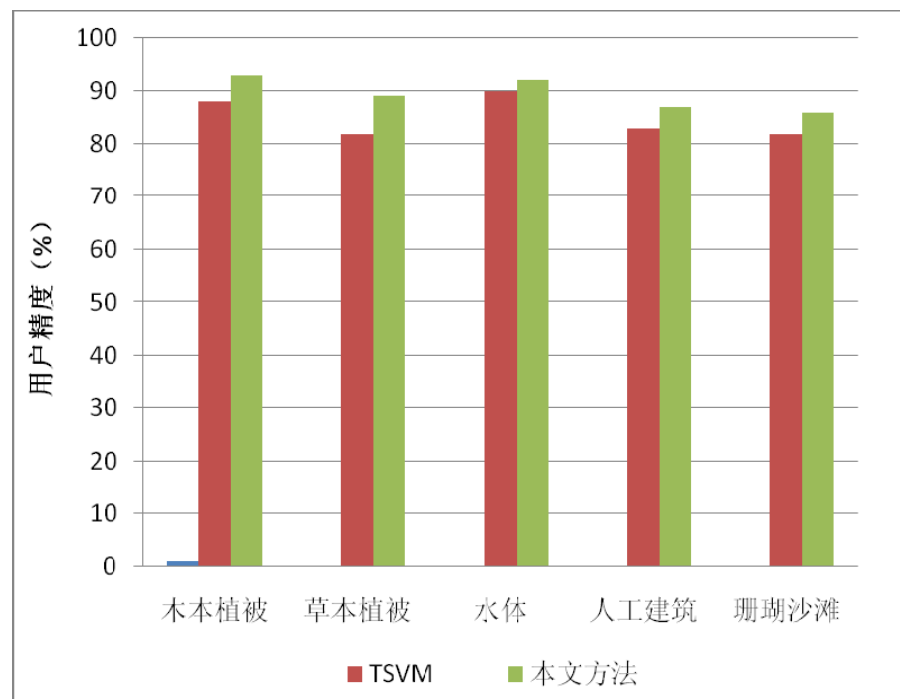
■ 水体 ■ 木本植被 ■ 草本植被 ■ 人工建筑 ■ 珊瑚沙滩

成对标注法半监督分类结果



聚类团簇法半监督分类结果

成对标注与聚类团簇标注对比



聚类团簇法因更好的描述了遥感影像的空间数据结构，而获得了更高的分类精度；分类效率方面，本实验中，聚类团簇标注法用时9s，成对标注法用时69s.

(二) D-S理论与海岸带遥感分类

1. D-S证据理论

由 **Dempster** 于 1967 年提出，其学生 **Shafer** 完善。

可看作是经典概率论的扩展，即：所描述的不确定性包括“不知道”或“不好说”。

概率分配函数

设 \mathbf{D} 是变量 \mathbf{x} 所有可能取值的集合，且 \mathbf{D} 中的元素是互斥的，即变量 \mathbf{x} 能且只能取 \mathbf{D} 中的某个元素为值，则称 \mathbf{D} 为 \mathbf{x} 的样本空间。 \mathbf{D} 的任一子集为一个命题。

概率分配函数 \mathbf{M} : $2^{\mathbf{D}} \rightarrow [0,1]$ ，且满足

$$M(\Phi) = 0$$

$$\sum_{A \subseteq \mathbf{D}} M(A) = 1$$

则称 \mathbf{M} 是 $2^{\mathbf{D}}$ 上的概率分配函数， $\mathbf{M}(\mathbf{A})$ 称为 \mathbf{A} 的基本概率数。

信任函数

Bel: $2^D \rightarrow [0, 1]$

$$Bel(A) = \sum_{B \subseteq A} M(B)$$

Bel(A)表示对命题A为真的信任程度。

容易推出对单元素子集，其基本概率数和信任程度是一致的。

另外，

$$Bel(\Phi) = M(\Phi) = 0$$

$$Bel(D) = \sum_{B \subseteq D} M(B) = 1$$

概率定义

定义在事件域 \mathcal{F} 上的一个集合函数 P 称为概率，如果它满足如下三个条件：

(1) $P(A) \geq 0$ ，对于一切 $A \in \mathcal{F}$ ；

(2) $P(\Phi) = 0$ ， $P(\Omega) = 1$ ；

(3) 若 $A_i \in \mathcal{F}$ ， $i = 1, 2, \dots$ ，且两两互不相容，则

$$P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

性质 (3) 称为可列可加性或完全可加性。

似然函数

Pl: $2^D \rightarrow [0, 1]$, 且

$$\mathbf{Pl(A)=1-Bel(-A)}$$

Pl(A)表示对A为非假的信任程度。

Pl(A) - Bel(A)表示对A不知道的程度，即既非对A信任又非不信任的那部分。

概率分配函数的正交和

设 M_1 和 M_2 是两个概率分配函数，则其正交和 $M = M_1 \oplus M_2$ 为

$$M(\Phi) = 0$$

$$M(A) = K^{-1} \times \sum_{X \cap Y = A} M_1(X) \times M_2(Y)$$

其中：

$$K = 1 - \sum_{X \cap Y = \Phi} M_1(X) \times M_2(Y) = \sum_{X \cap Y \neq \Phi} M_1(X) \times M_2(Y)$$

如果 $K \neq 0$ ，则正交和 M 也是一个概率分配函数；如果 $K = 0$ 则不存在正交和 M ，称 M_1 与 M_2 矛盾。

2. 证据理论的扩展

样本空间 \mathbf{D} 是变量 \mathbf{x} 所有可能取值的集合，且 \mathbf{D} 中的元素是互斥的，即变量 \mathbf{x} 能且只能取 \mathbf{D} 中的某个元素为值。应用到地物分类时，要求地物类型是完全已知的。

事实上，在对一景影像进行地物分类时，往往会存在一些人们不知道的地物类型，因此，证据理论无法应用。

下面就提出了一种能够处理样本空间元素不完全知道情况下扩展的证据理论。

概率分配函数

设样本空间 \mathbf{D} 中的元素满足互不相容的条件，但是存在 \mathbf{D} 中没有的未知类别，“变量 x 能且只能取 \mathbf{D} 中的某个元素为值”？

设函数 $M: 2^{\mathbf{D}} \rightarrow [0,1]$ ，且满足：

$$M(\Phi) \geq 0$$

$$\sum_{A \subseteq \mathbf{D}} M(A) = 1$$

则称 M 是 \mathbf{D} 上的概率分配函数， $M(A)$ 称为 A 的基本概率数， A 是 \mathbf{D} 的子集。

信任函数

Bel: $2^D \rightarrow [0, 1]$

$$Bel(A) = \sum_{B \subseteq A, B \neq \Phi} M(B)$$

Bel(A)表示对命题A为真的信任程度。

容易推出对单元素子集，其基本概率数和信任程度是一致的。另外，

$$Bel(\Phi) = M(\Phi) \geq 0$$

$$Bel(D) = \sum_{B \subseteq D, B \neq \Phi} M(B) = 1 - M(\Phi)$$

似然函数

Pl: $2^D \rightarrow [0, 1]$, 且

$$Pl(A) = 1 - M(\Phi) - Bel(-A)$$

Pl(A)表示对A为非假的信任程度。

Pl(A) - Bel(A)表示对A不知道的程度，即既非对A信任又非不信任的那部分。

概率分配函数的正交和

设 M_1 和 M_2 是两个概率分配函数，则其正交和 $M = M_1 \oplus M_2$ 为

$$M(\Phi) = K^{-1}[M_1(\Phi) + M_2(\Phi) - M_1(\Phi) \times M_2(\Phi)]$$

$$M(A) = K^{-1} \times \left[\sum_{x \cap y = A} M_1(X) \times M_2(Y) + M_1(A) \times M_2(\Phi) + M_1(\Phi) \times M_2(A) \right]$$

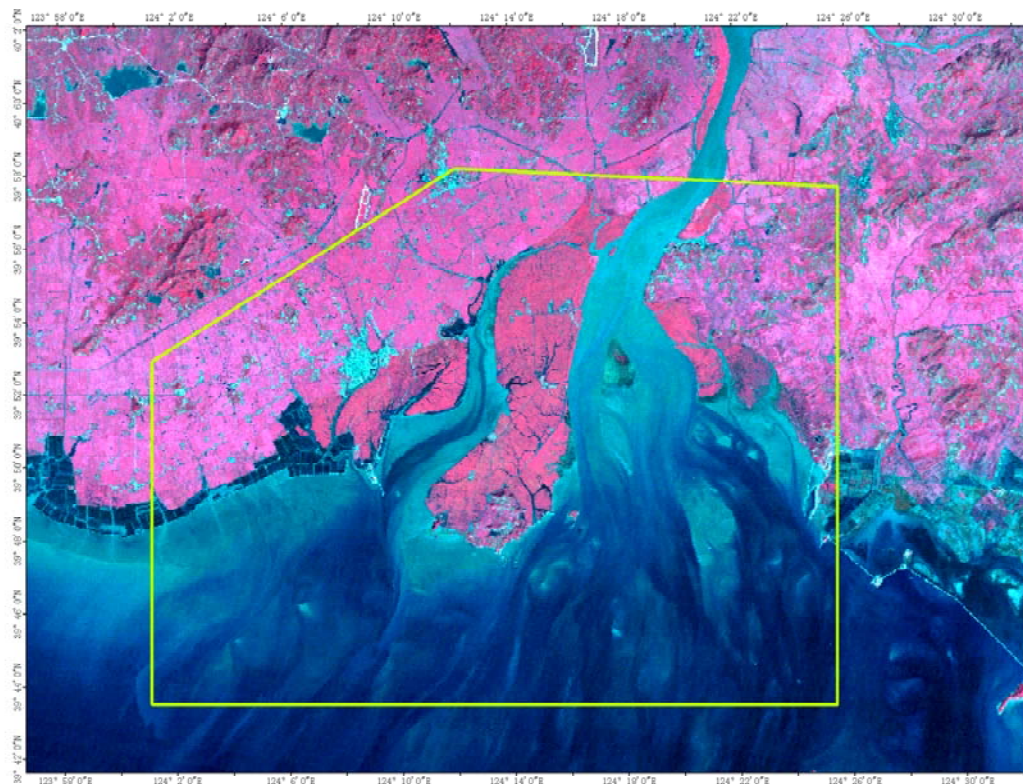
其中：

$$K = \sum_{X \cap Y \neq \emptyset} M_1(X) \times M_2(Y) + M_1(\Phi) + M_2(\Phi) - M_1(\Phi) \times M_2(\Phi)$$

如果 $K \neq 0$ ，则正交和 M 也是一个概率分配函数；如果 $K = 0$ 则不存在正交和 M ，称 M_1 与 M_2 矛盾。

3. 证据理论在遥感地物分类中的应用

选取鸭绿江河口
区域的SPOT-5影像
为数据源，其中的主
要地物类型为耕地、
水体、建设用地、潮
滩、荒地/林地等。



分类方法介绍




基于证据理论的高分辨率影像海岸带地物分类主要分三个步骤：

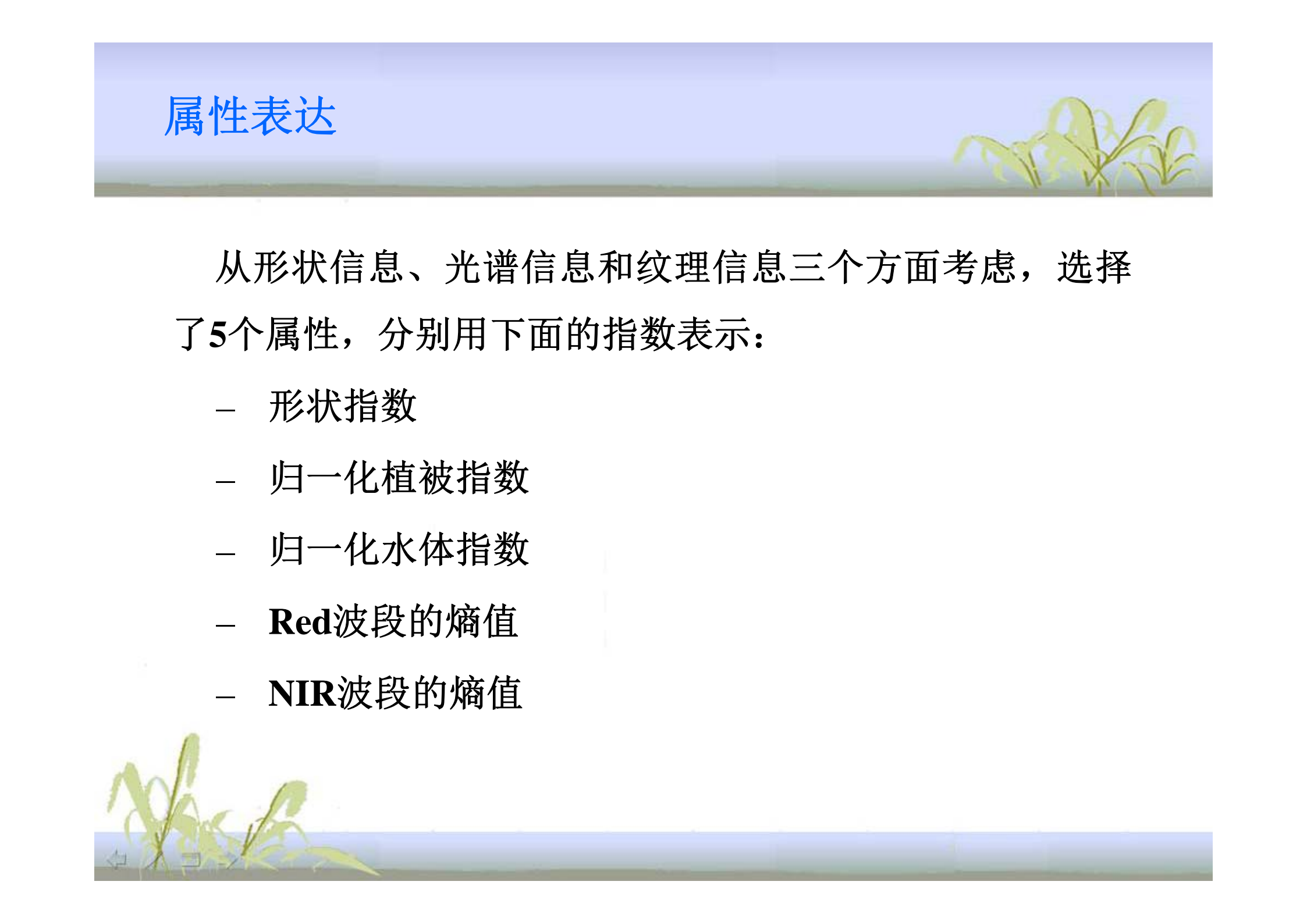
(1) **证据源的选择**：从形状、光谱、纹理选择分类特征，通过典型指数来实现。

(2) **分类规则的形成**：由训练影像中地物类型的特征属性的直方图分布，来定义概率分配函数，然后，计算各种证据源导出的概率分配函数正交和，求出各种地物类别的可信度，确定可信度最大的类别，得到“特征属性→地物类别”的规则。

(3) **基于规则的分类**：根据上面形成的规则，实现分类。



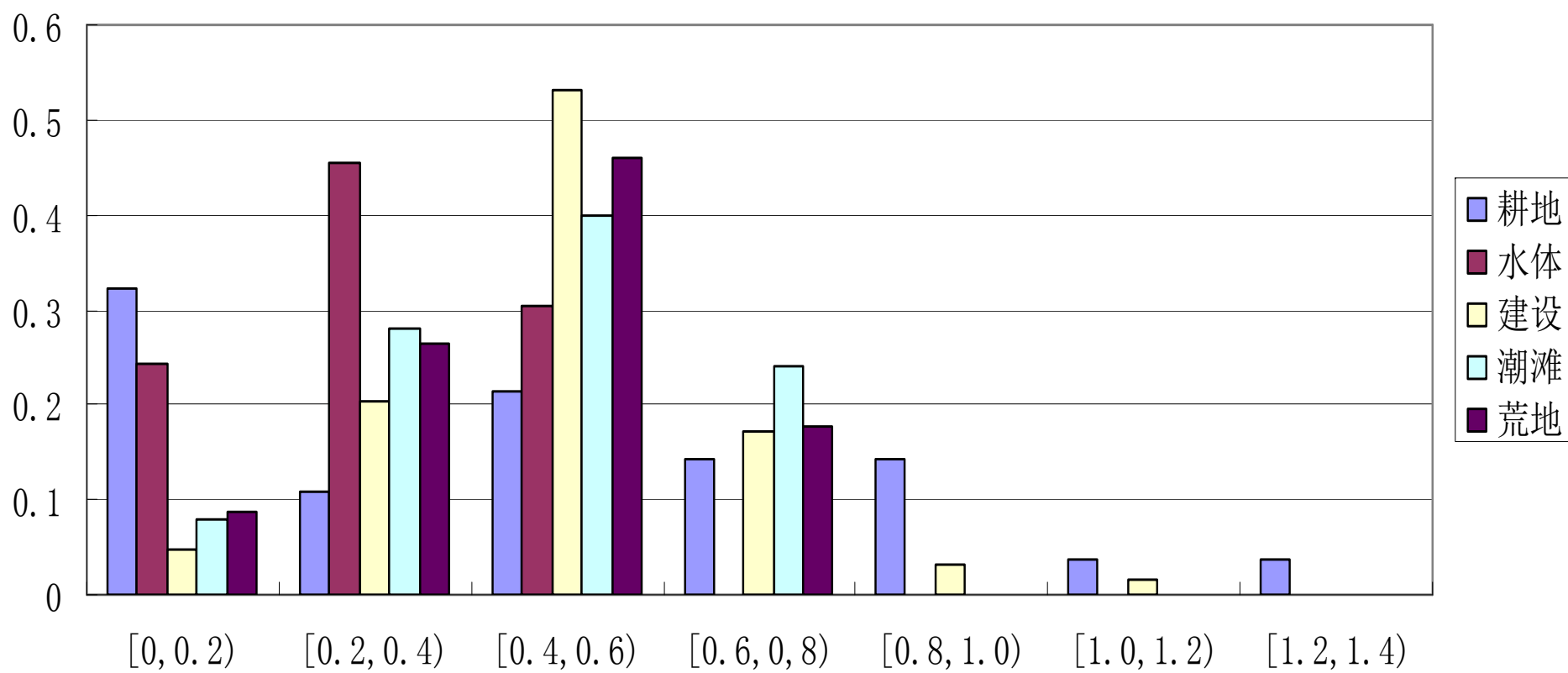
属性表达



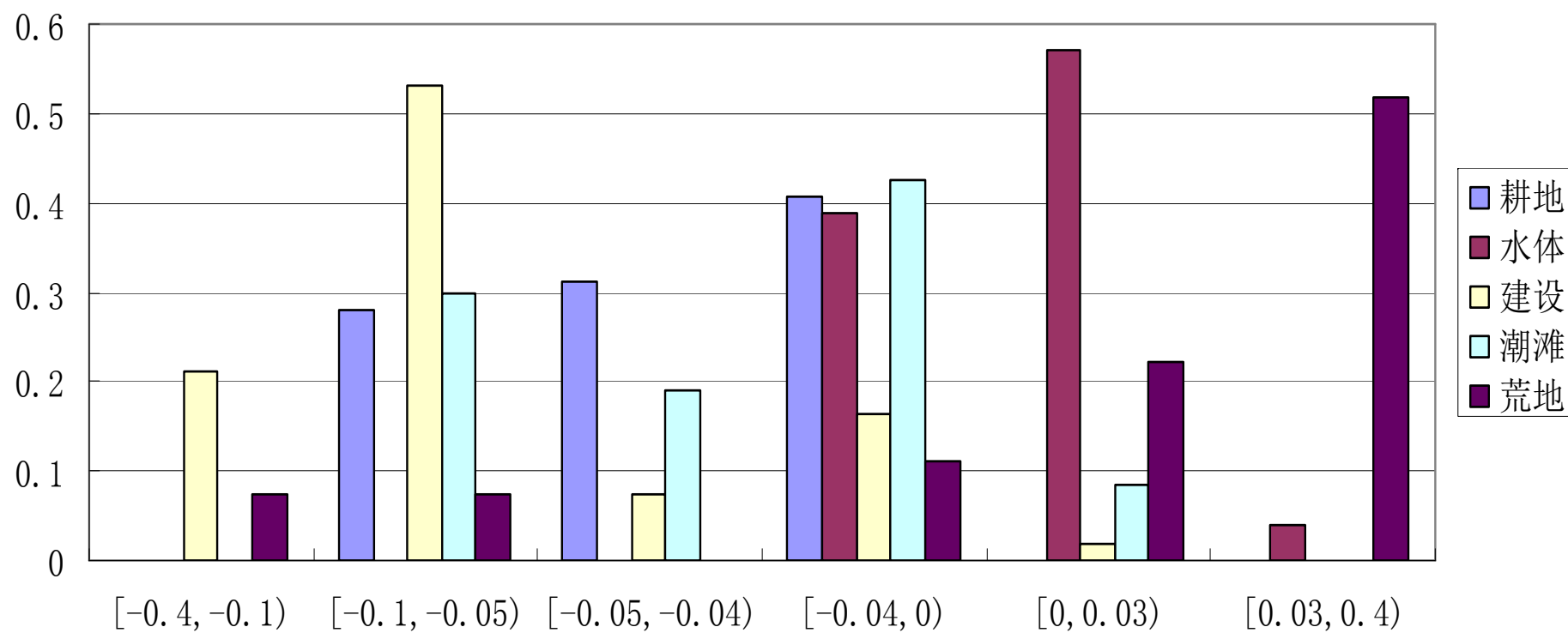
从形状信息、光谱信息和纹理信息三个方面考虑，选择了5个属性，分别用下面的指数表示：

- 形状指数
- 归一化植被指数
- 归一化水体指数
- **Red**波段的熵值
- **NIR**波段的熵值

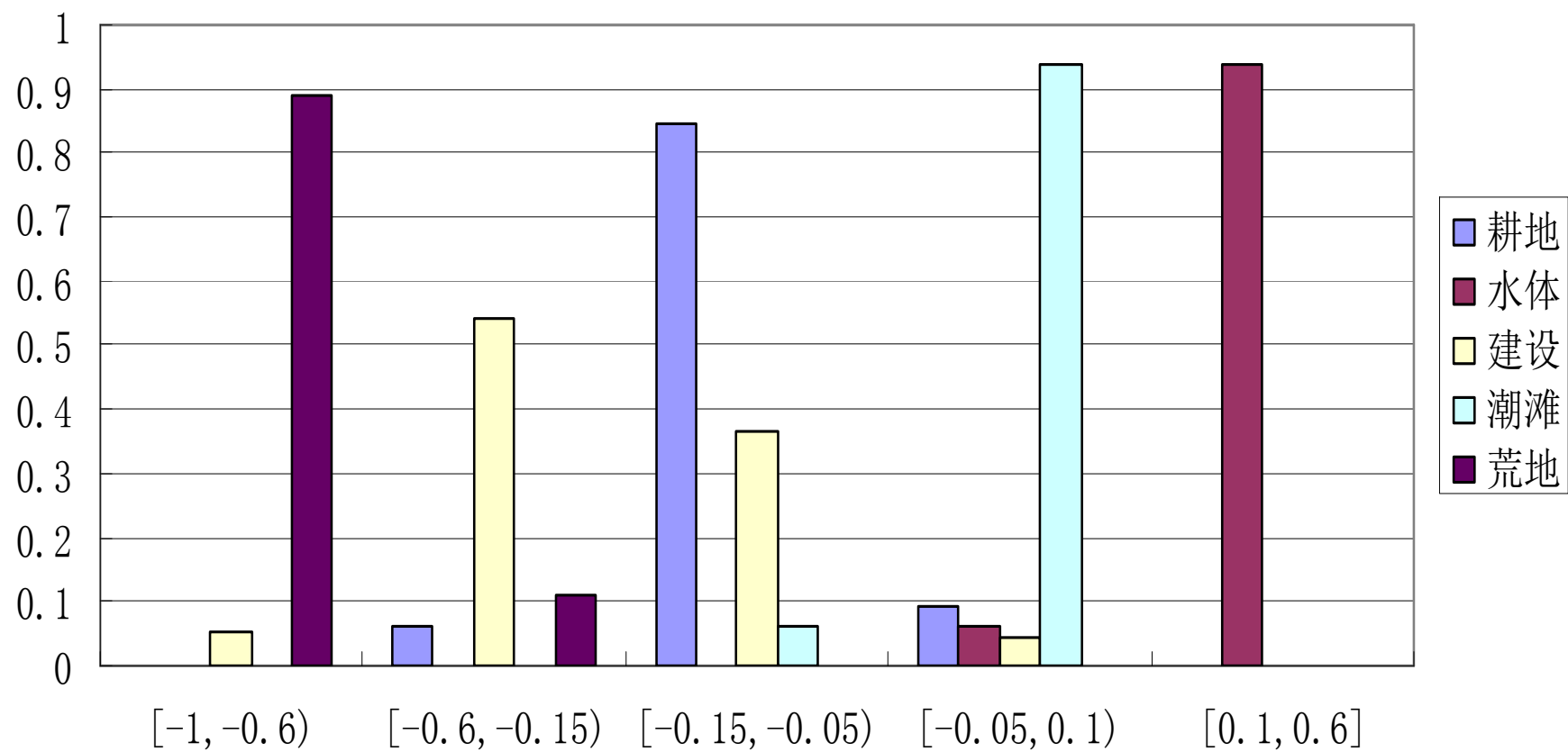
形状指数的基本概率分布



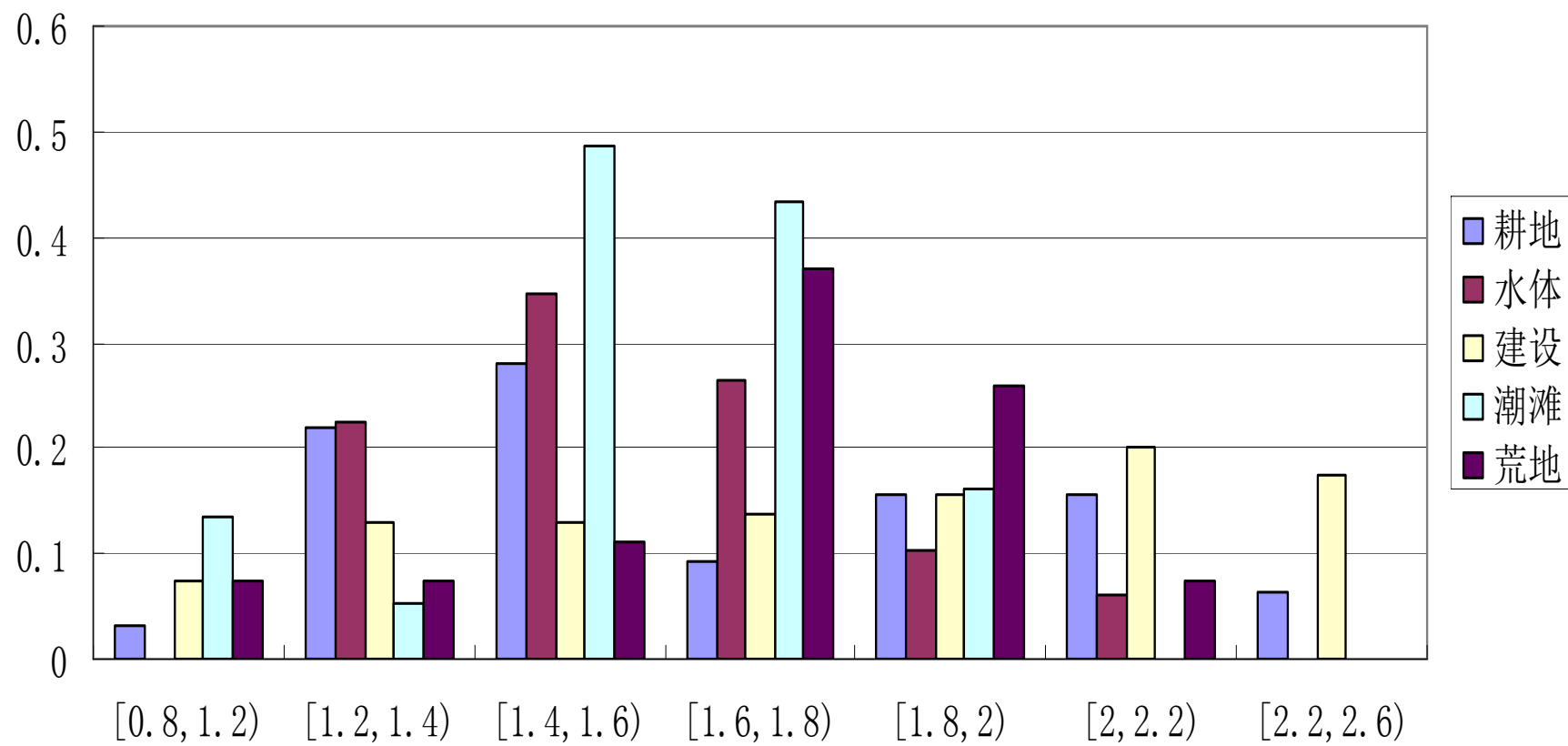
植被指数的基本概率分布



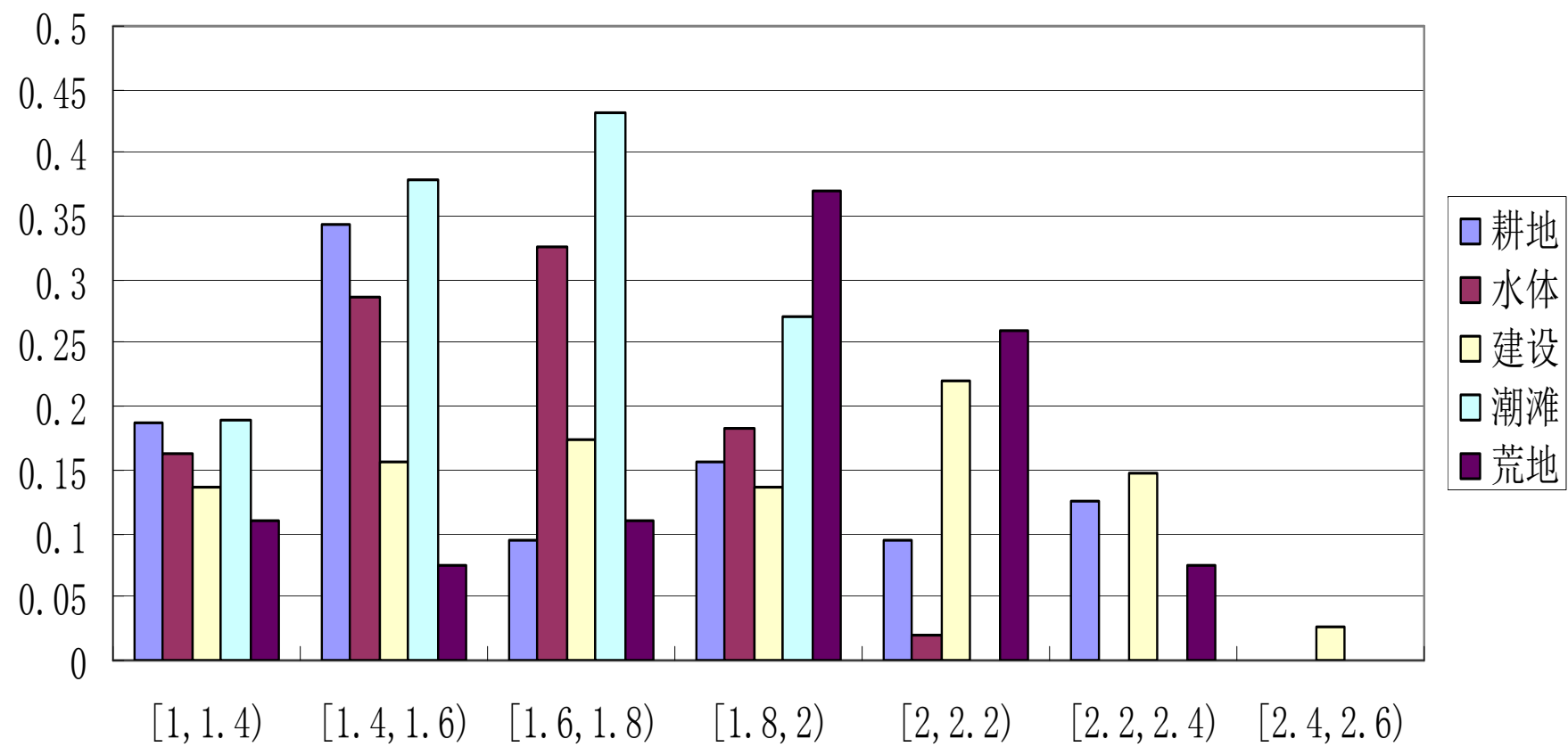
水体指数的基本概率分布



熵-Red



熵-NIR



分类规则形成

样本空间 $D=\{\text{耕地, 水体, 荒地, 潮滩, 建设}\}$ ，基于证据理论的地物分类方法的核心就是构造“特征 \rightarrow 类别”规则。

基于上述**5类32种**证据源，转化为**32种**概率分配函数，例如如下：

表1 形状指数为[0,0.2)的基本可信度分配 m_1

{耕地}	{耕地, 水体}	{耕地, 水体, 荒地}	{耕地, 水体, 荒地, 潮滩}	{耕地, 水体, 荒地, 潮滩, 建设}
0. 2458	0. 4797	0. 0256	0. 1031	0. 1458

表2 熵-red为[0,0.03)的基本可信度分配 m_2

{建设}	{耕地, 建设}	{耕地, 建设, 荒地}	{耕地, 水体, 建设, 荒地}
0. 2259	0. 4071	0. 0637	0. 3033

计算 m_1 与 m_2 的正交和 m

$$K=1-0.2259*(0.2458+0.4797+0.0256+0.1031)=0.8070;$$

$$m(\{\text{耕地}\})=1/0.8070*[0.2458*(0.4071+0.0637+0.3033)\\+0.4797*(0.4071+0.0637)+(0.0256+0.1031)*0.4071]=0.5805;$$

$$m(\{\text{建设}\})=1/0.8070*[0.1458*0.2259]=0.0408;$$

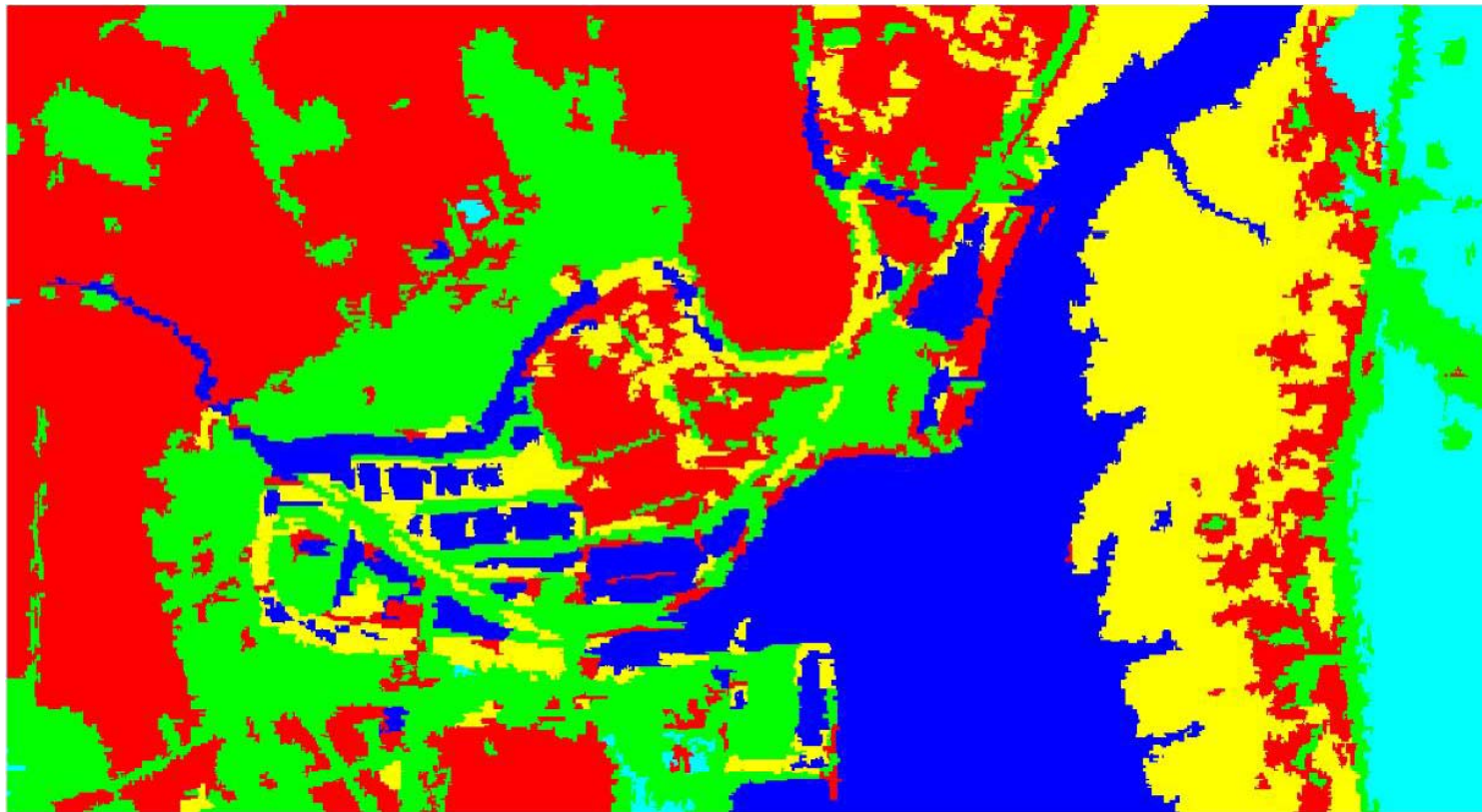
“形状指数为 $[0, 0.2)$ ” 且 “熵-red 为 $[0, 0.03)$ ” 的各类别信任值

{水体}	{耕地}	{荒地}	{潮滩}	{建设}
0	0.5805	0	0	0.0408

这样，就形成了特征到类别的分类规则：

“形状指数= $[0, 0.2)$ 且 熵-red= $[0, 0.03)$ ->类别={耕地}”

实施分类



耕地 建设用地 水体 潮滩 荒地或林地

测试数据识别精度

应用前面训练样本得到的地物分类规则进行类型识别；共计**271**个斑块，正确识别数目为**220**块，误识别斑块数目为**51**块。

实际类型 ↓ 识别类型 →	耕地	水体	建设用地	荒地或林地	潮滩	识别精度
耕地	17	0	0	0	1	17/18=94.44%
水体	1	47	0	3	0	47/51=92.16%
建设用地	0	9	93	1	20	93/123=75.61%
荒地或林地	0	5	6	11	0	11/22=50%
潮滩	2	0	1	0	54	54/57=94.74%
总识别精度：81.92%						

提出的识别方法能够高精度的识别出耕地、水体和潮滩，其识别精度分别为**94.44%**，**92.167%**和**94.74%**；能够较好的识别出建设用地，其识别精度为**75.61%**、而荒地或林地的识别结果不是非常理想；但总体识别精度达到**81.92%**。

(三) 最优化与赤潮生物量高光谱遥感反演

最优化基本理论与方法

- 最优化方法是用数值计算方法寻找最佳的选择，即目标函数满足一定约束条件下的极大或极小

- 无约束最优化问题

$$\min f(x), x \in R^n$$

- 经典算法：最速下降法、共轭梯度法
牛顿法、拟牛顿法

赤潮生物量高光谱遥感反演应用



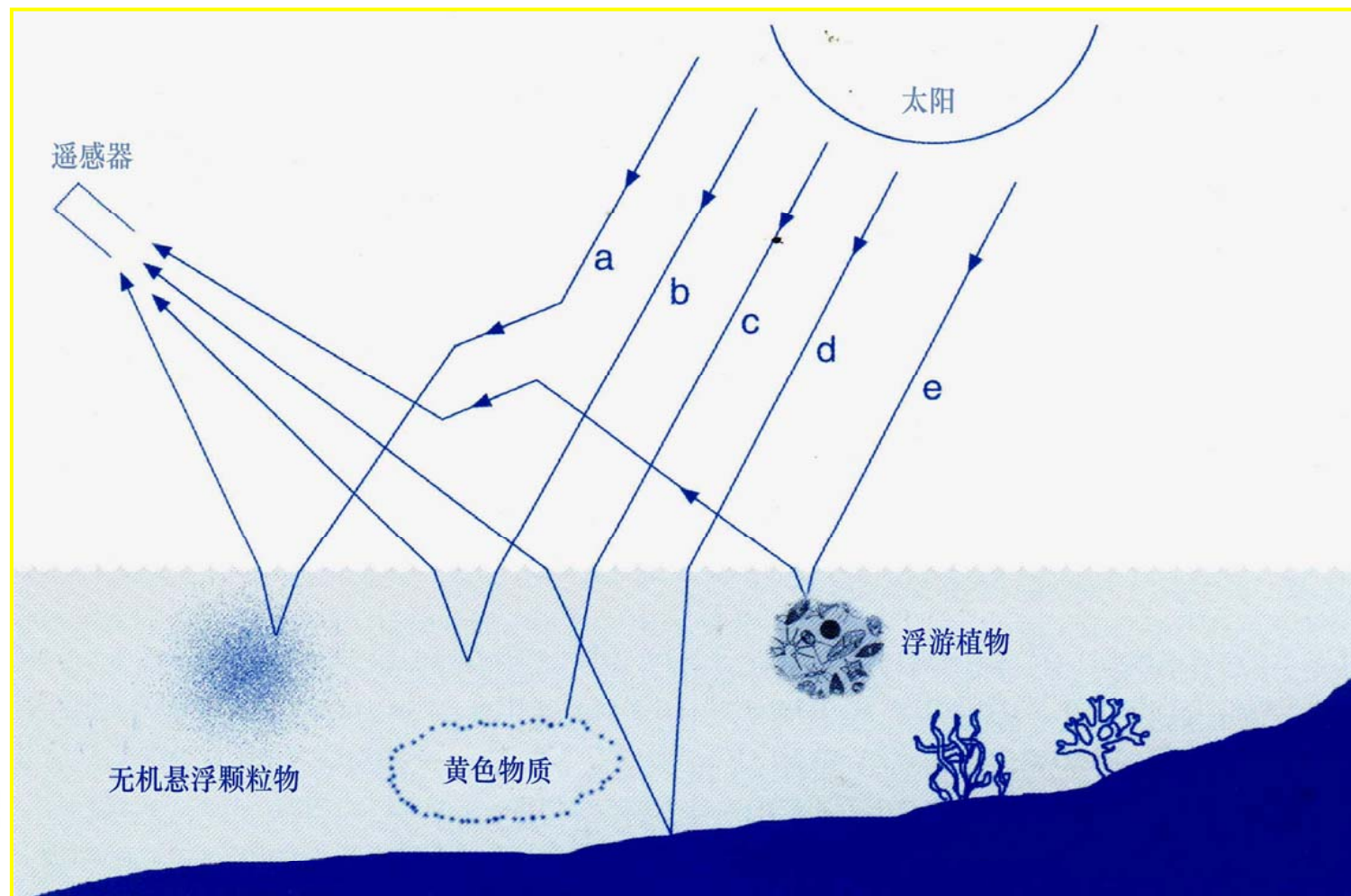
渤海赤潮航空照片



赤潮航空高光谱图像



水体光学成像机理



正问题模型

遥感反射率:

$$R(\lambda) = 0.051 \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}$$

吸收:

$$a(\lambda) = a_w(\lambda) + a_p(\lambda) + a_y(\lambda)$$

$$a_p(\lambda) = [a_0(\lambda) + a_1(\lambda) \ln(a_p(440))] a_p(440)$$

$$a_p(440) = 0.06 P^{0.65}$$

$$a_y(\lambda) = Y \exp(-0.014(\lambda - 440))$$

散射:

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bs}(\lambda)$$

$$b_{bw}(\lambda) = 0.5 b_w(\lambda)$$

$$b_w(\lambda) = 0.005862 \left(\frac{400}{\lambda} \right)^{4.322}$$

$$b_{bs}(\lambda) = B_1 b_1(\lambda) C_1 + B_2 b_2(\lambda) C_2$$

正问题模型参数优化

待优化的参数:

$$R(\lambda) = x_1 \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}$$

$$a_p(440) = x_2 P^{x_3}$$

$$a_y(\lambda) = x_4 \exp(-0.014(\lambda - 440))$$

$$b_1(\lambda) = x_7 \left(\frac{400}{\lambda} \right)^{x_8}$$

$$b_2(\lambda) = x_9 \left(\frac{400}{\lambda} \right)^{x_{10}}$$

$$b_{bs}(\lambda) = x_5 b_1(\lambda) C_1 + x_6 b_2(\lambda) C_2$$

偏差函数的最优化问题:

$$\min E(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}) = \sqrt{\sum_{i=1}^n [R_m(\lambda_i) - R_c(\lambda_i)]^2}$$

反问题模型

代价函数:

$$E(P, C, F) = \sqrt{\sum_{i=1}^n [R_m(\lambda_i) - R_c(\lambda_i, P)]^2}$$

偏差函数的最优化问题:

$$\min E(P, C, F) = \sqrt{\sum_{i=1}^n [R_m(\lambda_i) - R_c(\lambda_i, P, C, F)]^2}$$

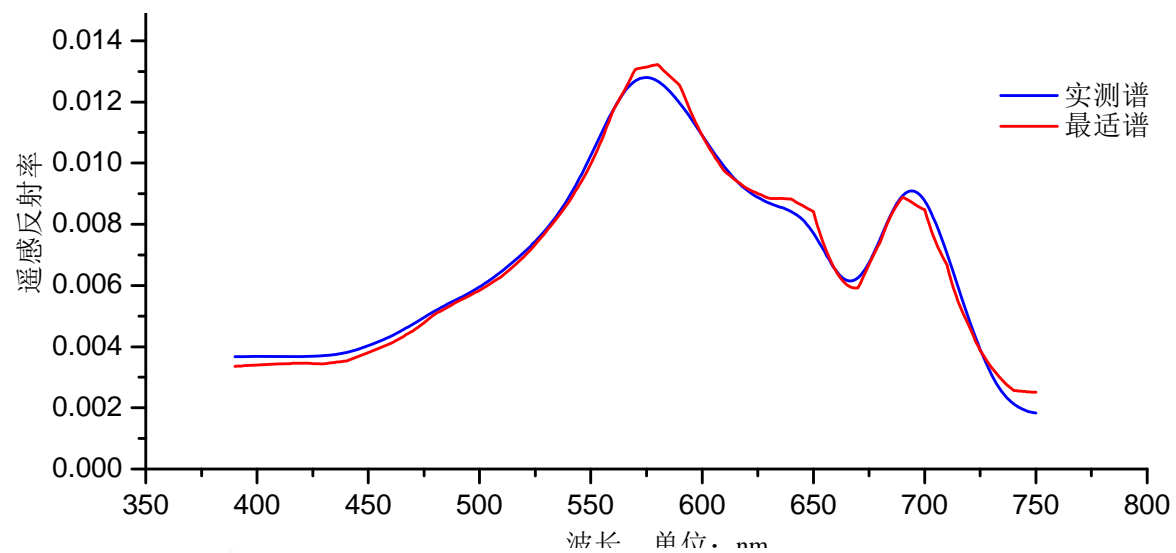
赤潮生物量高光谱遥感反演例子



$$P_0 = 1 \text{ mg/m}^3$$

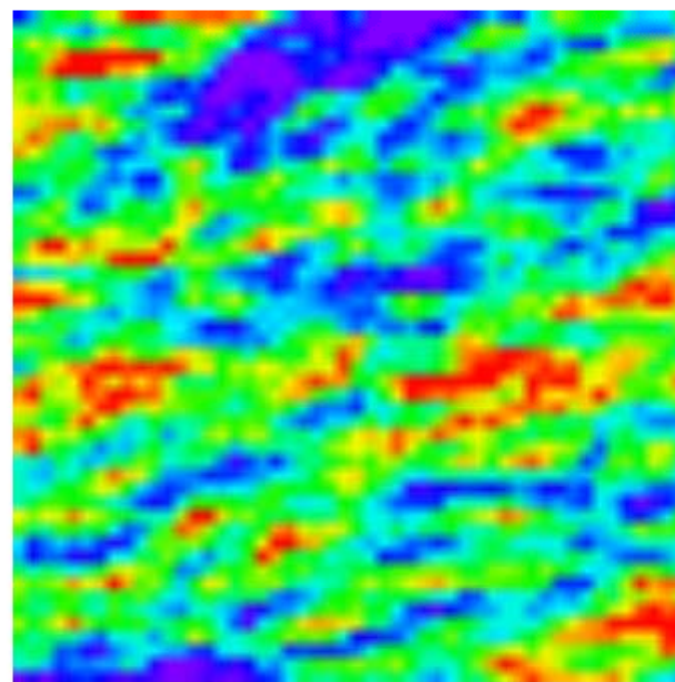
$$C_0 = 1 \text{ mg/m}^3$$

$$F_0 = 0.001$$



实测谱和最适谱





反演叶绿素浓度最大值为 13.3mg/m^3 ，最小值为 7.8mg/m^3 ，平均值为 10.2mg/m^3 。由于没有同步实测资料，以高光谱数据获取日期所在的7月上半月的平均值 6.5mg/m^3 、8月上半月的平均值 15.8mg/m^3 为参考，反演结果合理。

(四) 概率论中的De Morgan法则与 海島多源风险评价

1. 对事件 A 与 B ，若 $P(AB) = P(A) P(B)$ ，则称事件 A 与事件 B 是相互独立的
2. 对于三个事件 A 、 B 、 C ，若下列四个等式同时成立，则称它们相互独立。

$$P(AB) = P(A) P(B)$$

$$P(BC) = P(B) P(C)$$

$$P(AC) = P(A) P(C)$$

$$P(ABC) = P(A) P(B) P(C)$$

3. De Morgan法则

假设 A_i ($i=1, 2, \dots, m$) 相互独立, 则

$$P(\bigcup_{i=1}^m A_i) = 1 - \prod_{i=1}^m (1 - P(A_i))$$

概率型风险复合表征模型提出

1. 模型现状

- 风险计算的基本公式： $R = P \cdot D$
- 概率型风险评价可分为单源风险评价和多源风险复合评价
- 目前，多源风险复合评价采用的模型和算法基本都是线性的（权重求和），缺点：综合过程缺乏明确的物理意义，随意性大。
- 现阶段的研究已经指出：多源概率型风险的联合作用存在非线性效应。

2. 模型提出

我们的非线性模型包括四个算式：

(1) 评价单一风险源对单一生境类型的风险压力：

$$R_{ik} = 1 - \prod_{n=1}^j (1 - \lambda_n \beta_i D_k P_n)$$

(2) 评价单一风险源对评价区域的风险压力： $R_i = \sum_k R_{ik}$

(3) 评价单一生境受复合风险联合作用的风险压力：

$$R_k = 1 - \prod_{i=1}^m (1 - R_{ik})$$

(4) 评价区域综合风险压力： $R = \sum_k R_k$



国外审稿专家对模型的评价

审稿专家一：

This paper proposes a nonlinear model for assessing multiple probabilistic risks. Four types of habitats (Settlement, farmland and grassland, woodland and shrubbery, cobble tideland) are studied as risk receptors to 4 types of risks (drought, cold wave, pest plague, storm tide). This study used the bird species as the endpoint of assessment through roosting, feeding and breeding. The overall written of this paper was clear and easy to read, and the contribution of the study was original. A

审稿专家二：

This manuscript mainly focuses on evaluation of impacts from different risk sources, and tries to develop “integrated” approach for risk assessment. To make a case of ORIGINAL research, suggestions/comments are as follows:



海島概率型 風險評價框架

■ 問題形成

- (1) 區域背景分析
- (2) 區域生境結構分析
- (3) 風險源分析
- (4) 受體分析

■ 暴露與危害分析

- (1) 暴露分析：暴露—響應概念模型
- (2) 危害分析：潛在生態損失度

■ 風險計算

概率型複合風險表征模型應用

■ 風險管理

針對生態風險評估結果給出可能採取的對策

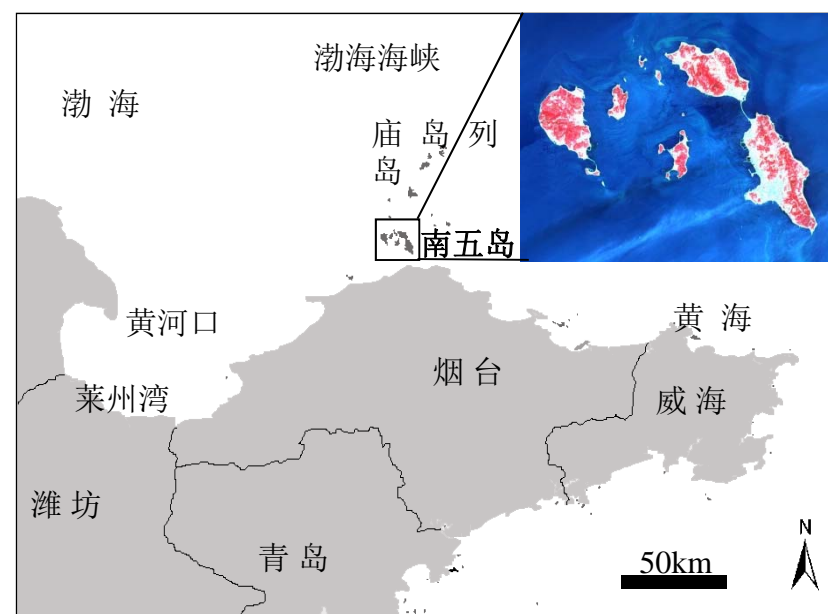
长岛南五岛概率型风险评价实例

1. 问题形成

■ 区域背景

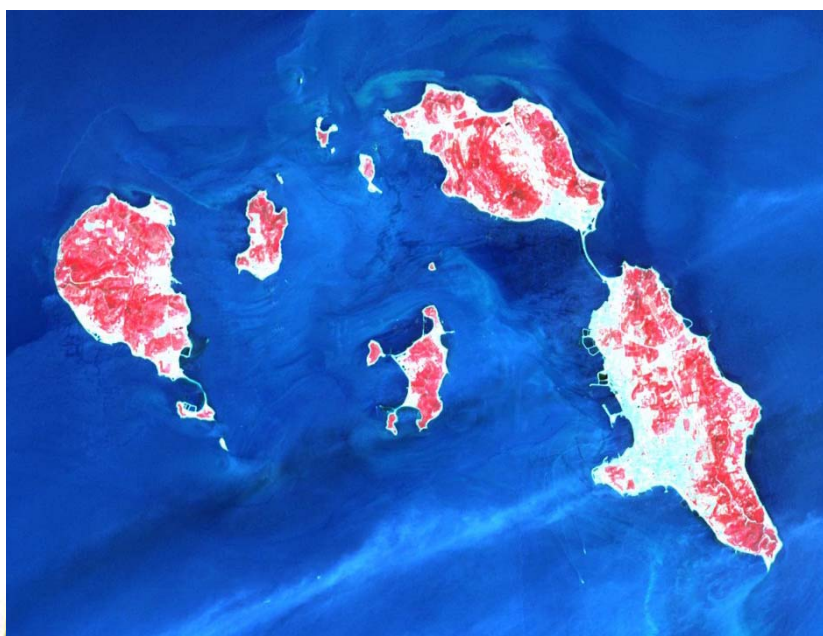
长岛南五岛位于黄、渤海交汇处，陆域面积 32.2km^2 ，南、北长山岛统称长山岛，为山东省第一、第二大岛。

南五岛是长岛国家自然保护区候鸟保护的核心区，该区共记录鸟类284种。国家一级保护鸟类9种，二级43种。此外，该区还是斑海豹和蝮蛇的重点保护区。

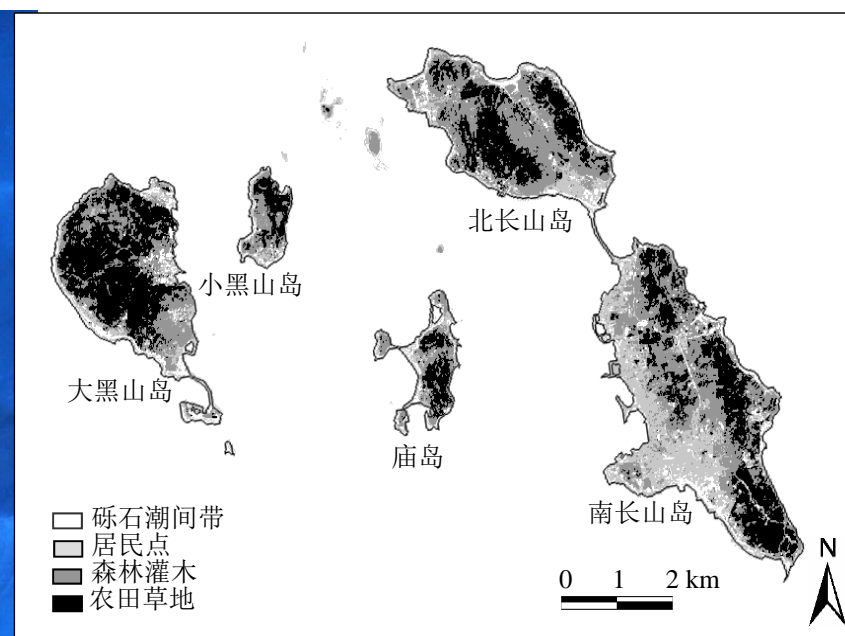


■ 区域生境结构

本文采用遥感手段进行区域生境结构分析，主要解译区分四类生境：林地/灌木、农田草地、居民点、潮滩。



ASTER遥感影像（2002. 6）



ASTER影像解译图

南五岛各岛生境类型面积GIS统计

	南长山岛	北长山岛	大黑山岛	小黑山岛	庙岛
居民点 (km ²)	3.896	1.425	0.832	0.349	0.503
农田草地 (km ²)	5.102	3.547	2.557	0.373	0.473
森林灌木 (km ²)	3.165	2.588	3.130	0.445	0.662
潮滩 (km ²)	1.312	0.604	0.943	0.169	0.163
合计 (km ²)	13.475	8.164	7.462	1.336	1.801

■ 风险源分析

选定干旱、寒潮、风暴潮和虫灾4类频发性灾害作为主要风险源。

南五岛主要自然灾害发生频次的历史统计


	干旱	寒潮	风暴潮	虫灾				
				南长山	北长山	大黑山	小黑山	庙岛
特大级（%）	12.0	16.0	8.0	—	—	—	—	—
重大级（%）	14.0	20.0	12.0	12.5	12.5	10.0	7.5	5.0
较大级（%）	22.0	56.0	18.0	22.5	22.5	17.5	20.0	20.0
总概率（%）	48.0	82.0	38.0	35.0	35.0	27.5	27.5	25.0



■ 受体分析

受体选择是针对对风险因子较敏感或在生态系统中具有重要地位的指示性物种或种群。鸟类应是能反映南五岛区生态状况最直接的物种。

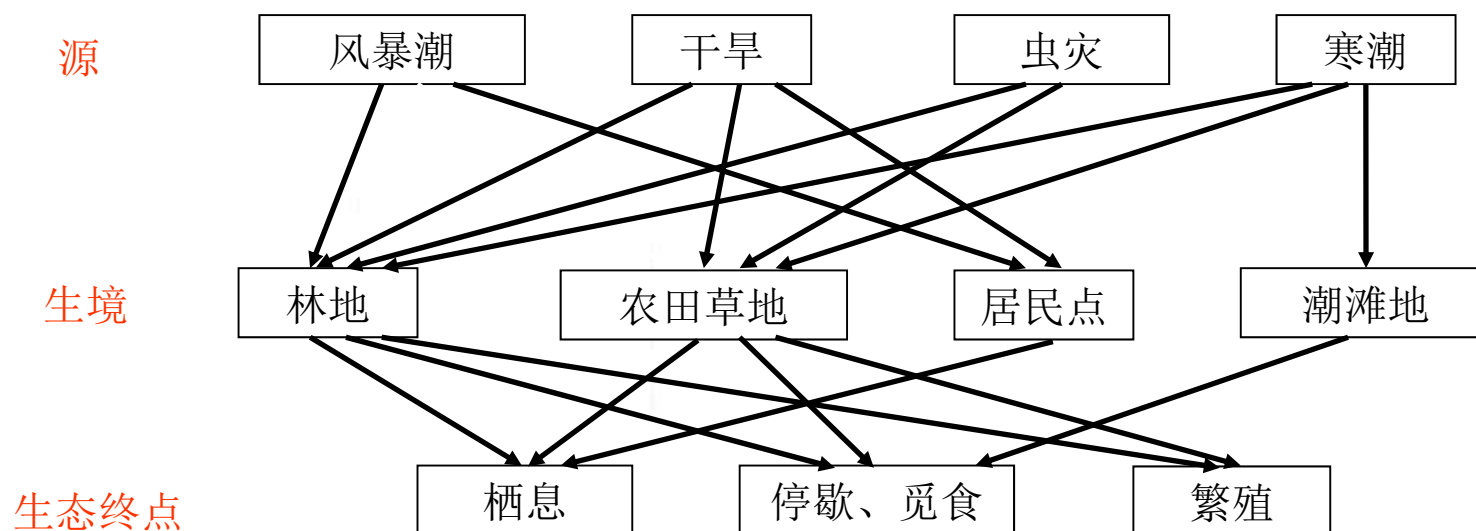
生态终点指在危害性和不确定性风险因素作用下，风险受体可能受到的损害。南五岛受体相应的生态终点为：鸟类种群及数量的减少、珍稀鸟类在本区的消失。



2. 暴露与危害分析

■ 暴露分析

分析4类概率型源到鸟类受体的接触暴露途径，构建风险源—生境—生态终点的暴露—响应概念模型。



风险源—受体—生态终点的暴露途径示意图

■ 危害分析

海岛概率型风险评价的危害分析采用“区域潜在生态损失度”度量法，计算公式为：

$$D = \sum_{i=1}^n (S_i / S) D_i = \sum_{i=1}^n (S_i / S) E_i F_i$$

式中，生态指数 $E = \sum_{i=1}^n \gamma_i U_i$ 由生物多样性指数 $DI = \sum_{i=1}^n N_i / N$ 和物种重要性指数 $EI = \sum_{i=1}^n \sum_{j=1}^m \alpha_j C_{ij} / C_j$ 联合计算得到；脆弱性指数 F 是通过生境脆弱性排序、归一化计算得到。

3. 风险计算

本文采用遥感解译手段和地方性统计数据相结合的方法，利用本文发展的复合风险表征模型进行风险表征。

遥感解译手段：

海岛面积；

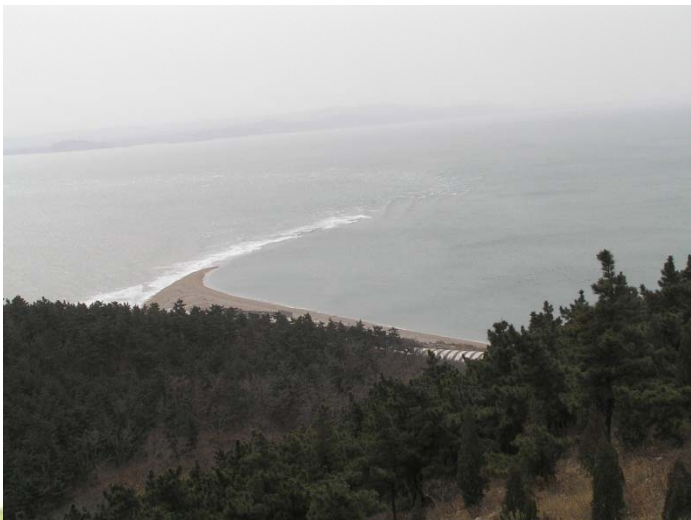
各类生境的面积、分布等。

地方性统计数据（长岛南五岛现场调研、踏勘）：

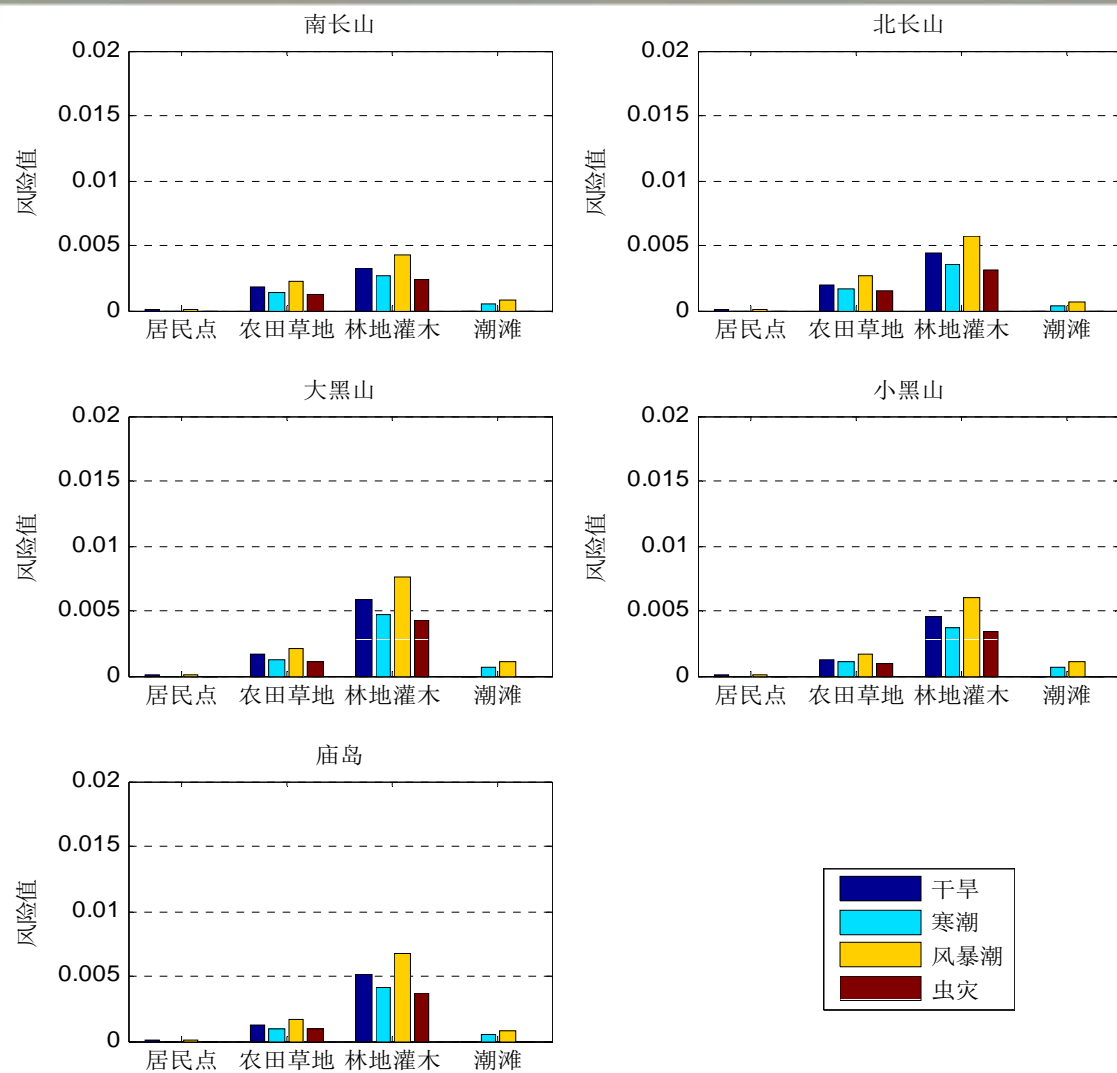
自然灾害历史发生概率统计、分布；

区域鸟类种数及其生活习性；

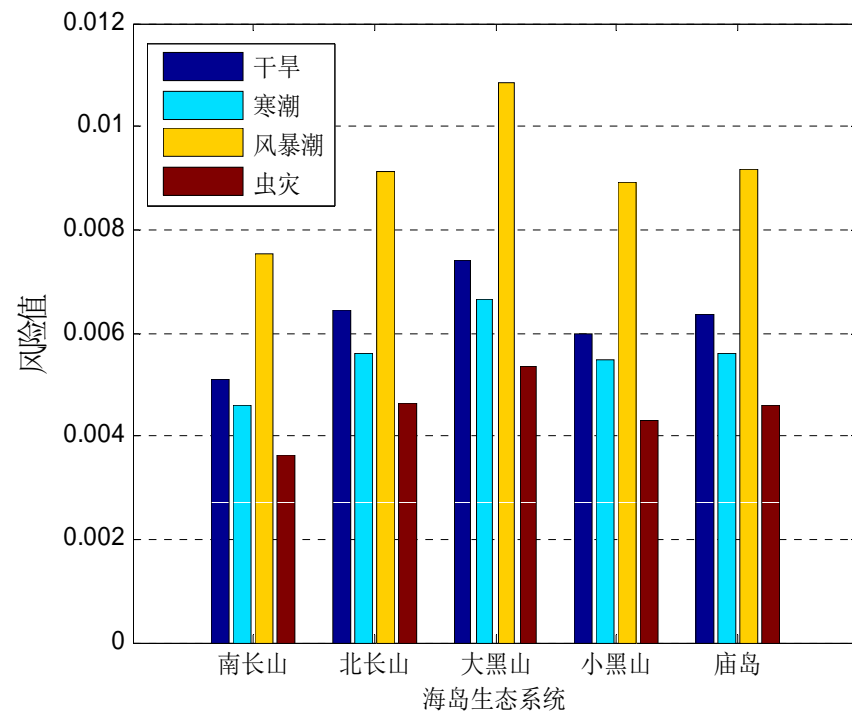
国家保护级鸟类种数及其生活习性等。



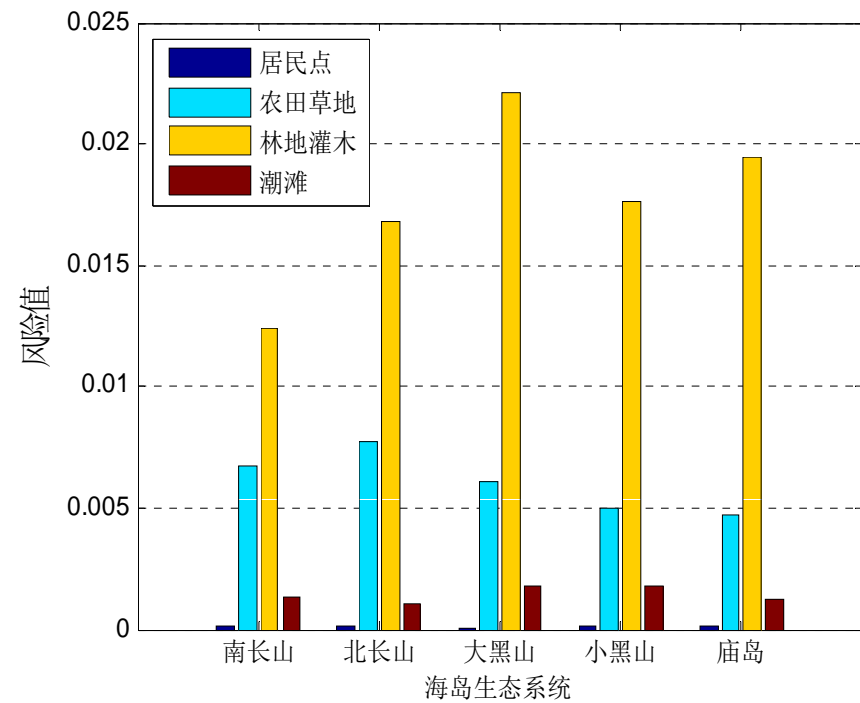
■ 评价结果



南五岛各岛生境潜在风险比较

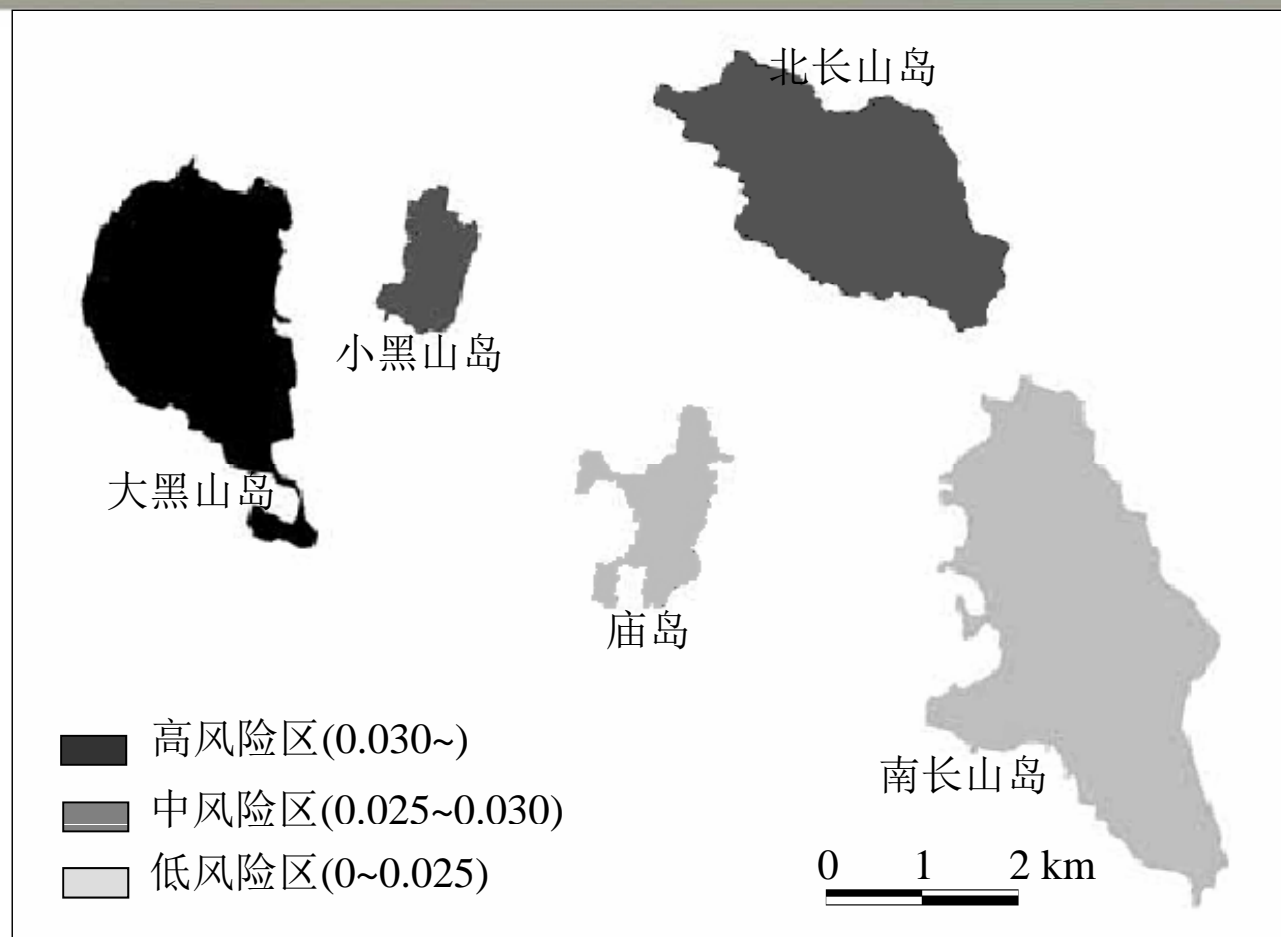


南五岛四类自然灾害风险比较



南五岛四类生境风险状况比较





长岛保护区南五岛自然灾害生态风险状况分级图

4. 风险管理

① 针对区域**首要风险源风暴潮**的风险管理对策

- 构筑防潮和防波堤，消减风暴潮引发的大浪及海水倒灌对近岸生境的影响
- 构建沿海防护林网，增强暴雨、大风抵御力

② 针对区域风险**最大生境类型林地灌木**的风险管理

- 调整林木种类和林带结构，加强水土保持林种植
- 大力建设蓄水池，储蓄天然降水
- 加强林地管理，减少拾柴、砍伐等人为破坏

③ 针对区域**综合风险**的管理对策

主要是加强大黑山岛抵御自然灾害的能力，把区域防治自然灾害的工作重点首要放在大黑山岛上



欢迎交流，实现信息学与遥感的交叉

马毅: mayimail@fio.org.cn

张杰: zhangjie@fio.org.cn

