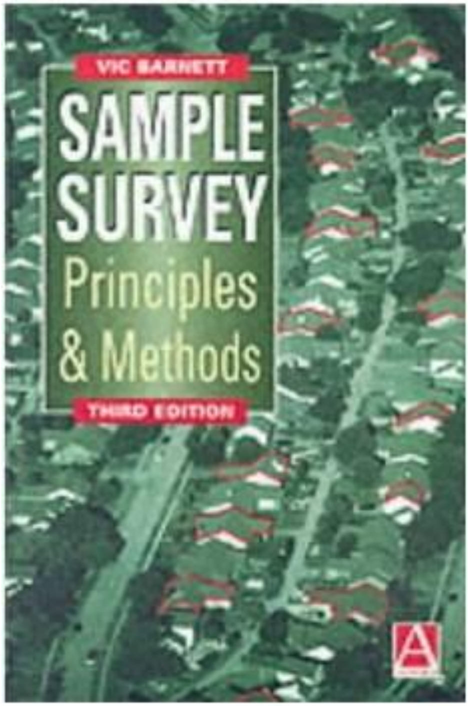
Sample survey



Barnett, Vic. 2002. Sample Survey: Principles & Methods (3rd ed.). New York: Oxford University Press

Inference for finite population

- Computation of population characteristics (census)
- Estimation of population characteristics (sample survey)

Census vs. sample survey

- Budget and time
- Coverage
- Accuracy
- Feasibility

Steps in sampling design

- What is the population?
- What are the parameters of interest?
- What is the sampling frame*?
- What size of the sample is needed?
- How much will it cost?

*the list of elements from which the sample is actually drawn

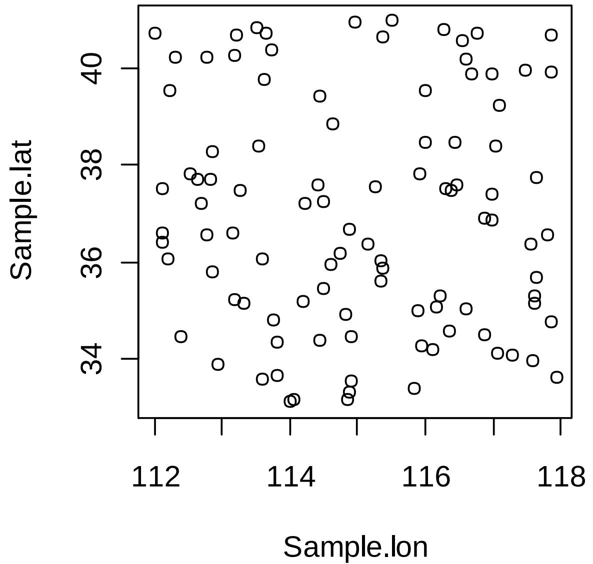
Fundamental Sampling Plans

- Simple Random Sampling (SRS)
- Stratified Sampling
- Systematic Sampling
- Cluster Sampling
- Multistage Sampling

Fundamental Concept of Simple Random Sampling

- A population unit is randomly selected from the population until a set of sample of size "n" is achieved
- At each of the selection process, the remaining population units have an equal chance of being selected
- A set of samples occurs with an equal probability

```r
# Simple Random Sampling for 100 locations
Sample.lat  <- runif(100,  min = 33,  max = 41)
Sample.lon <- runif(100, min = 112, max = 118)
plot (Sample.lon, Sample.lat)
```



Population Characteristics

- Population Total $$Y = \sum_{i=1}^{N} y_i$$
- Population Mean $$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$$
- Population Proportion $$P = \frac{A}{N}$$
- Ratio $$R = \frac{Y}{X}$$

Population characteristics and estimators under SRS

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$Y = \sum_{i=1}^{N} y_i \qquad \hat{Y} = N\overline{y}$$

$$P = \frac{A}{N} \qquad p = \frac{a}{n}$$

$$R = \frac{Y}{X} \qquad r = \frac{y}{x}$$

Quality of sampling

- Accuracy
  - Systematic variance
    - The variation in measures due to some known or unknown influences that “cause” the scores (results) to lean in one direction more than another
- Precision
  - Sampling error
    - The degree to which a given sample differs from the underlying population
    - Sampling error tends to be high with small sample sizes and will decrease as sample size increases

Properties of the estimator under SRS

$\overline{y}$ is an unbiased estimator for $\overline{Y}$ with the Variance

$$Var(\overline{y}) = (1 - f)\frac{S^2}{n};$$

where $f = \dfrac{n}{N}$ and $S^2 = \dfrac{\sum_{i=1}^{N}(y_i - \overline{Y})^2}{N-1}$

Proportion

- Basic Properties of the estimator

Sample proportion $p=\dfrac{a}{n}$ is an unbiased estimator

for the population proportion $P=\dfrac{A}{N}$ with the variance

$$Var(p)=\frac{PQ}{n}\left(\frac{N-n}{N-1}\right)=\frac{PQ}{n}(1-f)$$

where $Q=1\text{-}P$

Sampling error and sample size

Sampling error e when estimating a proportion p with a sample of size n taken from an infinite population

$$\mathrm{var}(p) = \frac{p(1-p)}{n}(1-f)$$

$$e = \sqrt{\frac{p(1-p)}{n}}$$

Confidence intervals

In a sample of 1,000 cages, 280 cages (28 percent) had captured a rodent individual.

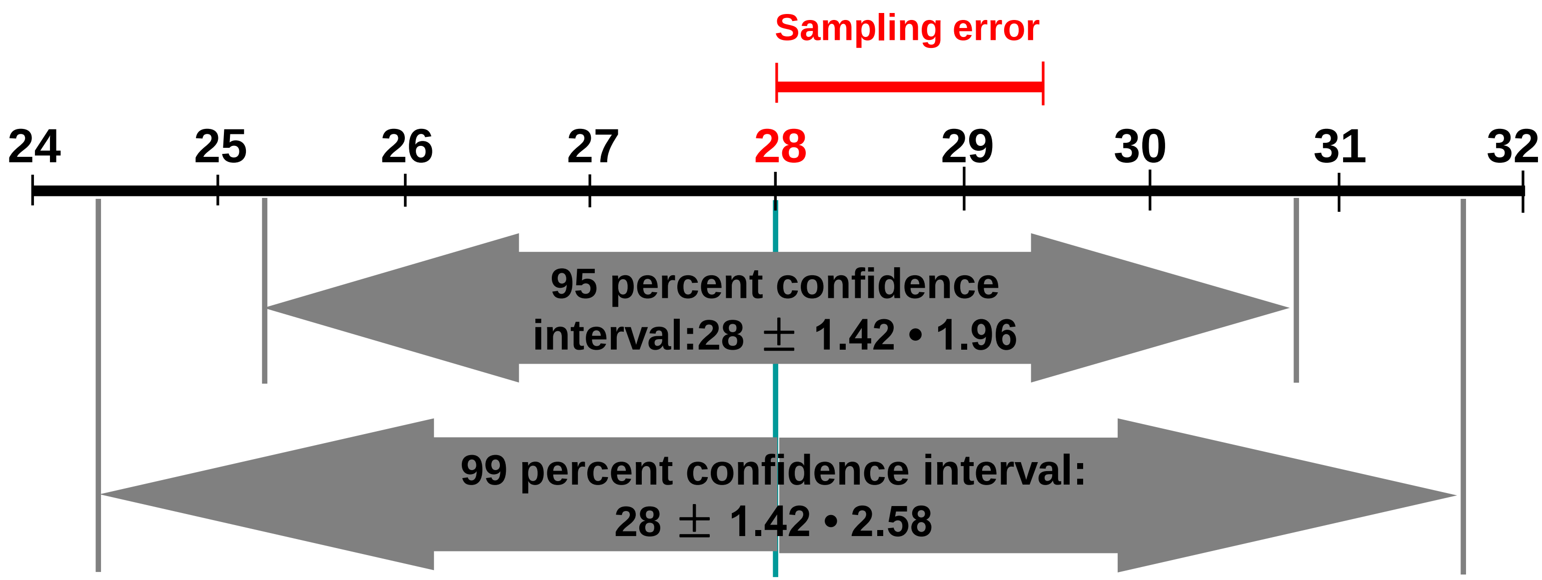
$$e = \sqrt{\frac{0.28 \times 0.72}{1{,}000}} = 0.0142$$

Sampling error is 1.42 percent.

Confidence intervals

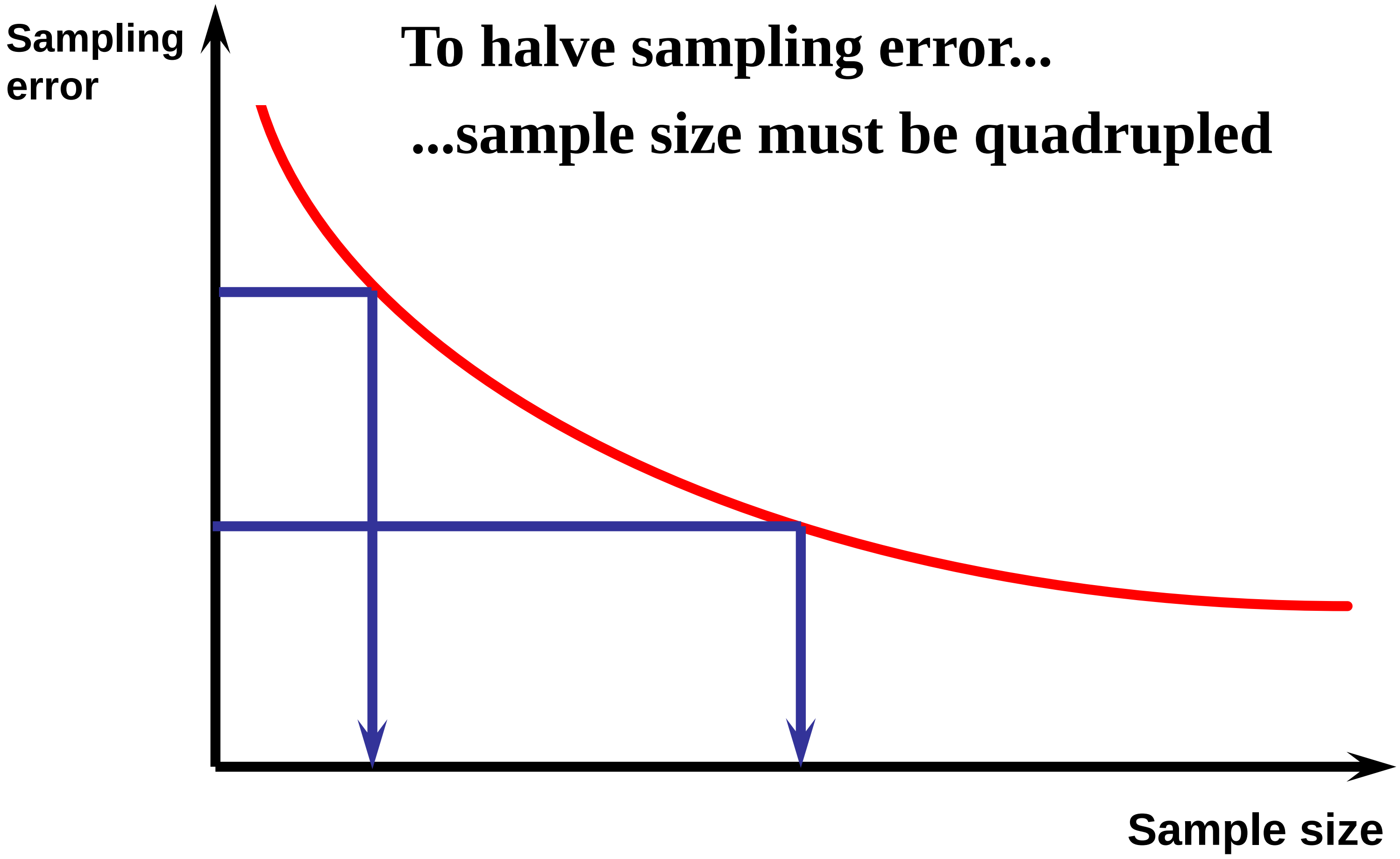
In a sample of 1,000 cages, 280 cages (28 percent) had captured a rodent individual.

Sampling error is 1.42 percent.



Sampling error and sample size

$$e = \sqrt{\frac{p(1-p)}{n}}$$



Sample size and population size

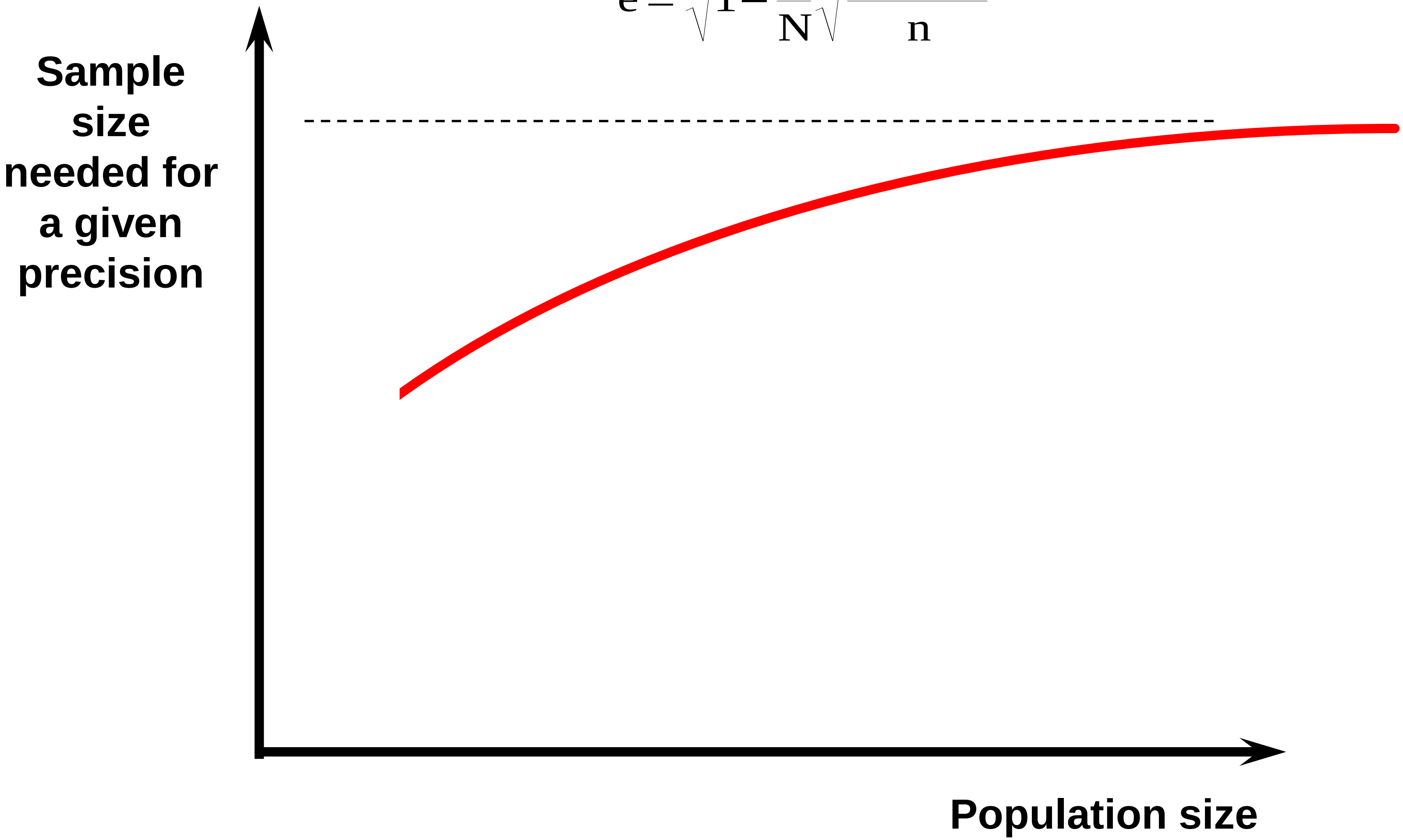
Sampling error e when estimating a proportion p with a sample of size n taken from a population of size N

$$e = \sqrt{1 - \frac{n}{N}} \sqrt{\frac{p(1-p)}{n}}$$

finite population correction

Sample size and population size

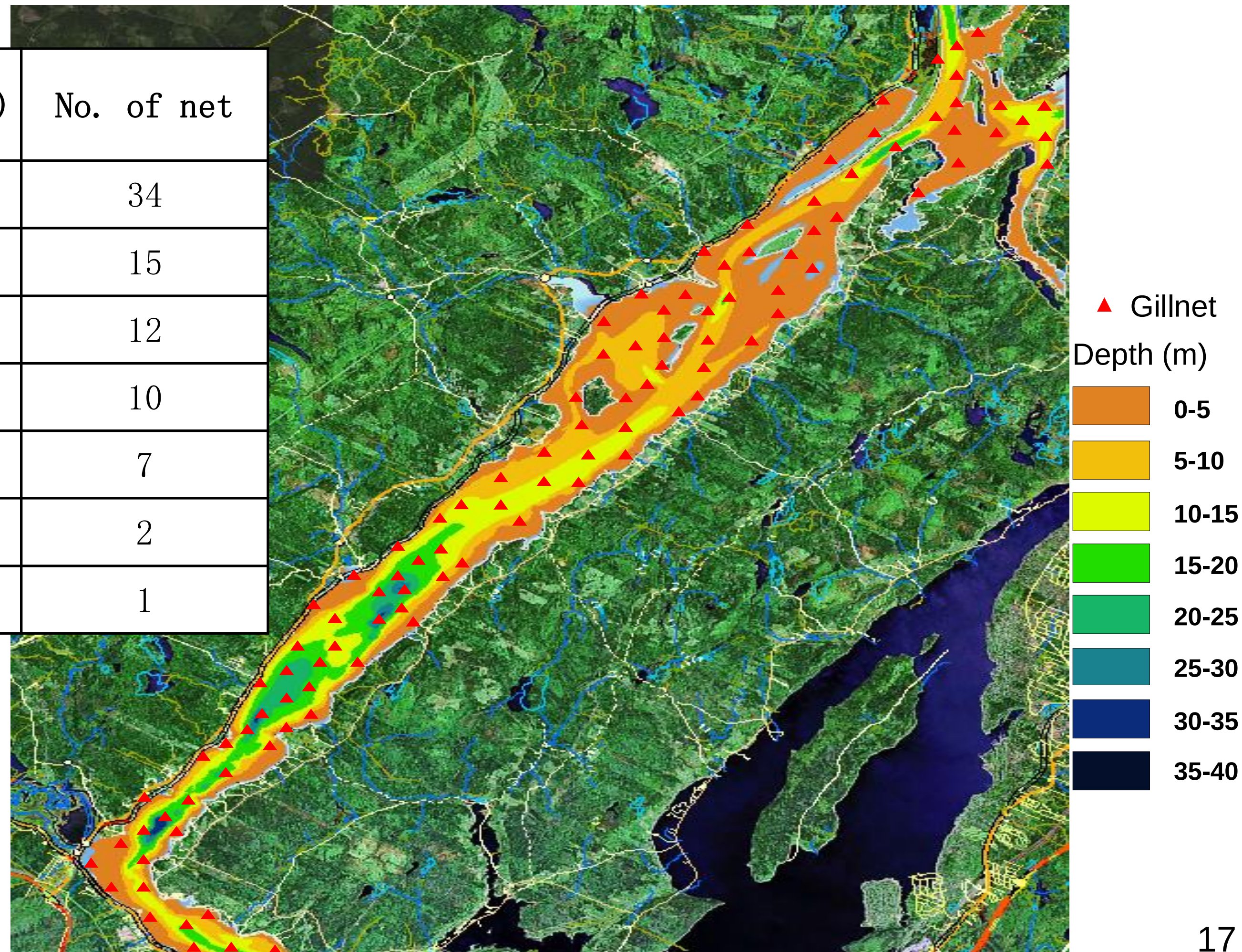
$$e = \sqrt{1 - \frac{n}{N}} \sqrt{\frac{p(1-p)}{n}}$$



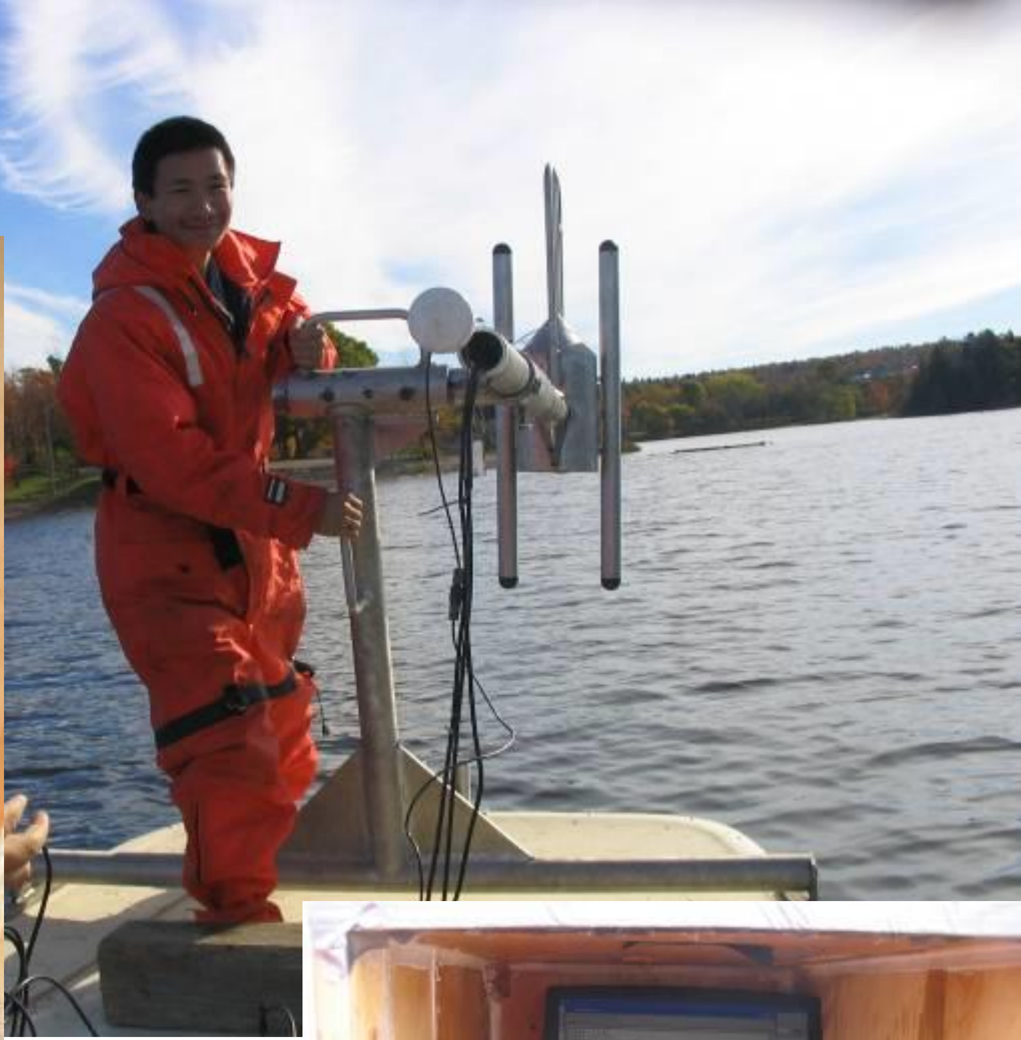
Stratified sampling

Sample the shortnose sturgeon using gillnet following a stratified sample design

Strata	Depth (m)	Area (km)	No. of net
Stratum1	0-5	68	34
Stratum2	5-10	31	15
Stratum3	10-15	25	12
Stratum4	15-20	20	10
Stratum5	20-25	15	7
Stratum6	25-30	5	2
Stratum7	>30	2	1

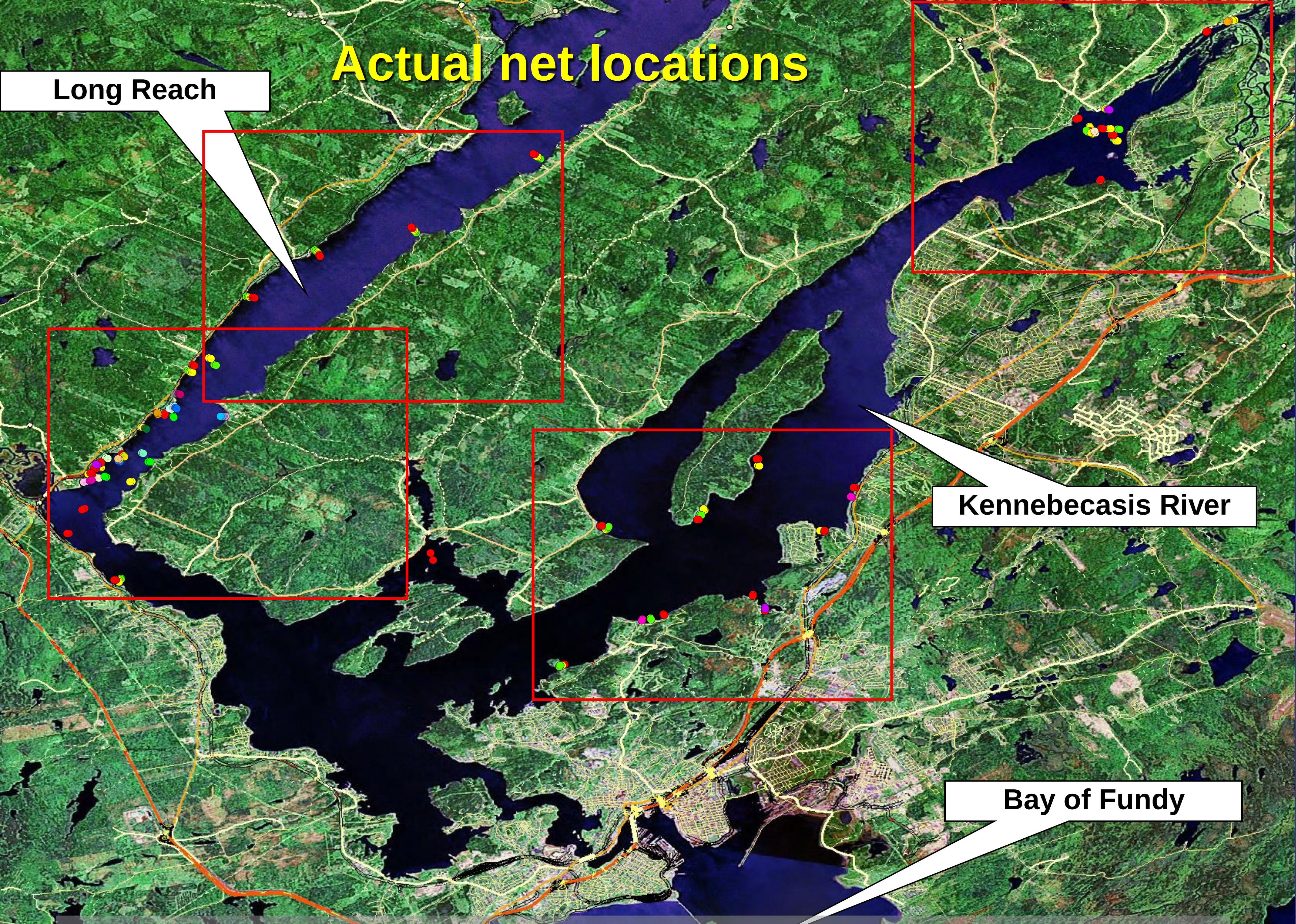


Sidescan sonar recording river bottom substrate and depth

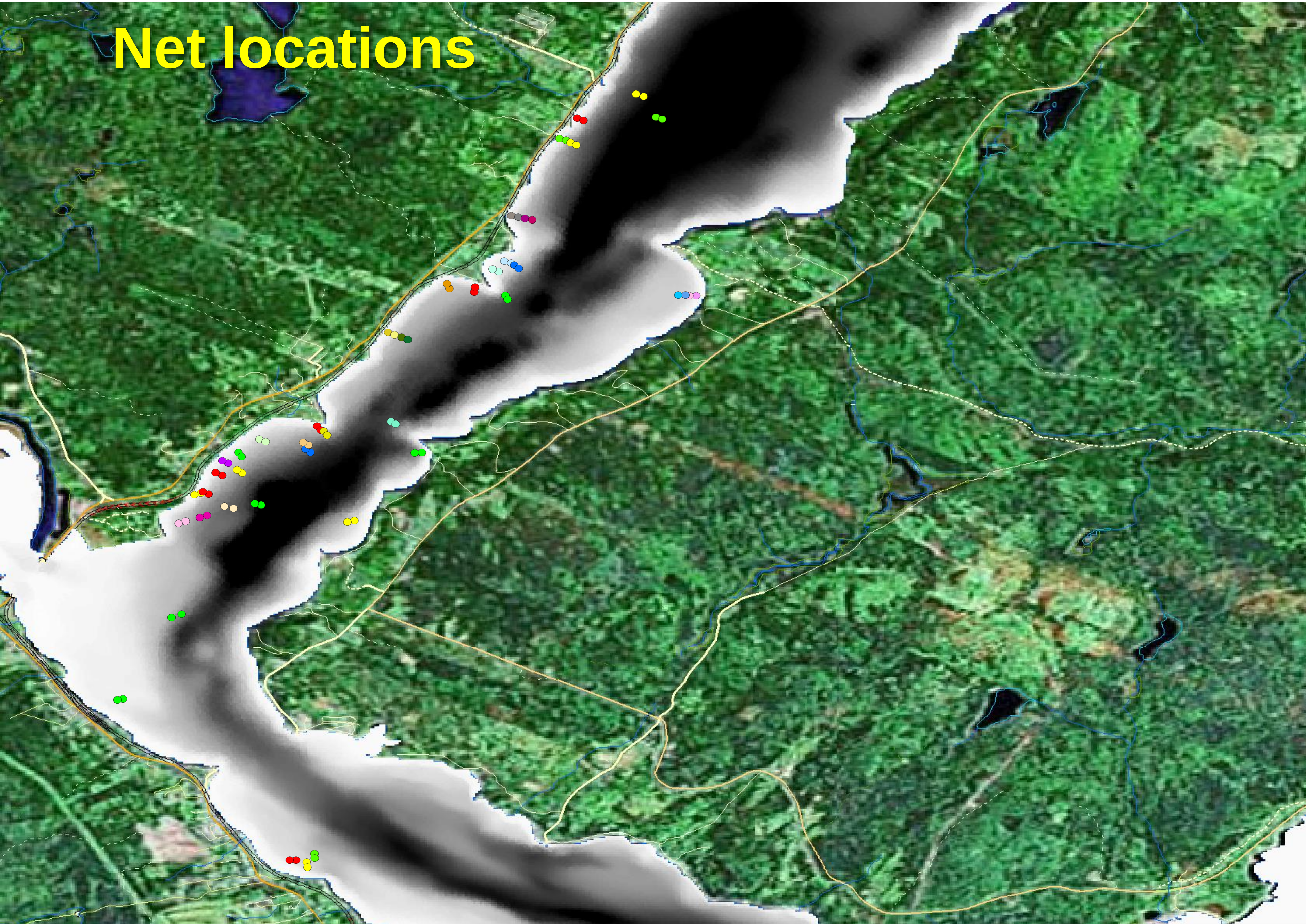


Detect depth

What we got from deep water



The locations of the gillnets that were set in Saint John River in 2005. The squares indicate the areas of upper Long Reach, Grand bay-Westfield, upper Kennebecasis River, and lower Kennebecasis River. The color points show the location of buoys at the two ends of the gillnets.


Net locations

Advantages of stratified sampling

1. Ensures that each strata (subpopulation) is well weighted

2. Can result in estimates with smaller standard errors if sampling is well allocated

3. Small Stratum will not be missed

Disadvantages of stratified sampling

1. More complicated than SRS

2. Need to identify strata ahead of time.
   Hence, more information needed prior to sampling than than for SRS

Systematic Sampling

A type of probability sampling in which every kth member of the population is selected

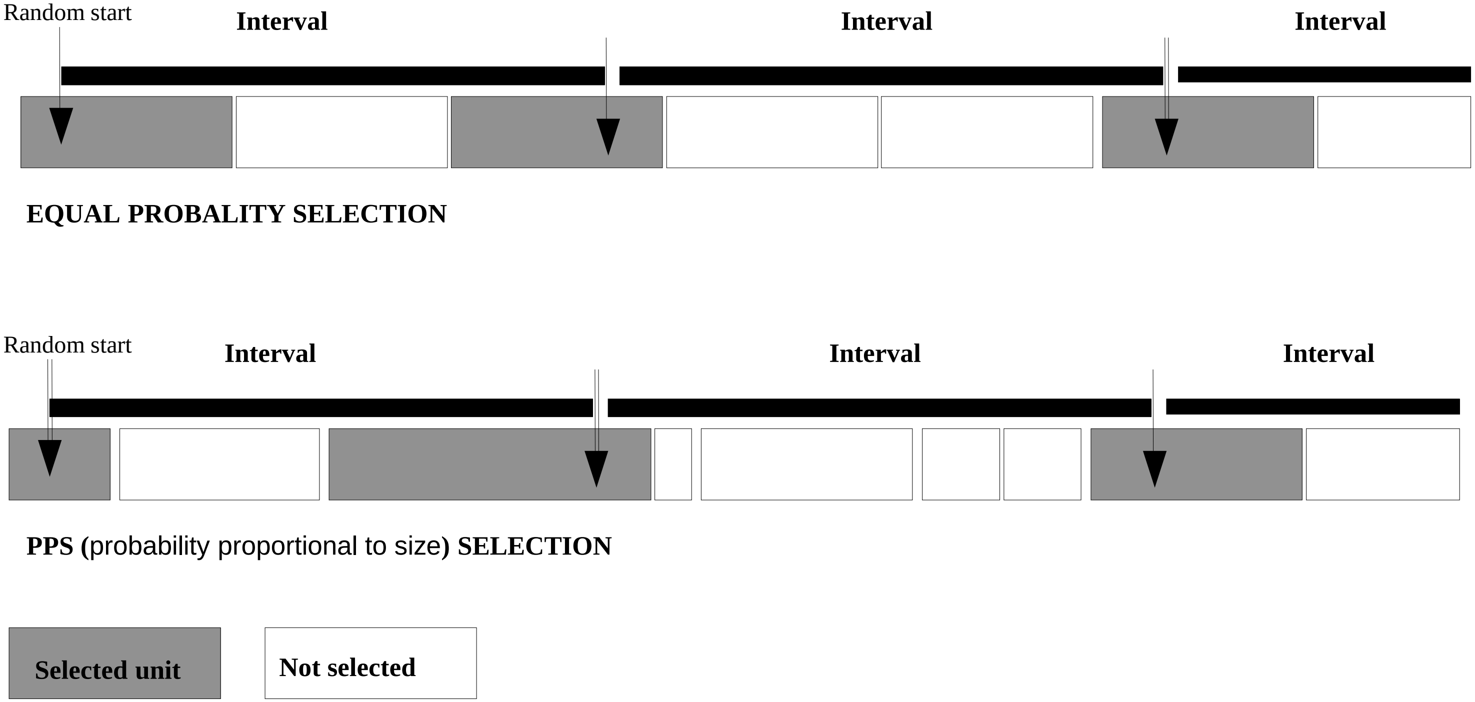
$k=N/n$

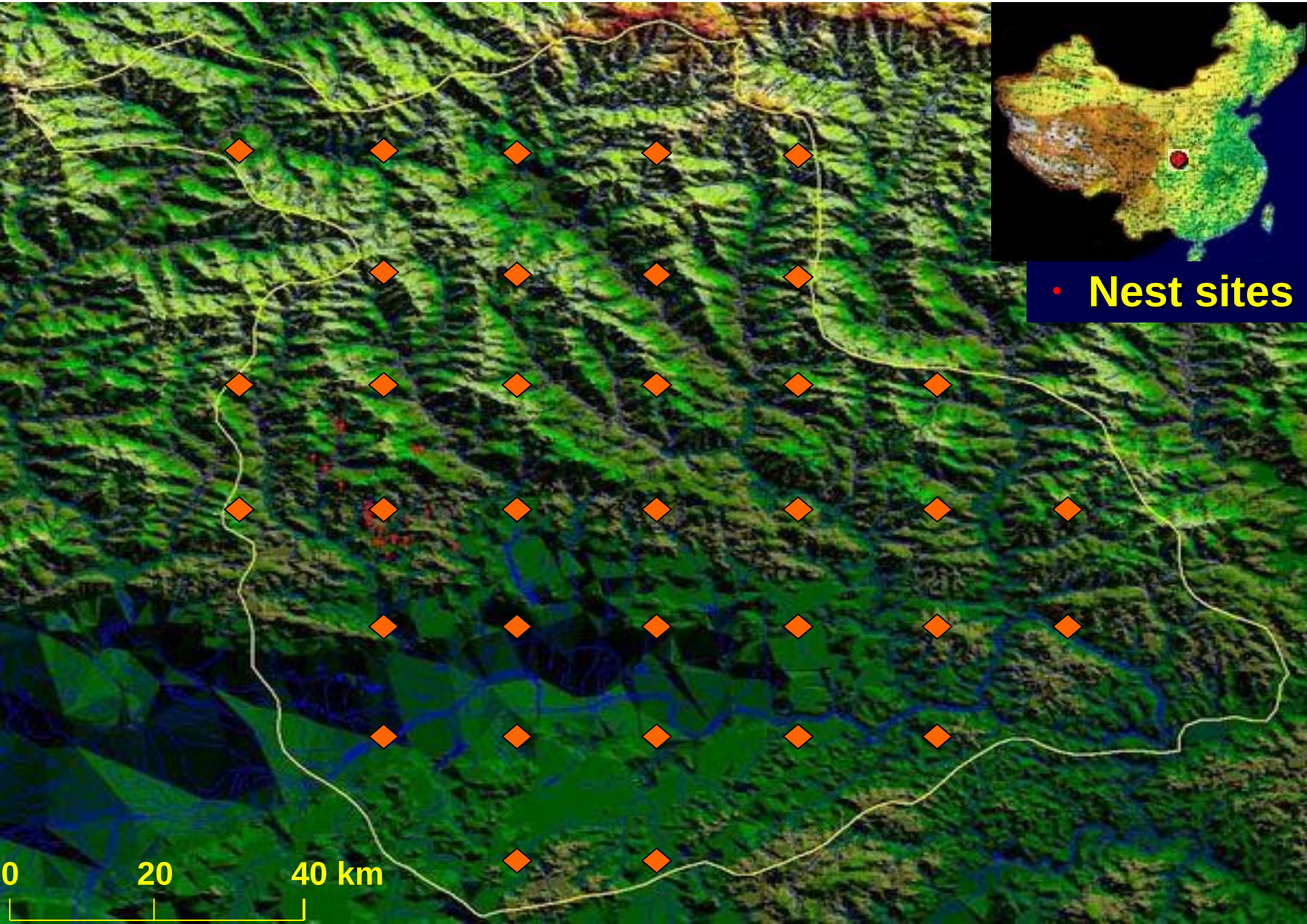
N = size of the population

n = sample size

Systematic Sampling

Illustration of systematic sampling procedure.



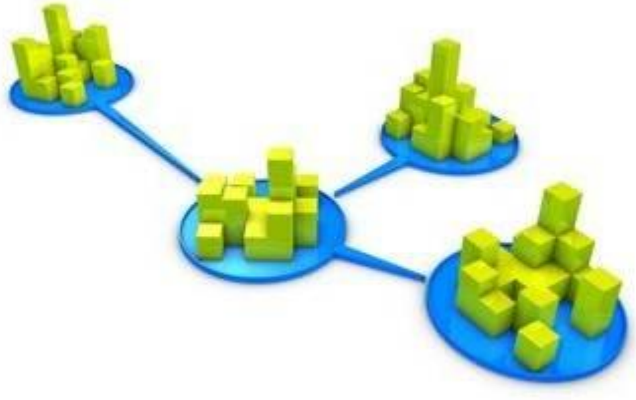

Nest sites
0
20
40 km

Systematic Sampling

- Advantages
  - Simplicity. It allows the researcher to add a degree of system or process into the random selection of subjects.
  - The assurance that the population will be evenly sampled. There exists a chance in simple random sampling that allows a clustered selection of subjects. This is systematically eliminated in systematic sampling
- Disadvantage
  - The process of selection can interact with a hidden periodic trait within the population. If the sampling technique coincides with the periodicity of the trait, the sampling technique will no longer be random and representativeness of the sample is compromised.

Cluster Sampling

- The sampling unit contains more than one population element.
- For simple cluster sampling, each cluster contain the same number of elements; clusters are chosen randomly; all selected elements are included in the sample.



Cluster Sampling

- Suppose there are A clusters in the population; a clusters are selected.
- Each cluster contains B elements.
- Thus, the sample size is: $n=aB$
- Population size is $N=AB$

Cluster Sampling

- Sample mean is also the mean of the a cluster means.

$$\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j = \frac{1}{aB}\sum_{\alpha=1}^{a}\sum_{\beta=1}^{B} y_{\alpha\beta} = \frac{1}{a}\sum_{\alpha=1}^{a}\left(\frac{1}{B}\sum_{\beta=1}^{B} y_{\alpha\beta}\right)$$

$$= \frac{1}{a}\sum_{\alpha=1}^{a} \bar{y}_{\alpha}$$

Cluster Sampling

In terms of variance of the estimator, the situation is exactly the same as in SRS.

Property: An unbiased estimator of sample variance is:

$$Var(\bar{y}) = Var(\frac{1}{a}\sum_{\alpha=1}^{a}\bar{y}_{\alpha})$$

$$= \frac{1-f}{a}\frac{1}{A-1}\sum_{\alpha=1}^{A}(\bar{y}_{\alpha}-\bar{y})^2$$

$$= \frac{1-f}{a}\frac{S_a^2}{B}$$

Key points

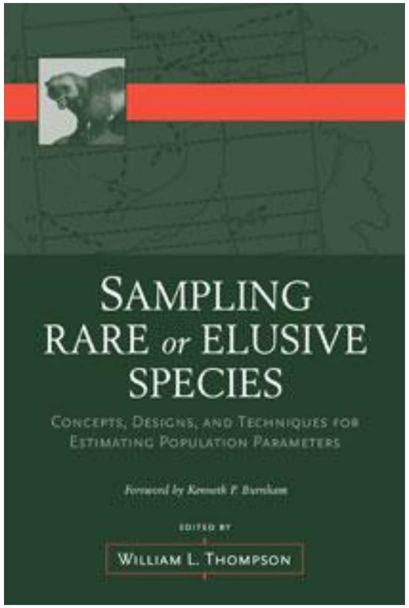
- The precision of the estimator depends on between cluster variance only. Thus, when selecting clusters, we want to “minimize” between variance, or equivalently, “maximize” within variance.

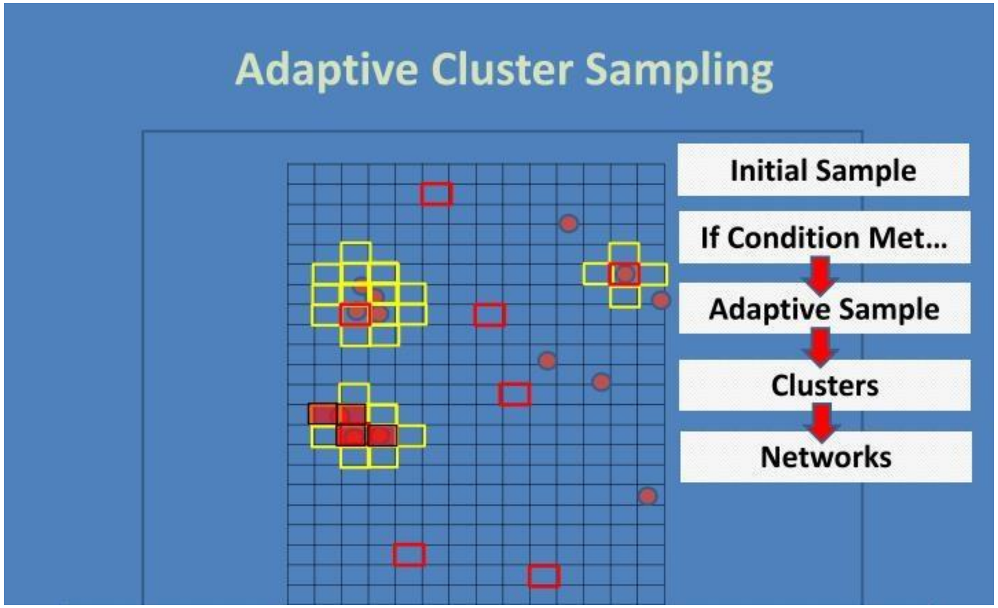
- Unfortunately in many cases, clusters are naturally formed. For example, county, classes, etc.

Multistage Sampling

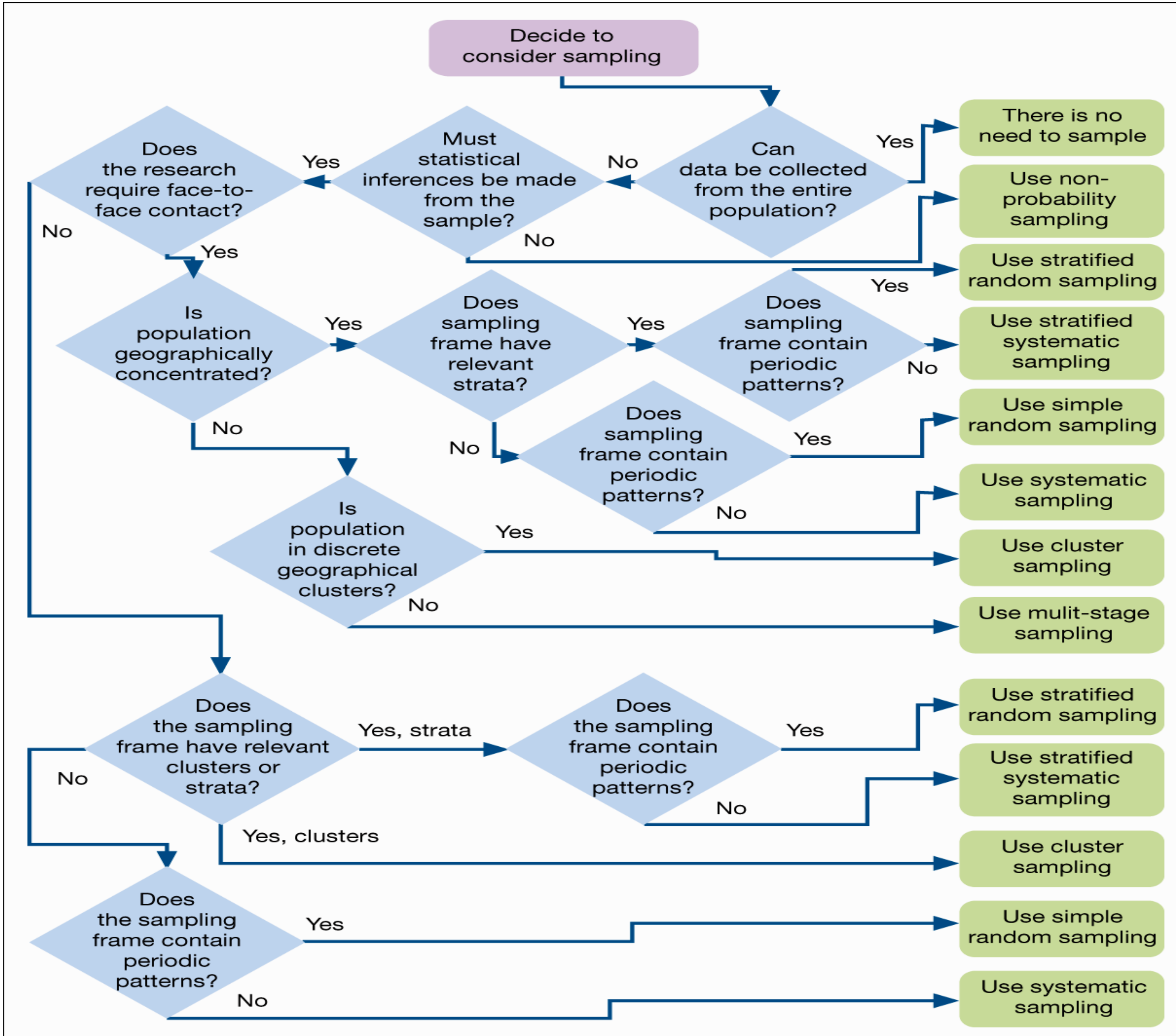
- Stage 1
  - randomly sample clusters (or apply other sampling methods)
- Stage 2
  - randomly sample individuals from the cluster selected

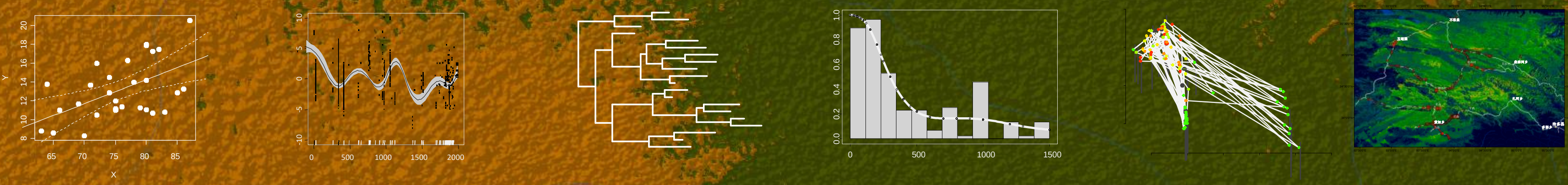
Multistage Sampling





Overview of probability sampling

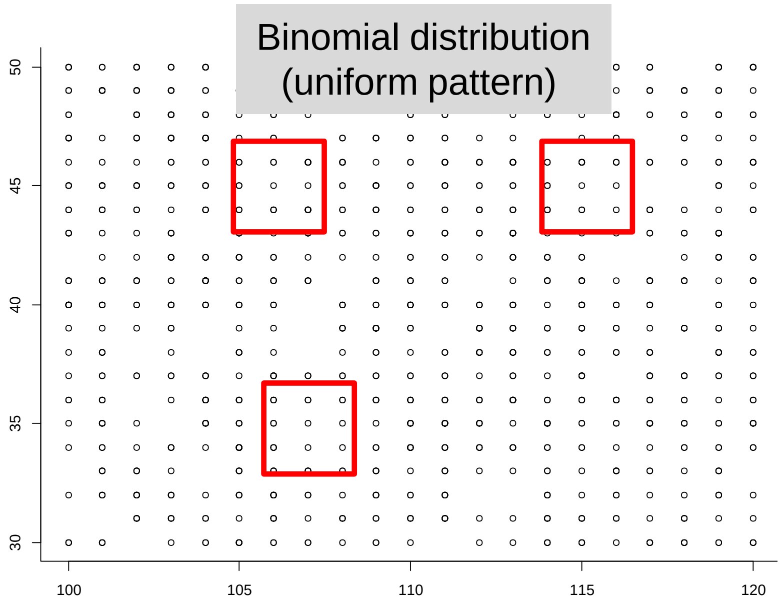


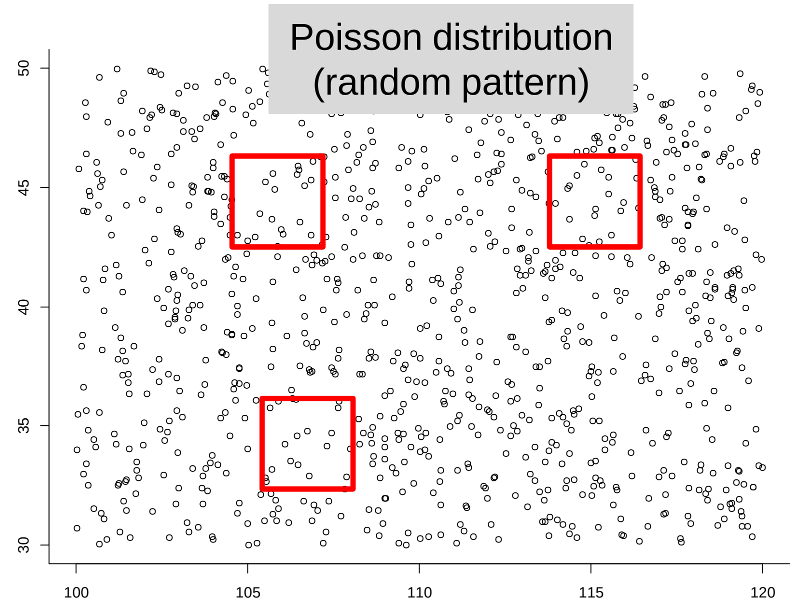


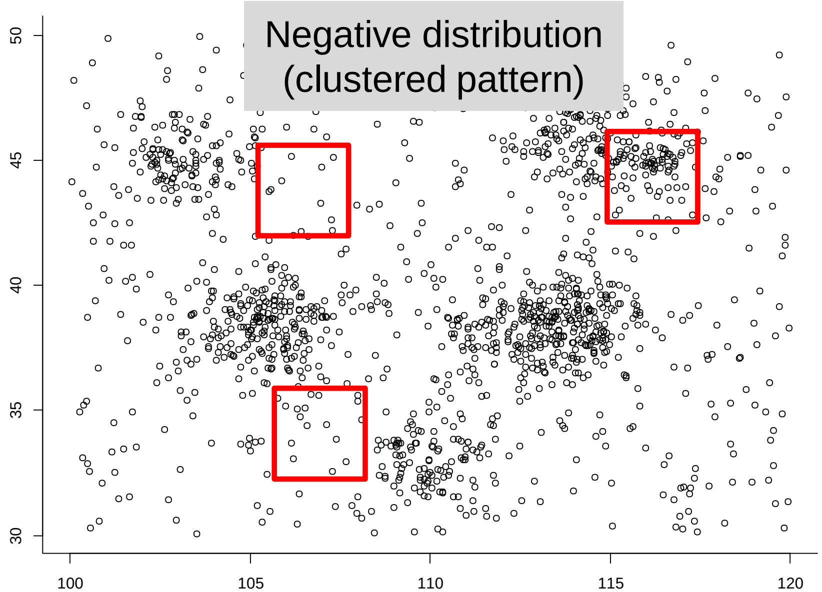
Wildlife survey

- Sampling an infinite population to estimate animal distribution and abundance, as well as population dynamics.
- Estimation accuracy depends on the spatial distribution patterns of the animals.
- Survey protocols are important.

Spatial point patterns

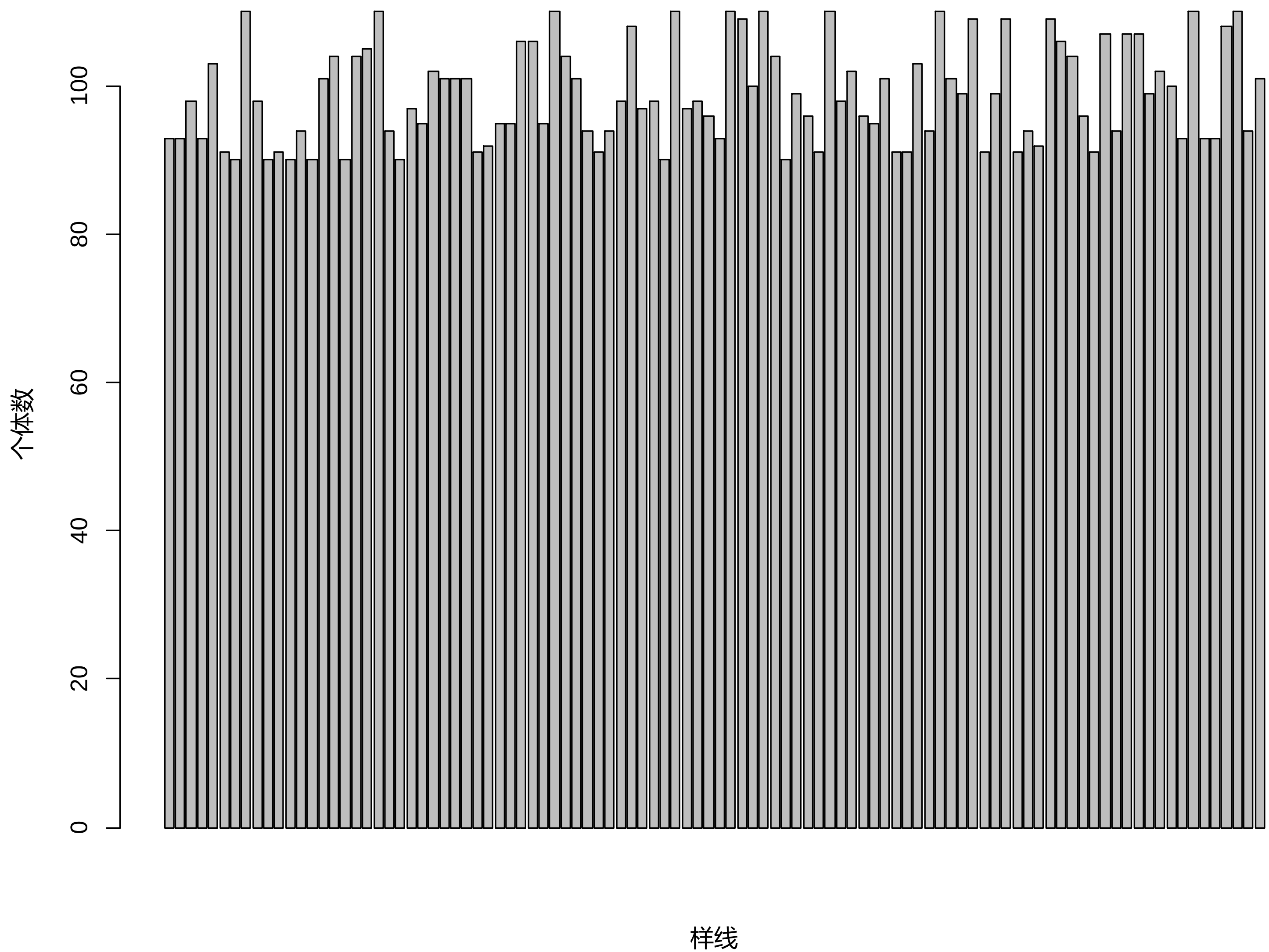




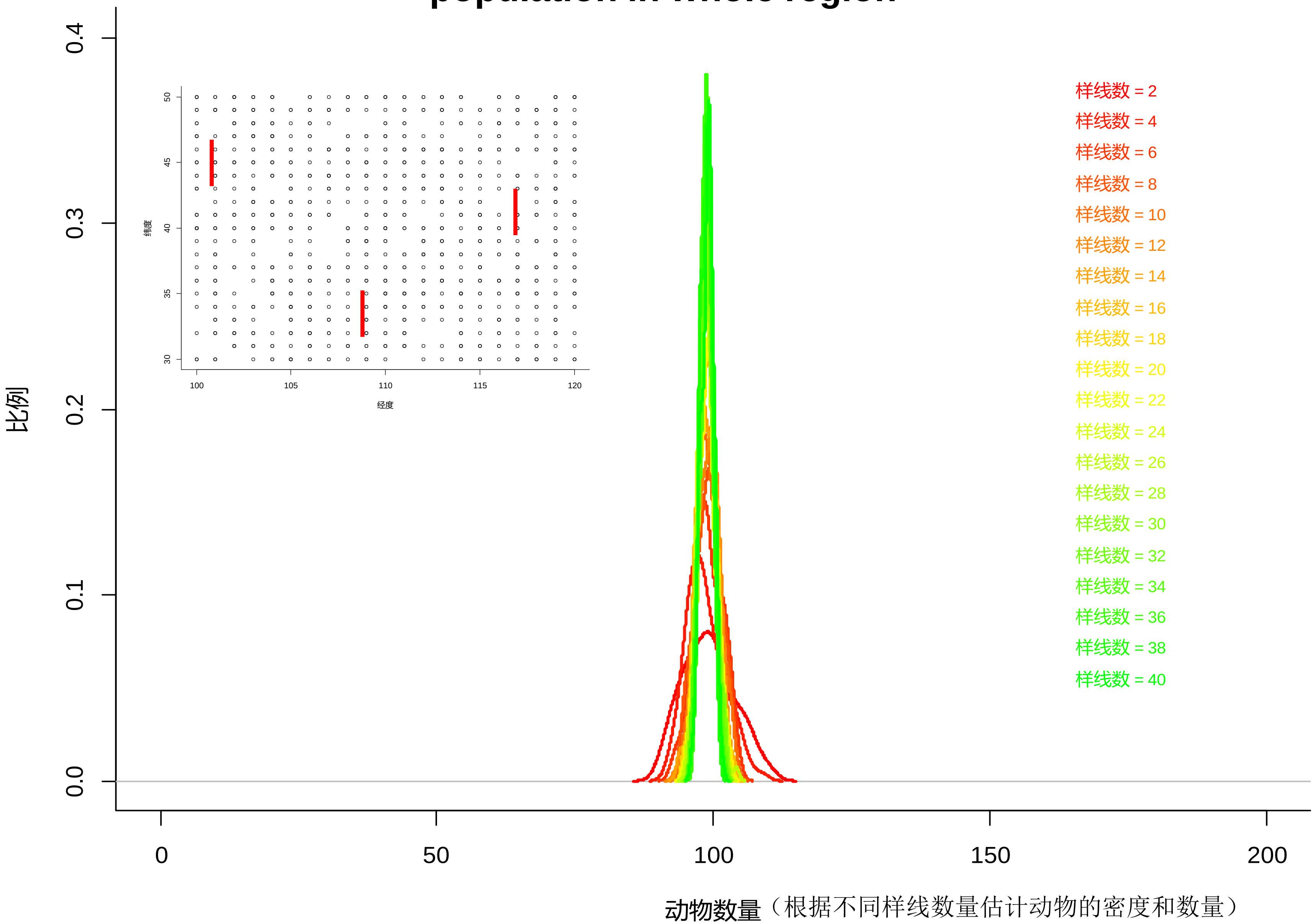


One question:
how many samples (survey routes) needed for estimating population size

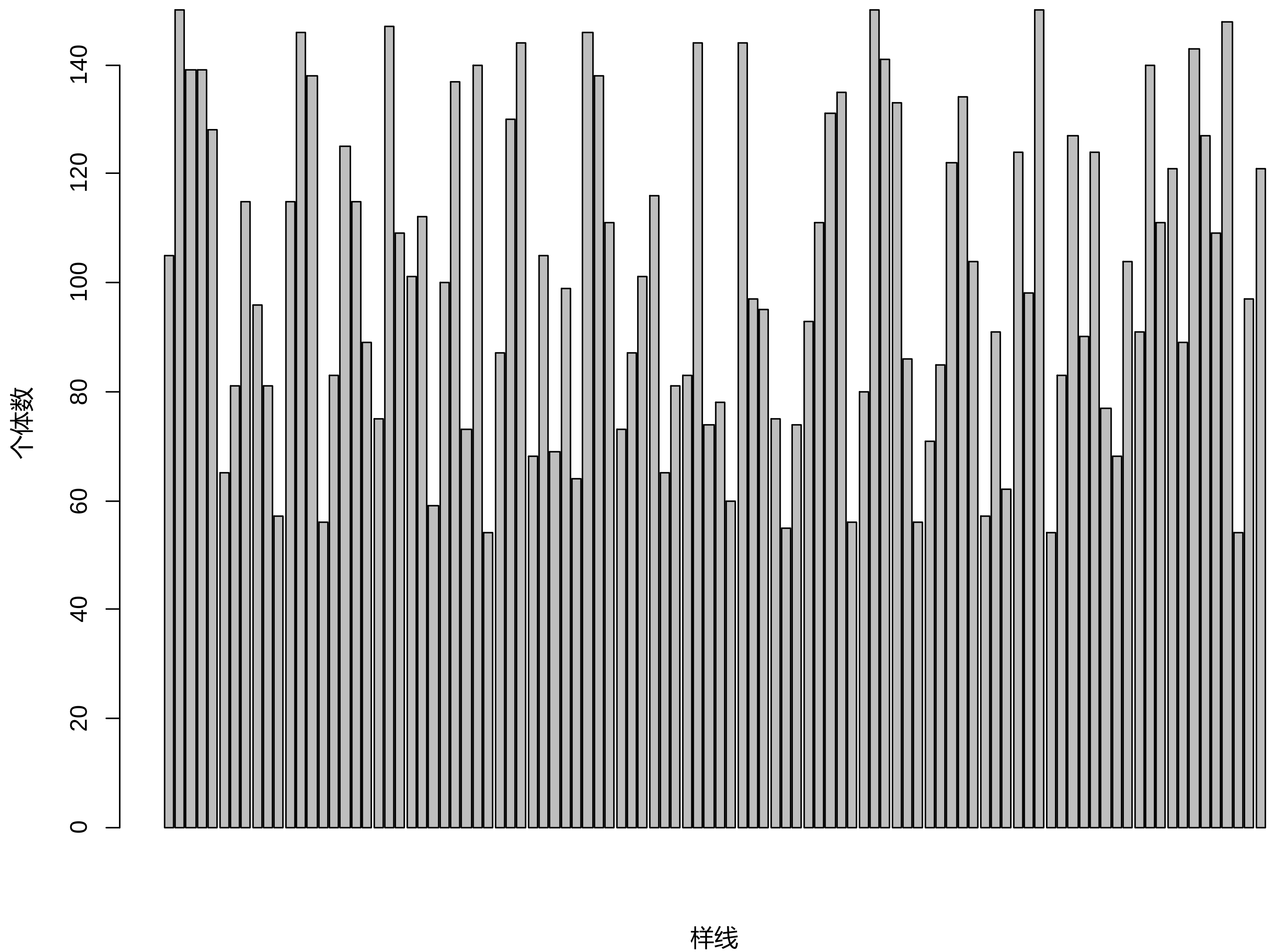
Uniform distribution



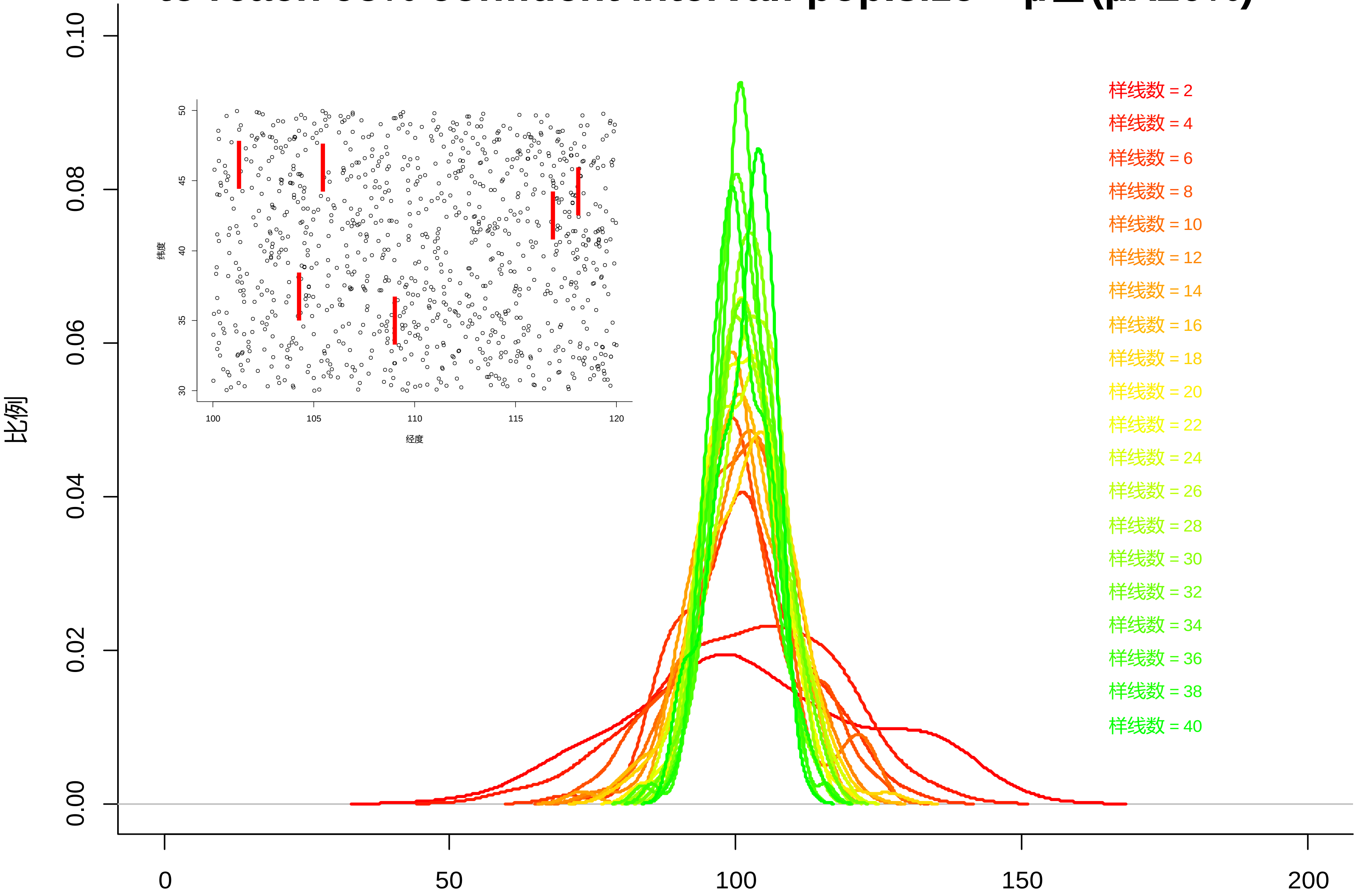
For uniform distribution, three survey routes are enough for estimating the population in whole region



Random distribution (variance = 10μ)

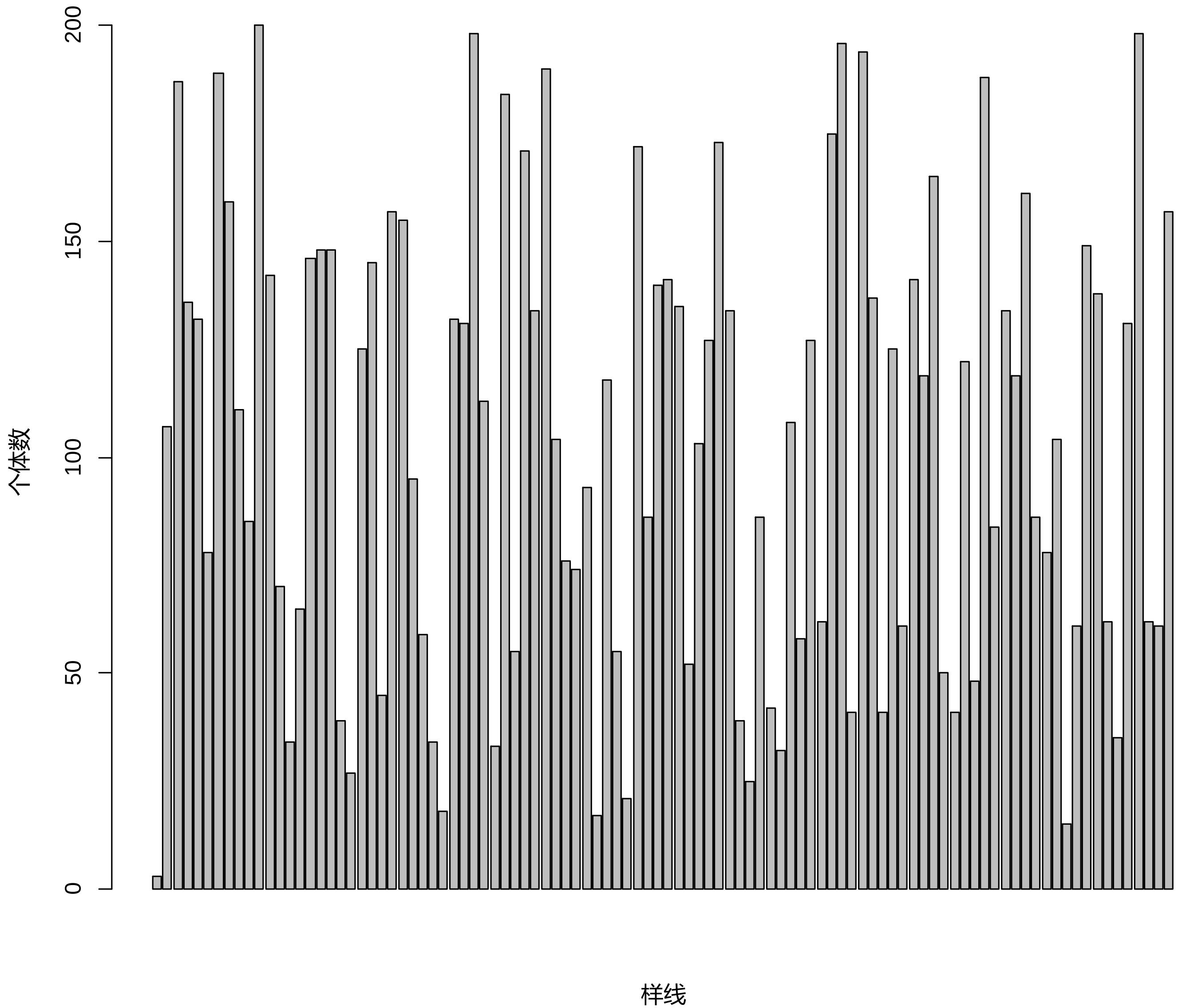


Random distribution (variance = 10μ), 10-16 survey routes needed to reach 95% confident interval: pop.size = μ±(μX20%)

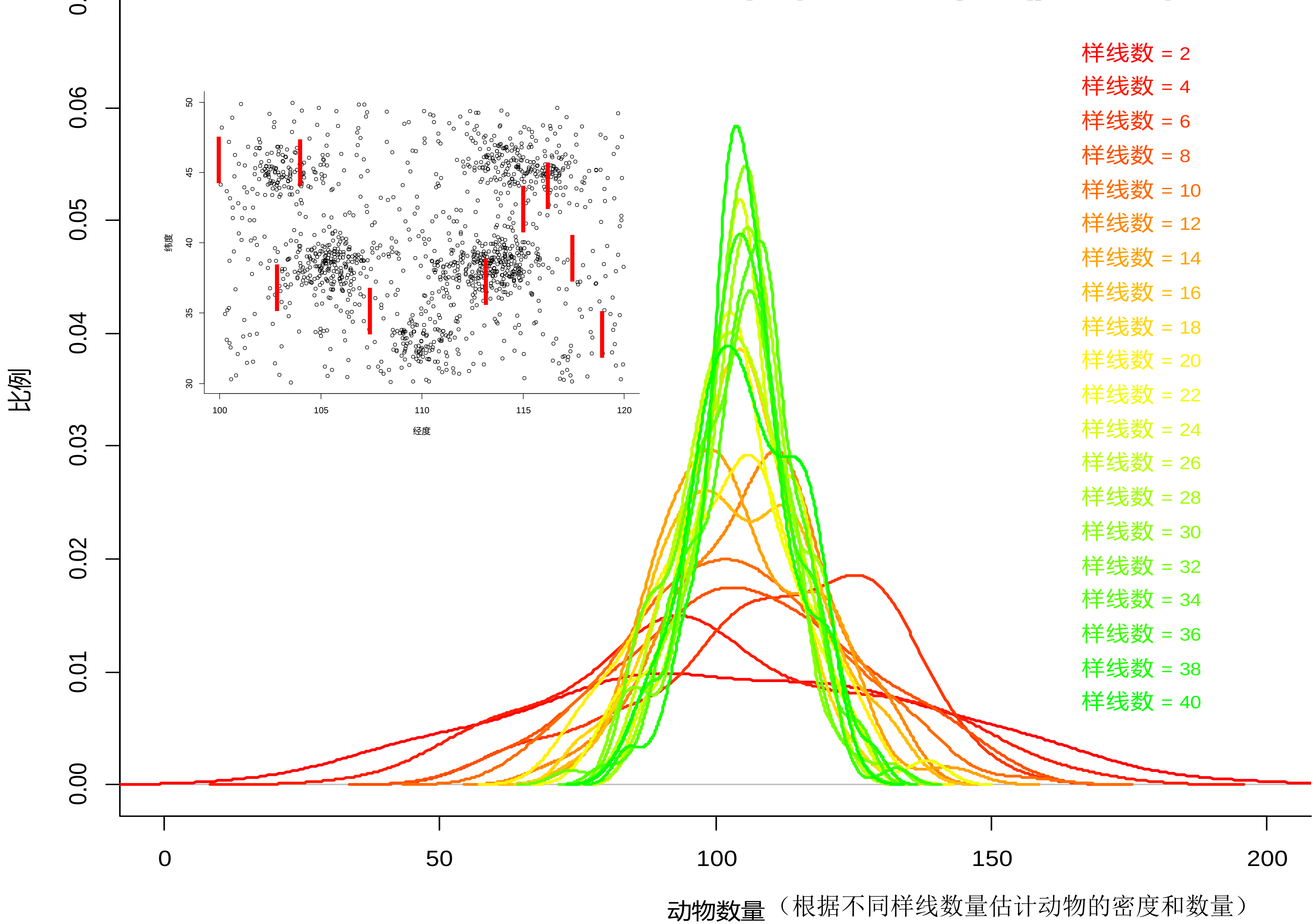


动物数量 （根据不同样线数量估计动物的密度和数量）

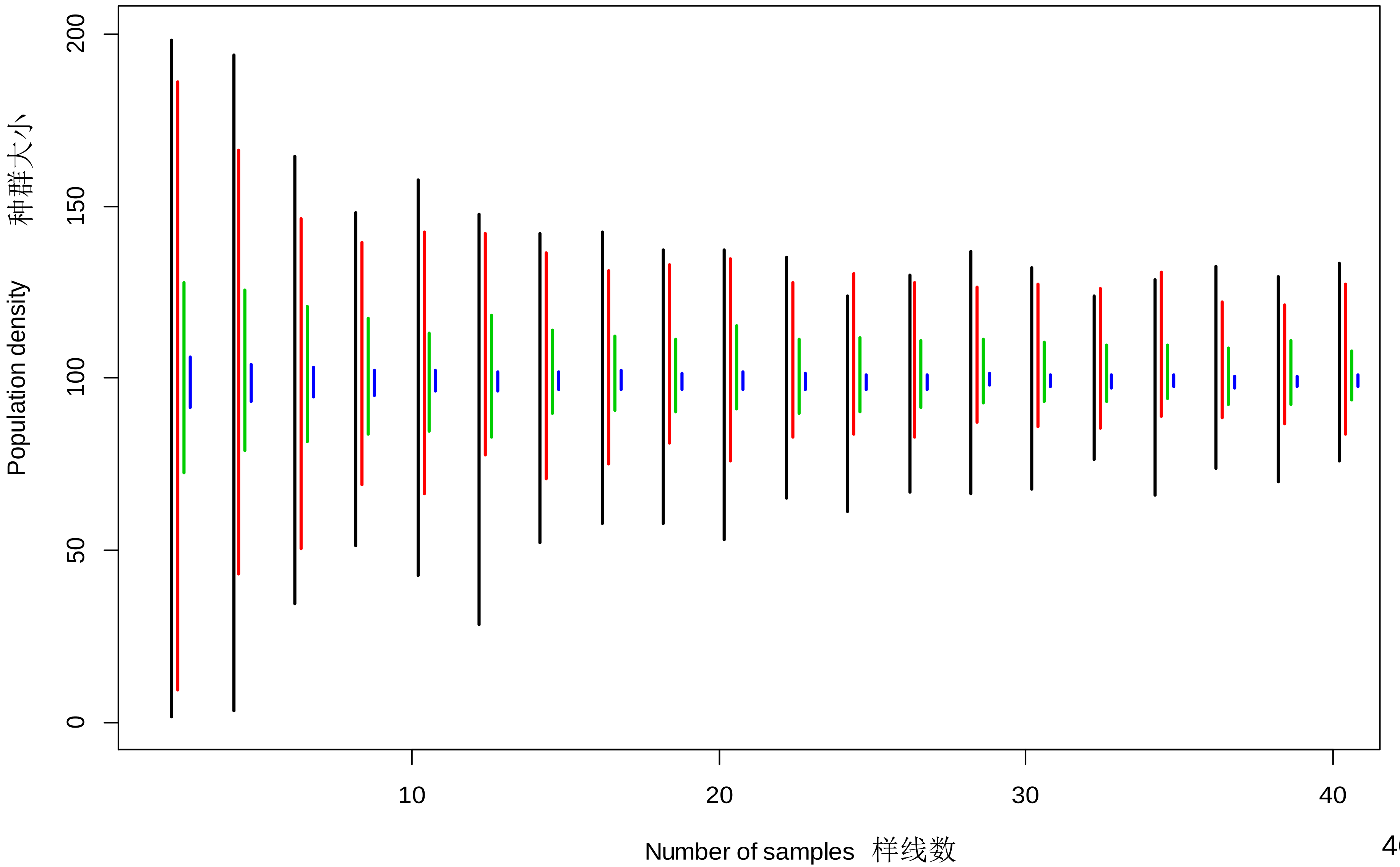
Cluster distribution (variance = 30μ)

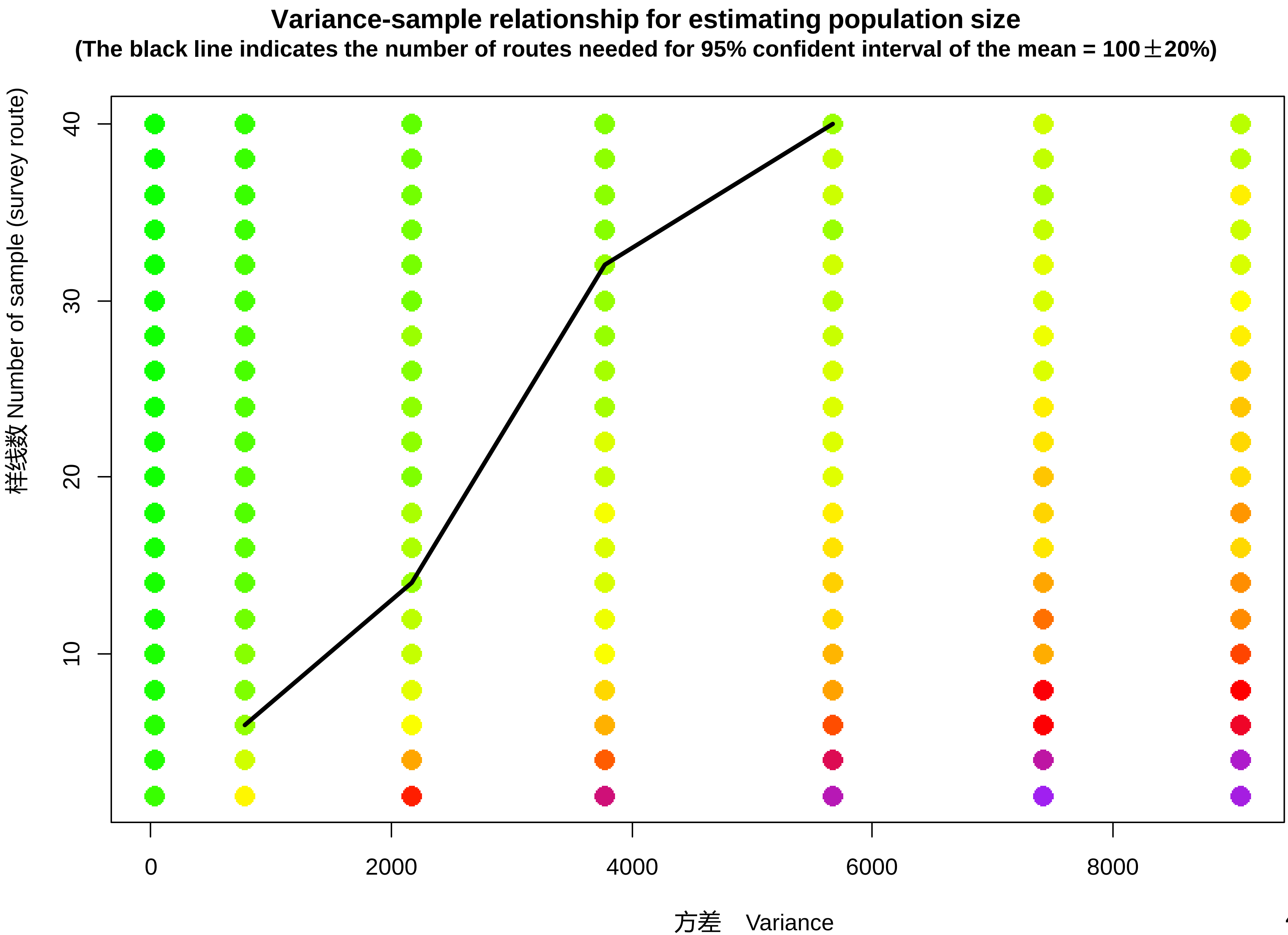


Cluster distribution (variance = 30μ), 24-32 survey routes needed to reach 95% confident interval: pop.size = μ±(μX25%)



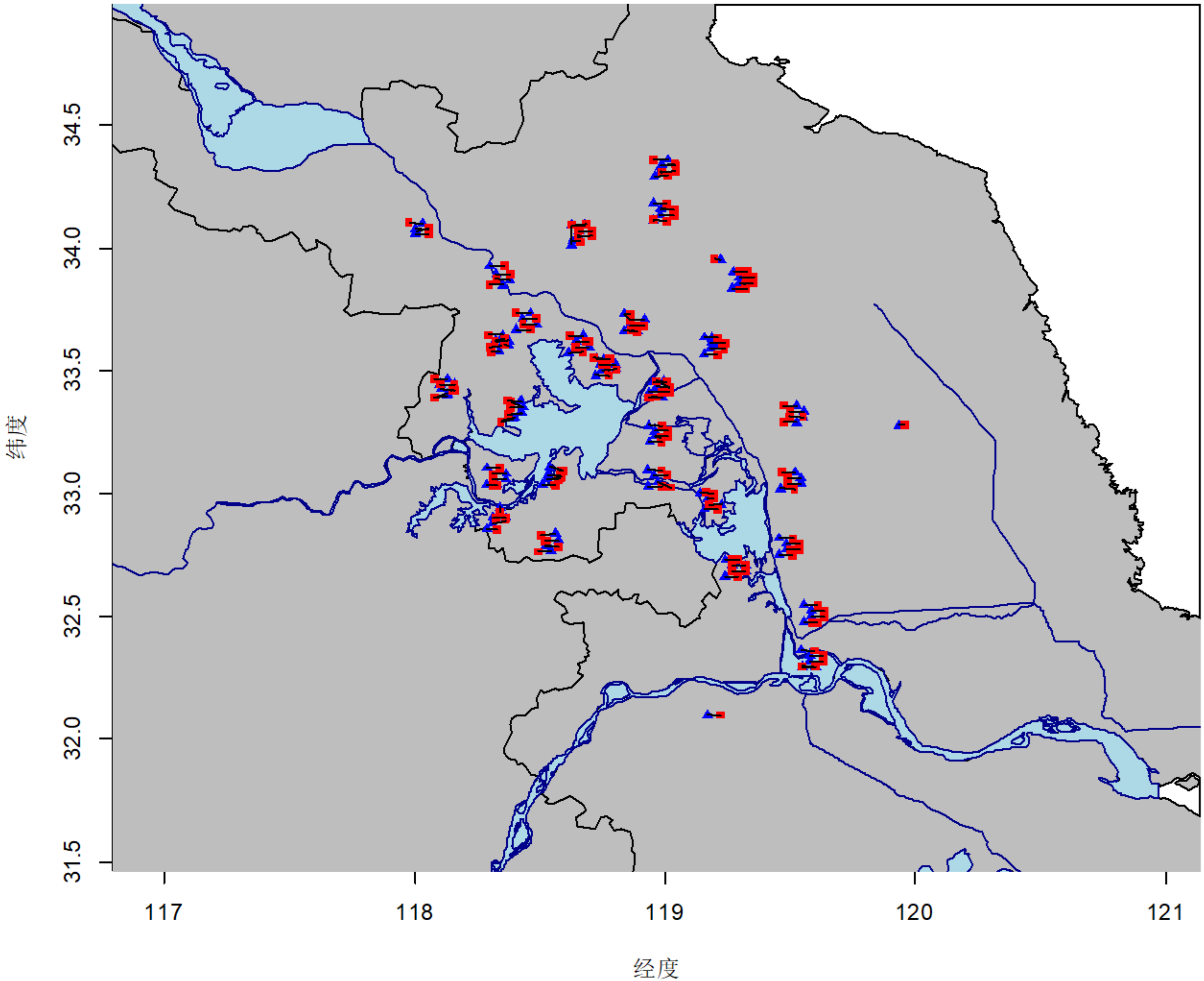
Confident intervals (95%) for estimating population size when the variance of animal count at each sample (survey route) is 90 (black), 30 (red), 10 (green) and 0.3 (blue) times as same as the mean, respectively

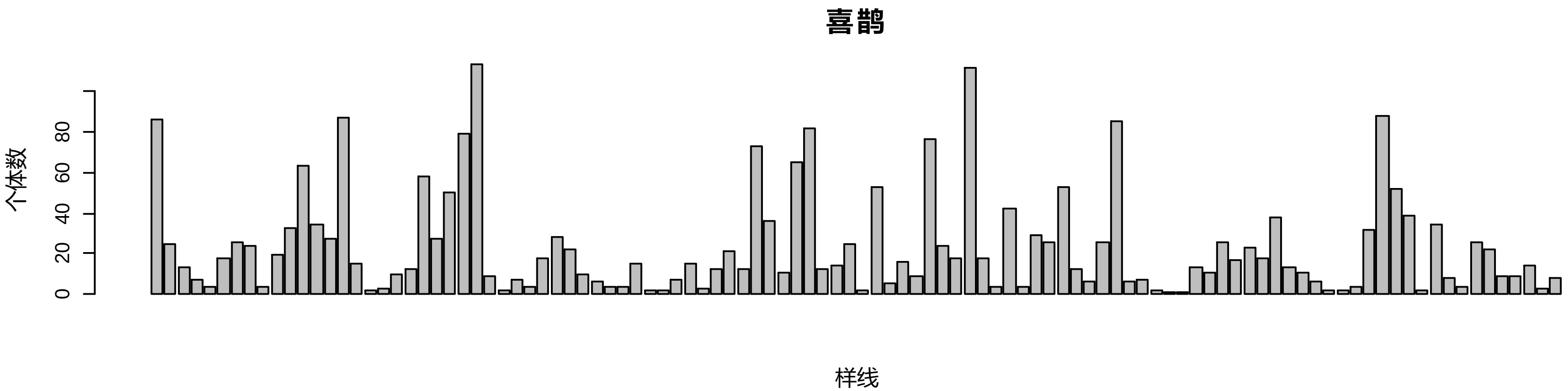


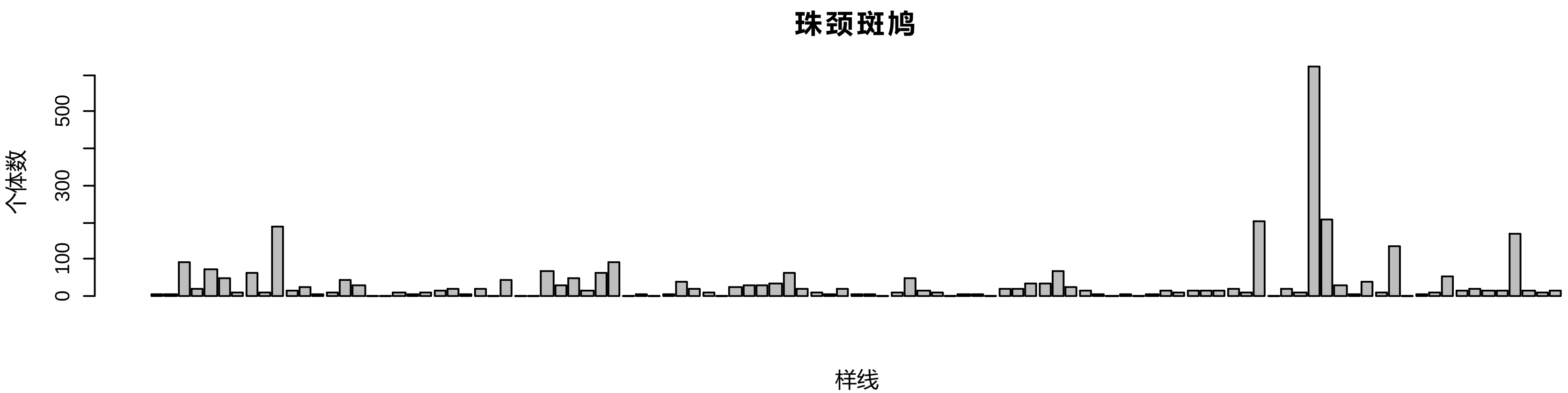


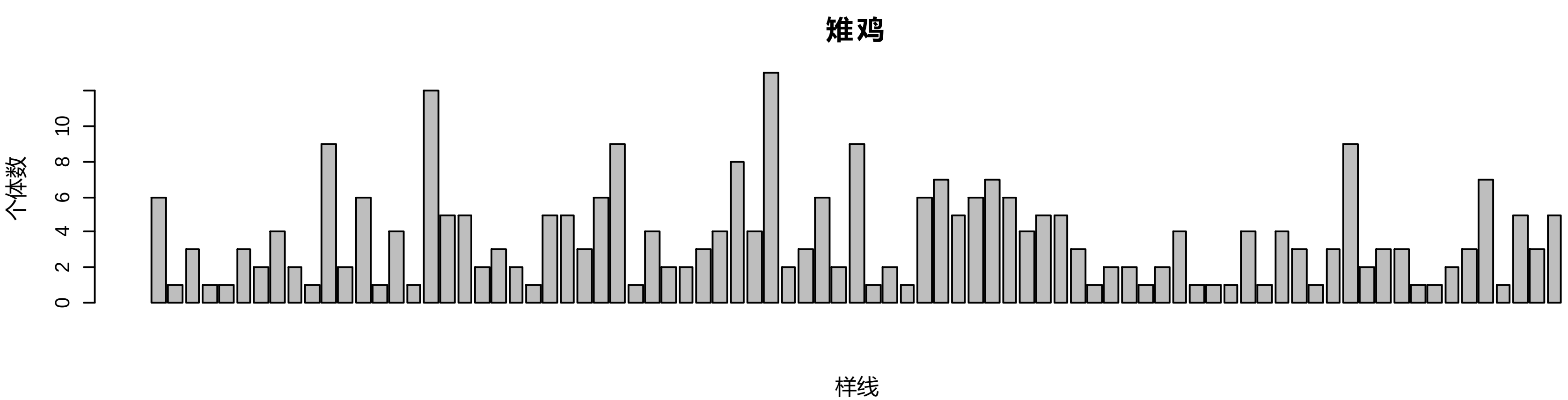
Variance-sample relationship for estimating population size
(The black line indicates the number of routes needed for 95% confident interval of the mean = 100±20%)
样线数 Number of sample (survey route)
10
20
30
40
0
2000
4000
6000
8000
方差 Variance

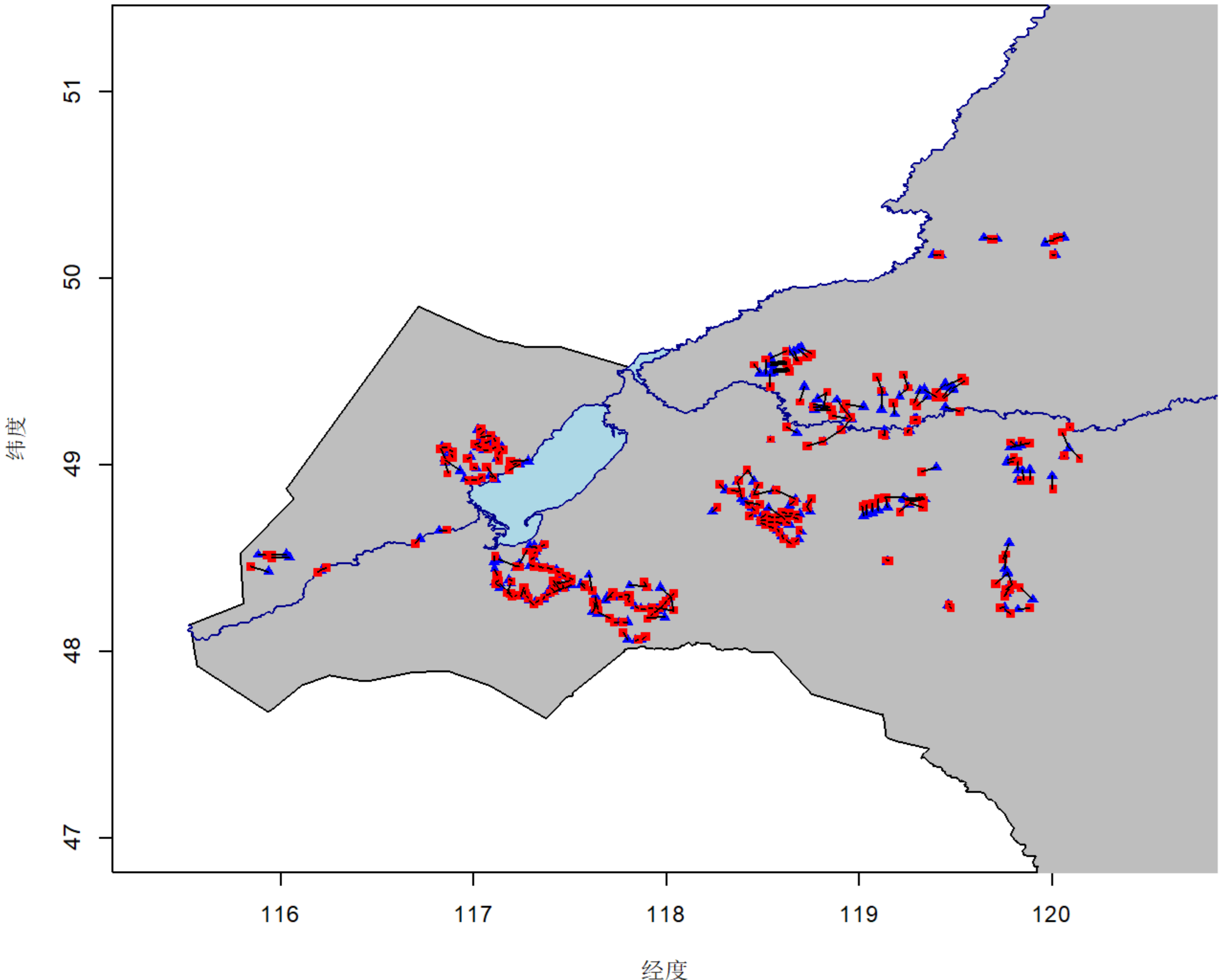
Actual survey results

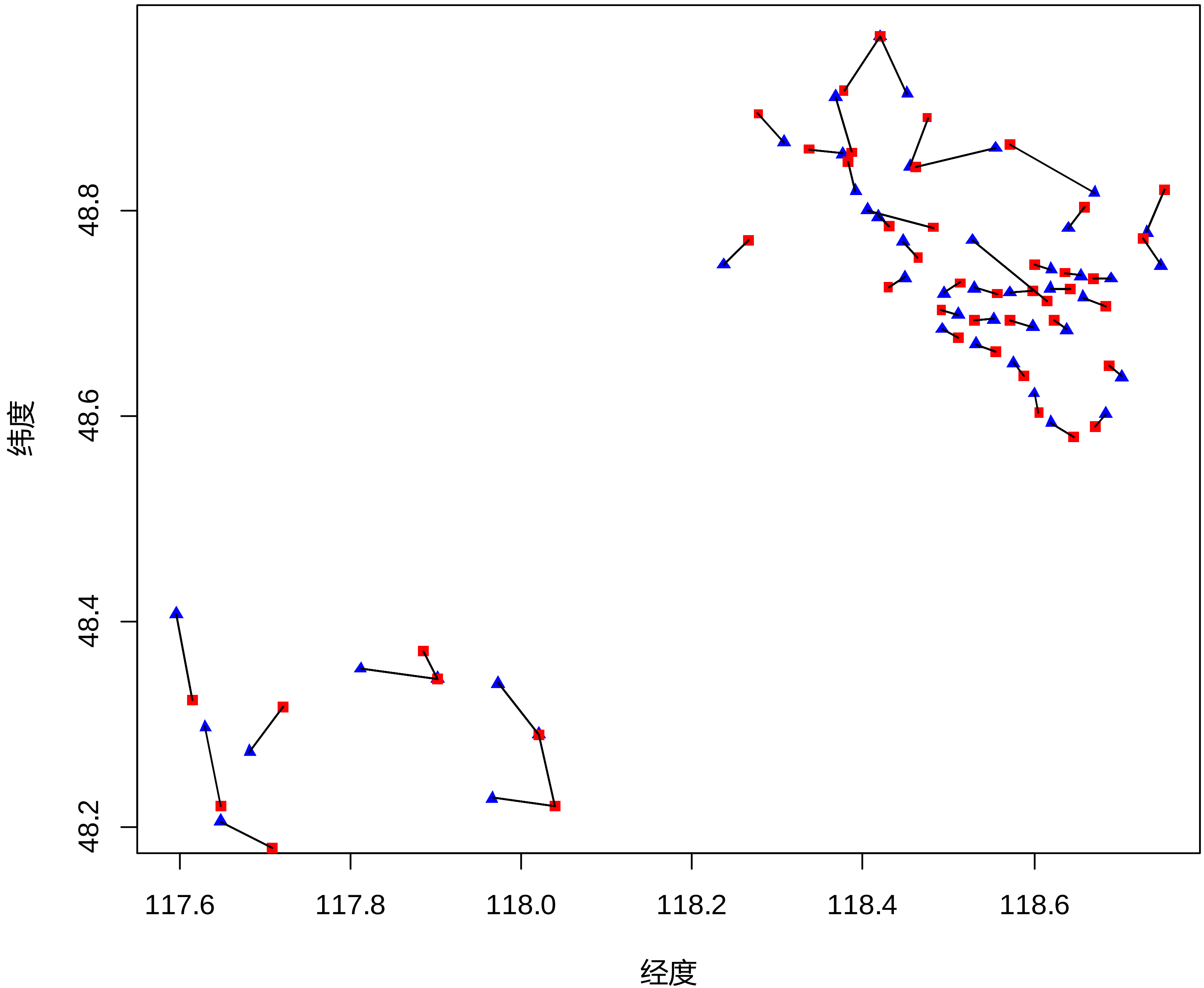

34.5
34.0
33.5
33.0
32.5
32.0
31.5
纬度
117
118
119
120
121
经度


喜鹊
个体数
0
20
40
60
80
样线

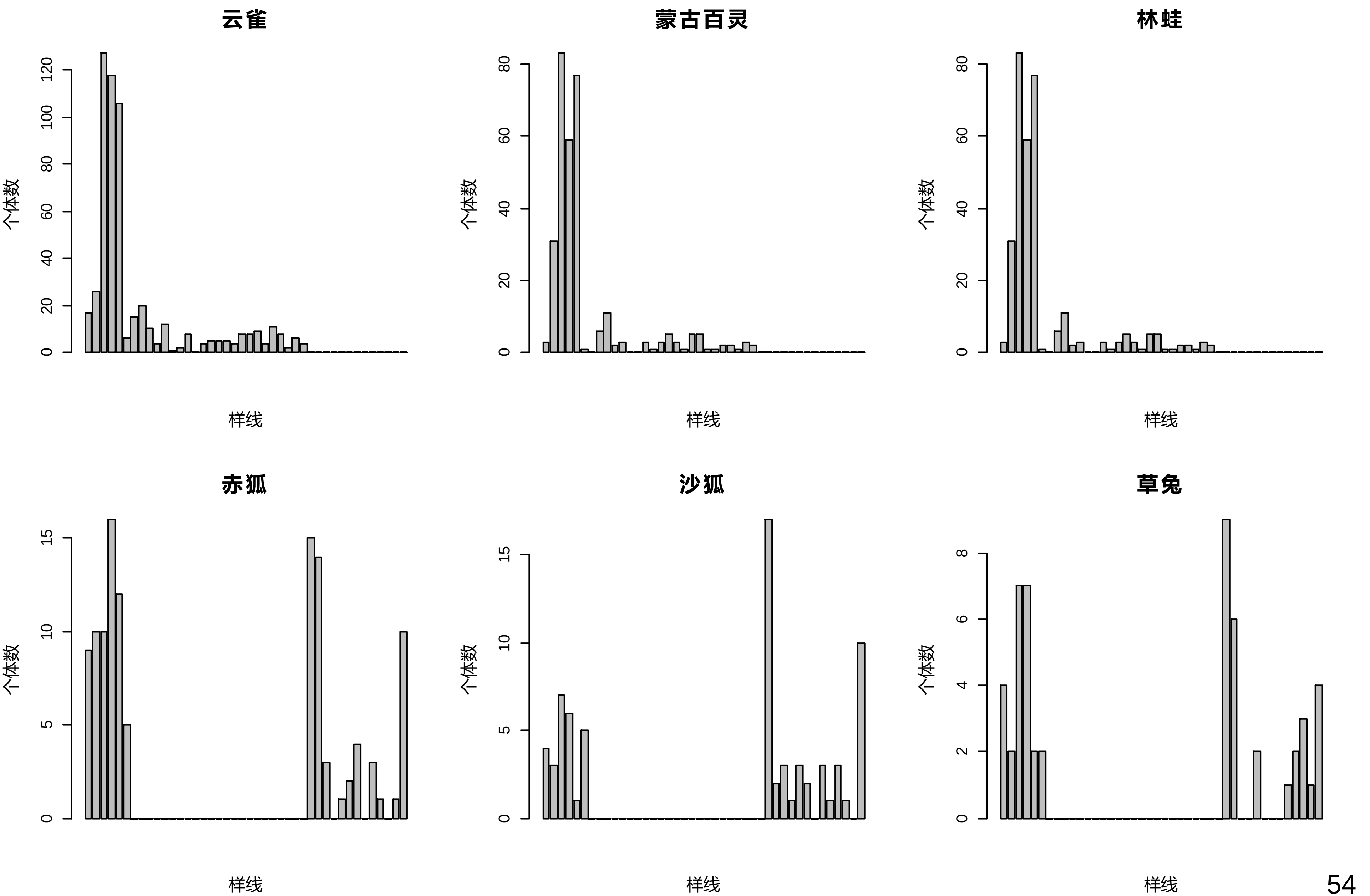

珠颈斑鸠
个体数
0
100
300
500
样线


雉鸡
个体数
0
2
4
6
8
10
样线


51
50
49
48
47
纬度
116
117
118
119
120
经度

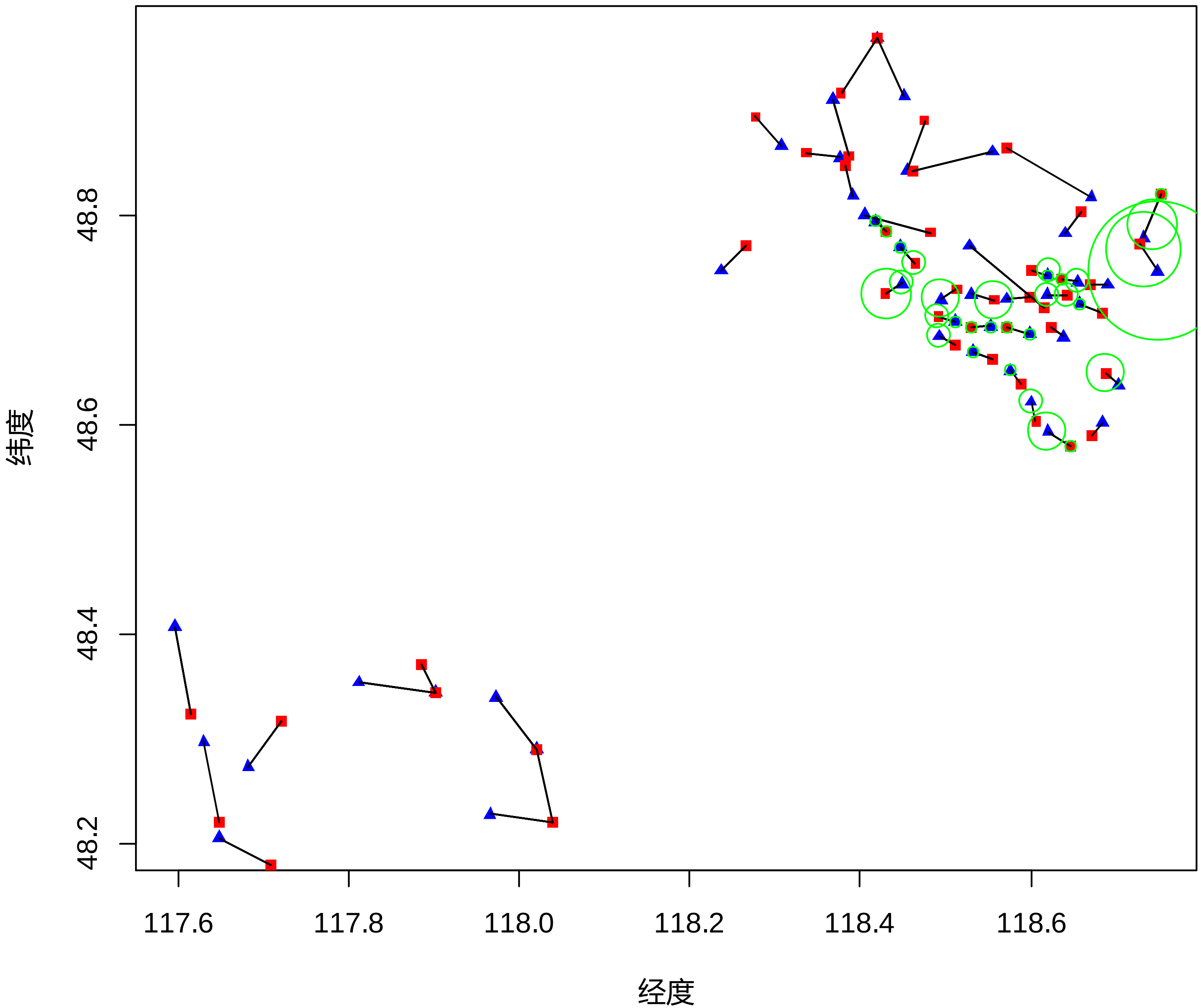

纬度
48.2
48.4
48.6
48.8
117.6
117.8
118.0
118.2
118.4
118.6
经度

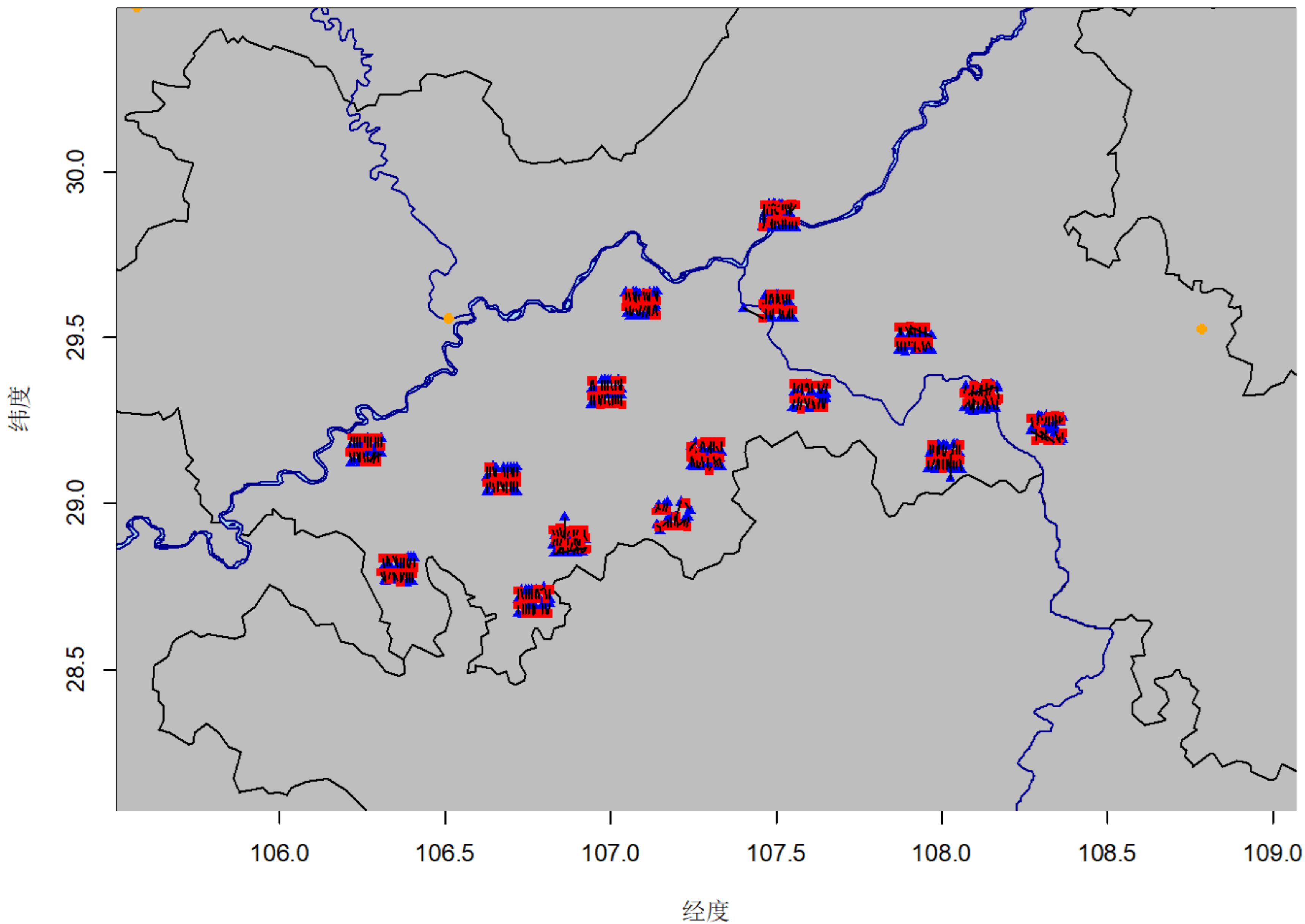
	样线	云雀	蒙古百灵	林蛙	赤狐	沙狐	草兔
1	3-06-079-201-101	17	3	3	9	4	4
2	3-06-079-201-102	26	31	31	10	3	2
3	3-06-079-201-103	127	83	83	10	7	7
4	3-06-079-201-104	118	59	59	16	6	7
5	3-06-079-201-105	106	77	77	12	1	2
6	3-06-079-201-106	6	1	1	5	5	2
7	3-06-079-201-107	15	0	0	0	0	0
8	3-06-079-201-108	20	6	6	0	0	0
9	3-06-079-201-109	10	11	11	0	0	0
10	3-06-079-201-110	4	2	2	0	0	0
11	3-06-079-201-111	12	3	3	0	0	0
12	3-06-079-201-112	1	0	0	0	0	0
13	3-06-079-201-113	2	0	0	0	0	0
14	3-06-079-201-114	8	3	3	0	0	0
15	3-06-079-201-115	0	1	1	0	0	0
16	3-06-079-201-116	4	3	3	0	0	0
17	3-06-079-201-117	5	5	5	0	0	0
18	3-06-079-201-118	5	3	3	0	0	0
19	3-06-079-201-119	5	1	1	0	0	0
20	3-06-079-201-120	4	5	5	0	0	0
21	3-06-079-201-121	8	5	5	0	0	0
22	3-06-079-201-122	8	1	1	0	0	0
23	3-06-079-201-123	9	1	1	0	0	0
24	3-06-079-201-124	4	2	2	0	0	0
25	3-06-079-201-125	11	2	2	0	0	0
26	3-06-079-201-126	8	1	1	0	0	0
27	3-06-079-201-128	2	3	3	0	0	0
28	3-06-079-201-129	6	2	2	0	0	0
29	3-06-079-201-130	4	0	0	0	0	0
30	3-06-079-201-201	0	0	0	15	17	9
31	3-06-079-202-101	0	0	0	14	2	6
32	3-06-079-202-201	0	0	0	3	3	0
33	3-06-079-202-202	0	0	0	0	1	0



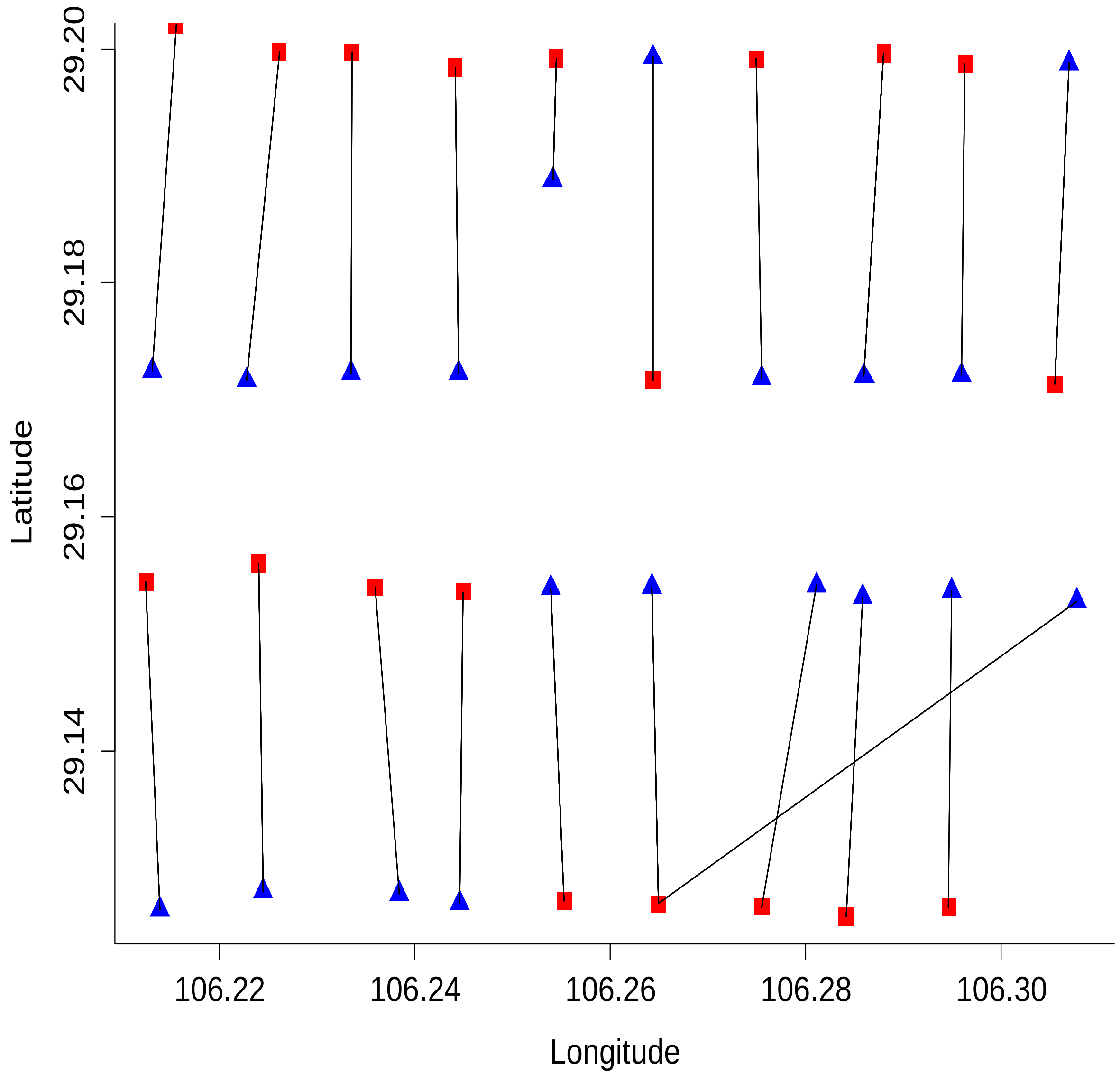
云雀
蒙古百灵
林蛙
赤狐
沙狐
草兔
个体数
样线

云雀

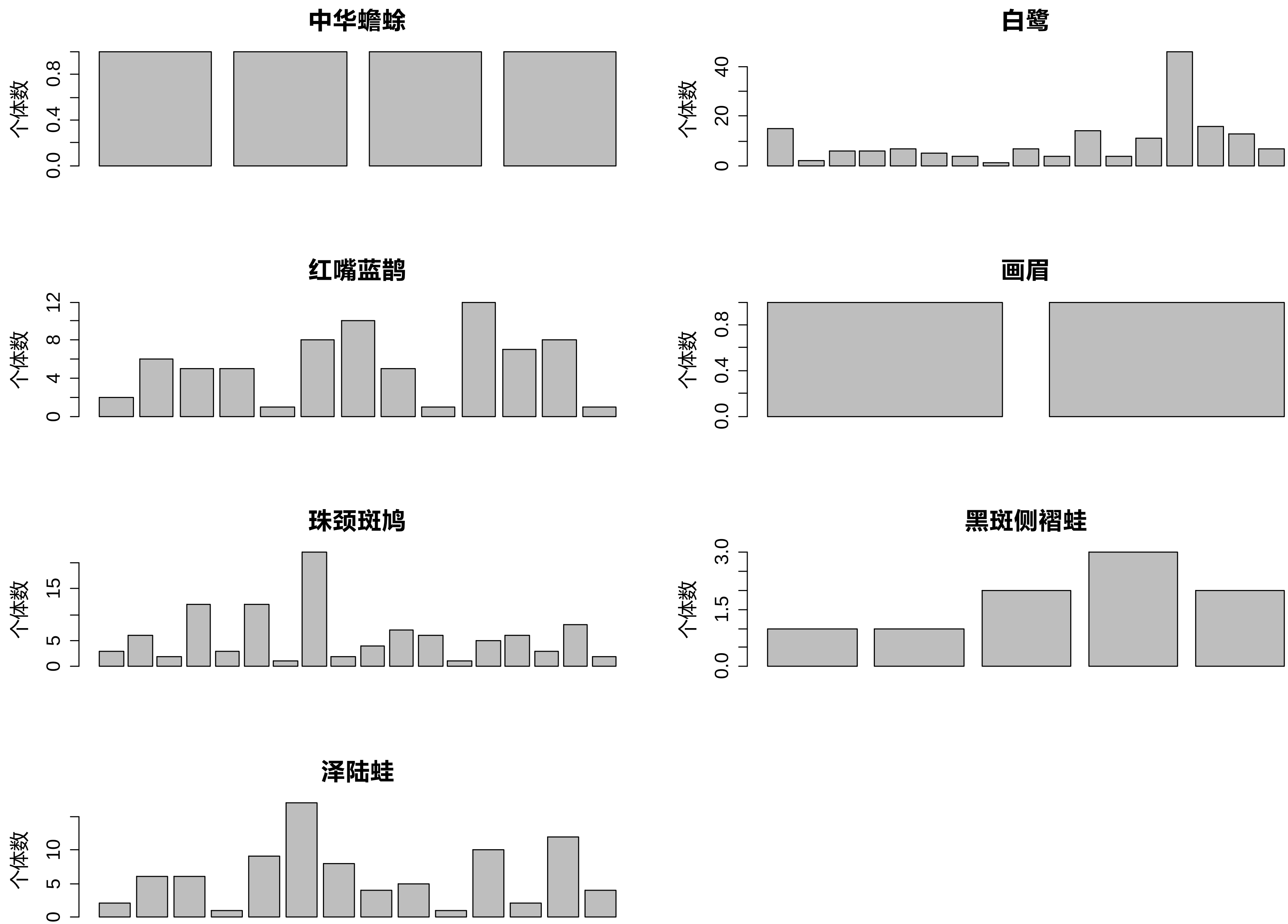



30.0
29.5
29.0
28.5
纬度
106.0
106.5
107.0
107.5
108.0
108.5
109.0
经度

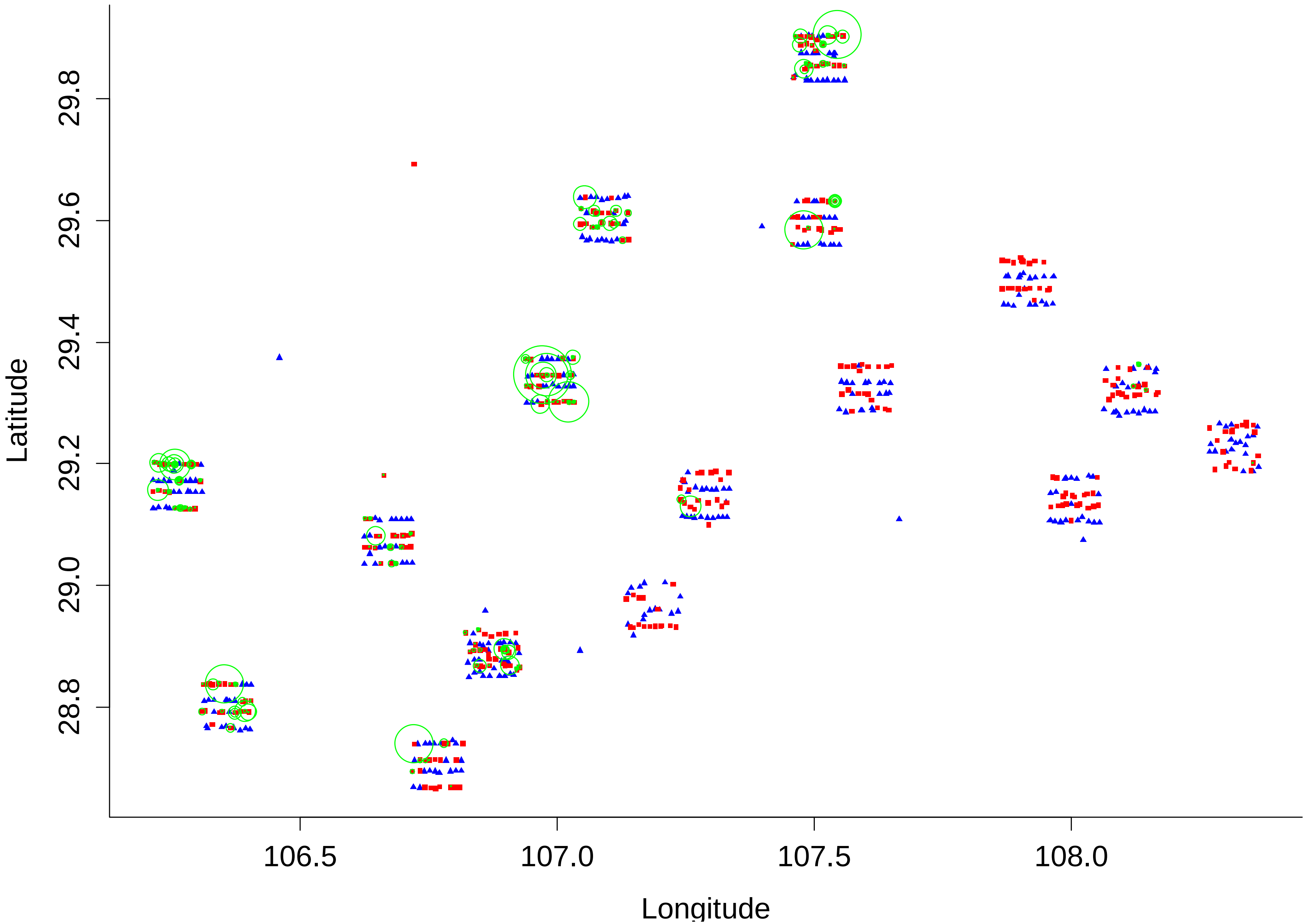
样区10的样线


29.20
29.18
29.16
29.14
Latitude
106.22
106.24
106.26
106.28
106.30
Longitude

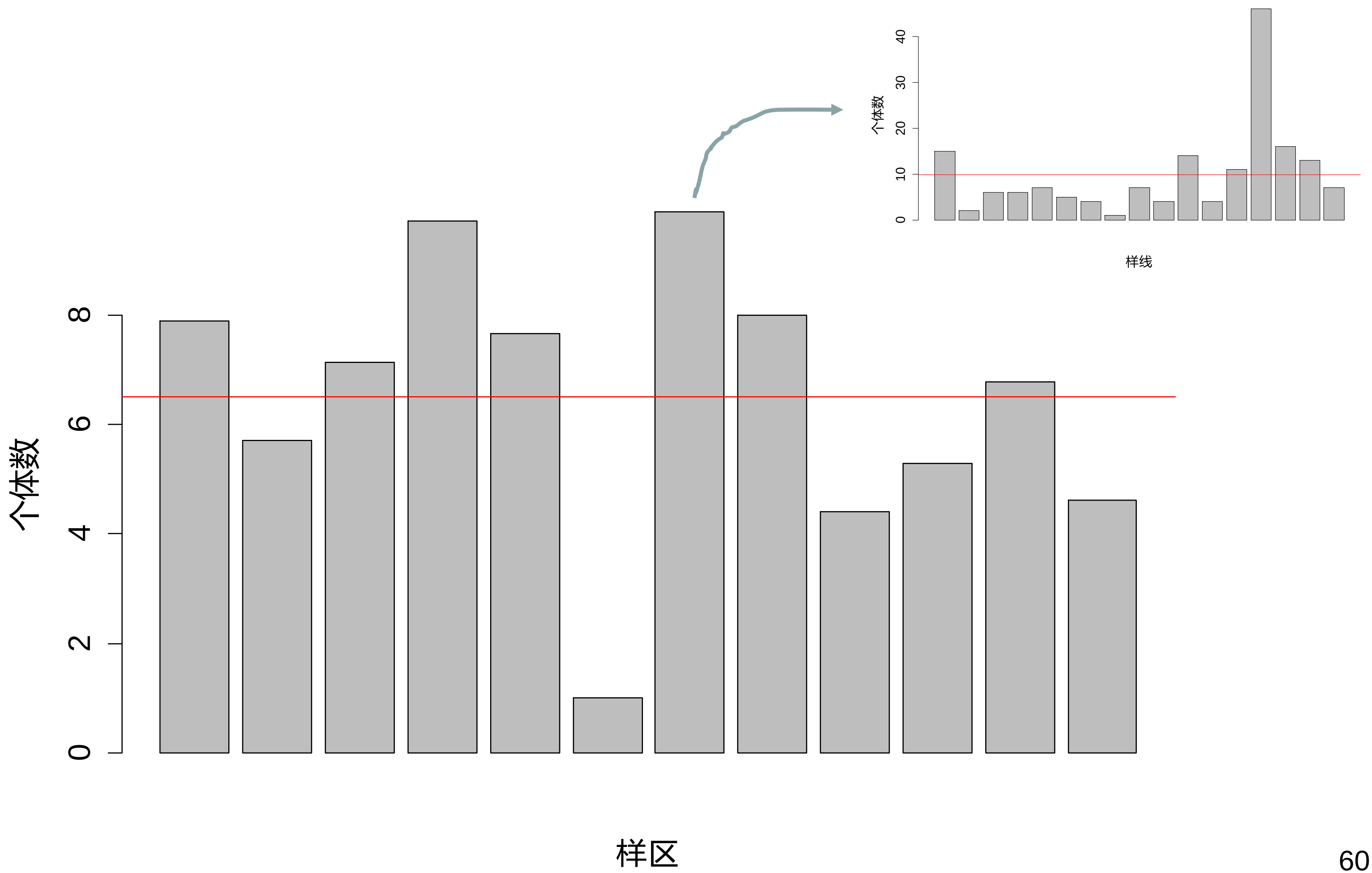
样区10各样线的物种数



小白鹭在各样区每条样线的数量



小白鹭在各样区每条样线的平均数量



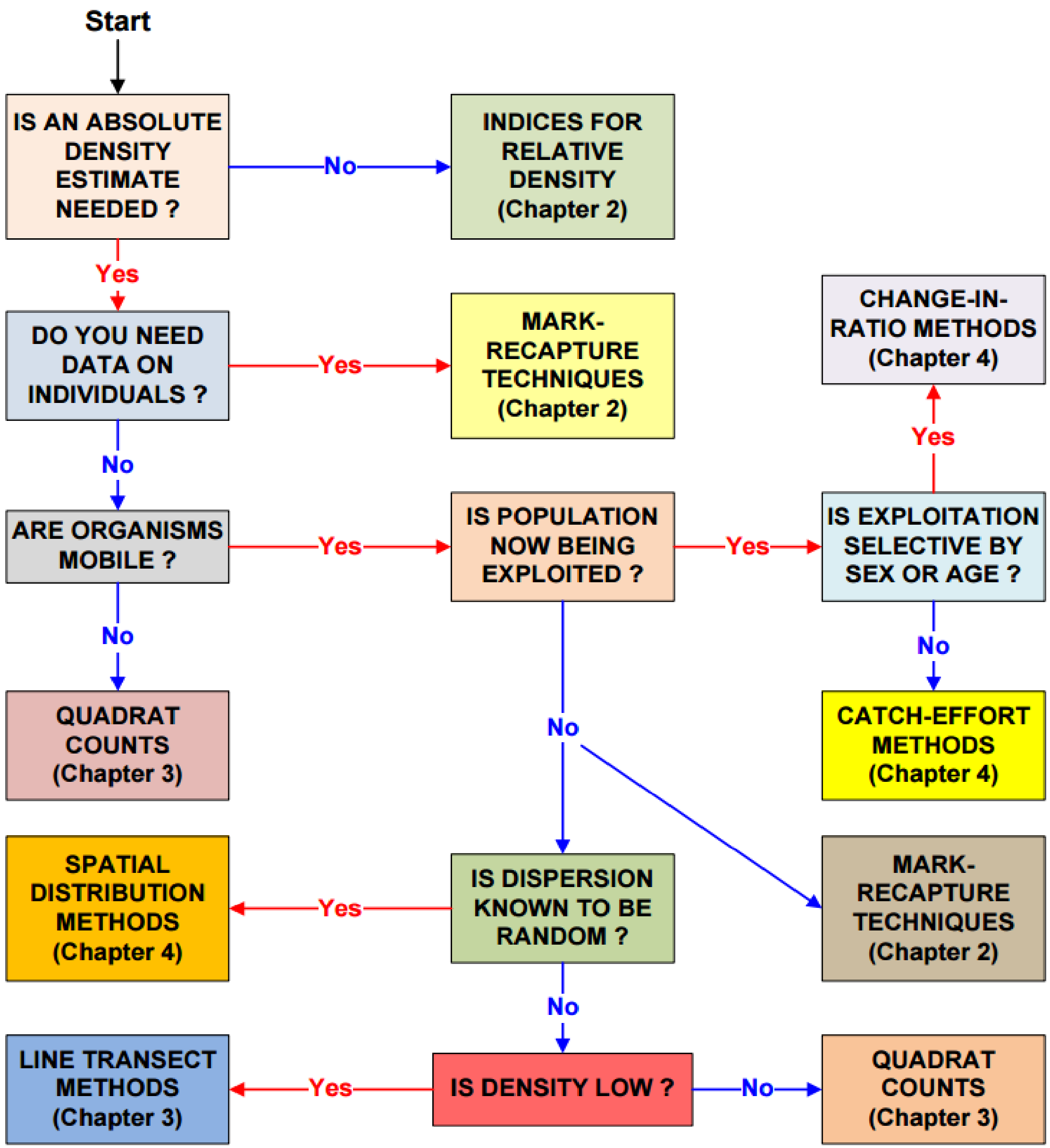
All models are wrong; some models are useful.

George E. P. Box, William Hunter and Stuart Hunter, *Statistics for Experimenters*, second edition, 2005, page 440.

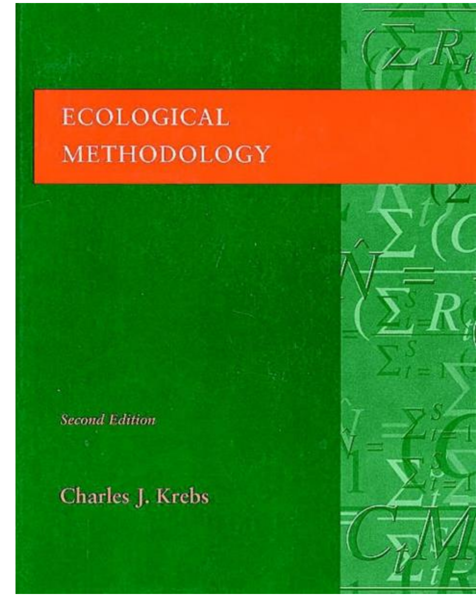
- George Edward Pelham Box (October 18, 1919 – March 28, 2013).
- A British mathematician and professor of statistics at the University of Wisconsin.
- A pioneer in the areas of quality control, time series analysis, design of experiments and Bayesian inference.
- He was the son-in-law of Sir Ronald Fisher.

http://cen.acs.org/content/cen/articles/91/i13/Design-Experiments-Makes-Comeback/_jcr_content/articlebody/subpar/boxed_text_0/par/media_0.img.jpg/1466564026093.jpg

Wildlife survey methods



Page 22. **Figure A**. Sequence of decisions by which a technique for estimating abundance can be chosen. (Modified from Caughley 1977.)



Krebs, Charles J. 2014. Ecological Methodology. Third edition.

abundanceR

install_github("Xinhai-Li/abundanceR")

中国科学: 生命科学 2019 年 第 49 卷 第 2 期: 151 ~ 162
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

论 文

用物种分布模型和距离抽样估计三江源藏野驴、藏原羚和藏羚羊的数量

李欣海[1,2*], 郜二虎[3], 李百度[4], 詹祥江[1,2*]

1. 中国科学院动物研究所动物生态与保护生物学院重点实验室, 北京 100101;
2. 中国科学院大学, 北京 100049;
3. 国家林业和草原局调查规划设计院, 北京 100714;
4. 中国人民大学附属中学, 北京 100030
* 联系人, E-mail: lixh@ioz.ac.cn; zhanxj@ioz.ac.cn

收稿日期: 2018-08-22; 接受日期: 2018-10-30; 网络版发表日期: 2019-01-14
中国科学院科技服务网络(STS)计划重点项目(批准号: KFJ-STS-ZDTP-013)、UNDP-GEF青海三江源生物多样性保护项目和国家自然科学基金面上项目(批准号: 31772479, 31572287)资助

摘要 三江源国家公园成立后, 人们需要了解该区域关键有蹄类物种如藏野驴(*Equus kiang*)、藏原羚(*Procapra picticaudata*)和藏羚羊(*Pantholops hodgsonii*)等物种的分布和数量, 以便制定相应的保护规划. 我们于2014~2017年夏季在三江源53.8×10^4 km^{2}的范围内进行了样线调查, 考察行程总计14597.8 km, 记录到藏野驴3711头, 藏原羚1187只, 藏羚羊423只. 为了准确估计有蹄类的数量, 我们用随机森林模型量化了物种分布与22个环境变量的关系, 预测了三大有蹄类在整个区域的分布和数量, 并通过样线调查的数据进行校正, 得到藏野驴、藏原羚和藏羚羊在三江源研究区域的总数分别为44240头、13162只和2390只. 四年来三大有蹄类数量稳定. 我们应用距离抽样的探测函数、随机森林模型中环境变量对物种数量的解释程度以及调查结果和模型结果的匹配程度进行不确定性分析, 计算了动物估计数量的置信区间. 我们建立了新的动物数量估计方法, 适合于动物分布与环境变量关系密切并有样线调查结果的情况.

关键词 三江源国家公园, 野生动物数量估计, 物种分布模型, 距离抽样, 有蹄类

三江源地区具有丰富而独特的物种资源和重要的生态系统服务功能[1–3], 然而人们对其关键物种的分布和数量了解得还很不充分. 三江源国家公园成立后, 人们迫切需要了解关键物种的分布和数量, 以便制定国家公园的详细规划, 达到长期保护的目的. 藏野驴(*Equus kiang*)、藏原羚(*Procapra picticaudata*)和藏羚羊(*Pantholops hodgsonii*)是青藏高原的特有物种, 也是三江源的优势物种和关键物种. 为了掌握这些物种的分布、数量和动态, 我们在2014~2017年进行了调查并估计了藏野驴、藏原羚和藏羚羊的分布和数量. 这些数据将为三江源国家公园的建设和生物多样性保护提供决策支持.

引用格式: 李欣海, 郜二虎, 李百度, 等. 用物种分布模型和距离抽样估计三江源藏野驴、藏原羚和藏羚羊的数量. 中国科学: 生命科学, 2019, 49: 151–162
Li X H, Gao E H, Li B D, et al. Estimating abundance of Tibetan wild ass, Tibetan gazelle and Tibetan antelope using species distribution models and distance sampling (in Chinese). Sci Sin Vitae, 2019, 49: 151–162, doi: 10.1360/N052018-00171

李欣海，郜二虎，李百度和詹祥江。2019。用物种分布模型和距离抽样估计三江源藏野驴、藏原羚和藏羚羊的数量。中国科学: 生命科学，49(2): 151-162。

land MDPI

Article

AbundanceR: A Novel Method for Estimating Wildlife Abundance Based on Distance Sampling and Species Distribution Models

Xinhai Li [1,2,*,†], Ning Li [3,†], Baidu Li [4], Yuehua Sun [1] and Erhu Gao [5]

1 Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; sunyh@ioz.ac.cn
2 College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
3 Institute of Applied Ecology, Nanjing Xiaozhuang University, Nanjing 211171, China; lining@njxzc.edu.cn
4 Lassonde School of Engineering, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada; baidu@my.yorku.ca
5 Academy of Inventory and Planning, National Forestry and Grassland Administration, Beijing 100714, China; gaoerhu2022@126.com
* Correspondence: lixh@ioz.ac.cn; Tel.: +86-10-64807898
† These authors contributed equally to this work.

Abstract: Appropriate field survey methods and robust modeling approaches play an important role in wildlife protection and habitat management because reliable information on wildlife distribution and abundance is important for conservation planning and actions. However, accurately estimating animal abundance is challenging in most species, as usually only a small proportion of the population can be detected during surveys. Species distribution models can predict the habitat suitability index, which differs from species abundance. We designed a method to adjust the results from species distribution models to achieve better accuracy for abundance estimation. This method comprises four steps: (1) conducting distance sampling, recording species occurrences, and surveying routes; (2) performing species distribution modeling using occurrence records and predicting animal abundance in each quadrant in the study area; (3) comparing the difference between field survey results and predicted abundance in quadrats along survey routes, adjusting model prediction, and summing up to obtain total abundance in the study area; (4) calculating uncertainty from three sources, i.e., distance sampling (using detection rate), species distribution models (using R squared), and differences between the field survey and model prediction [using the standard deviation of the ratio (observation/prediction) at different zones]. We developed an R package called abundanceR to estimate wildlife abundance and provided data for the Tibetan wild ass (*Equus kiang*) based on field surveys at the Three-River-Source National Park, as well as 29 layers of environmental variables covering the terrestrial areas of the planet. Our method can provide accurate estimation of abundance for animals inhabiting open areas that can be easily observed during distance sampling, and whose spatial heterogeneity of animal density within the study area can be accurately predicted using species distribution models.

Keywords: abundance; distance sampling; population density; R package; species distribution models; wildlife survey; uncertainty

check for updates

Citation: Li, X.; Li, N.; Li, B.; Sun, Y.; Gao, E. AbundanceR: A Novel Method for Estimating Wildlife Abundance Based on Distance Sampling and Species Distribution Models. *Land* **2022**, *11*, 660. https://doi.org/10.3390/land11050660

Academic Editors: Juan F. Beltrán, Pedro Abellán, John Litvaitis and Francesco Maria Angelici

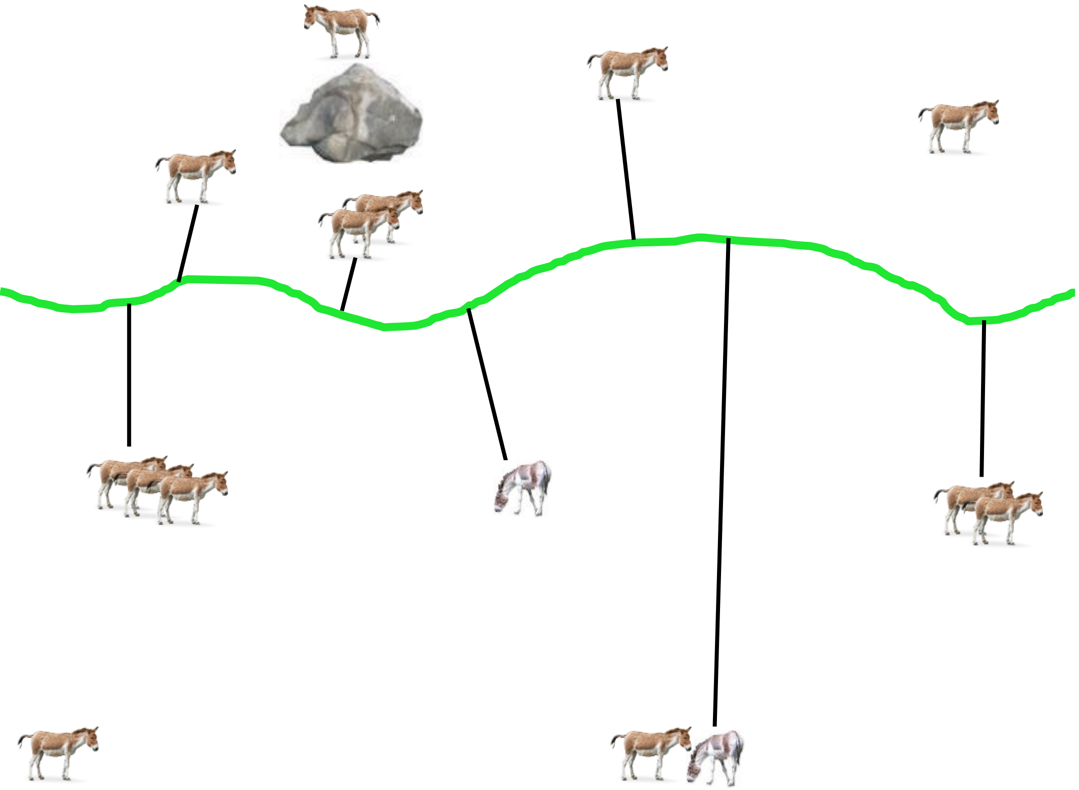
Received: 10 February 2022
Accepted: 27 April 2022
Published: 29 April 2022

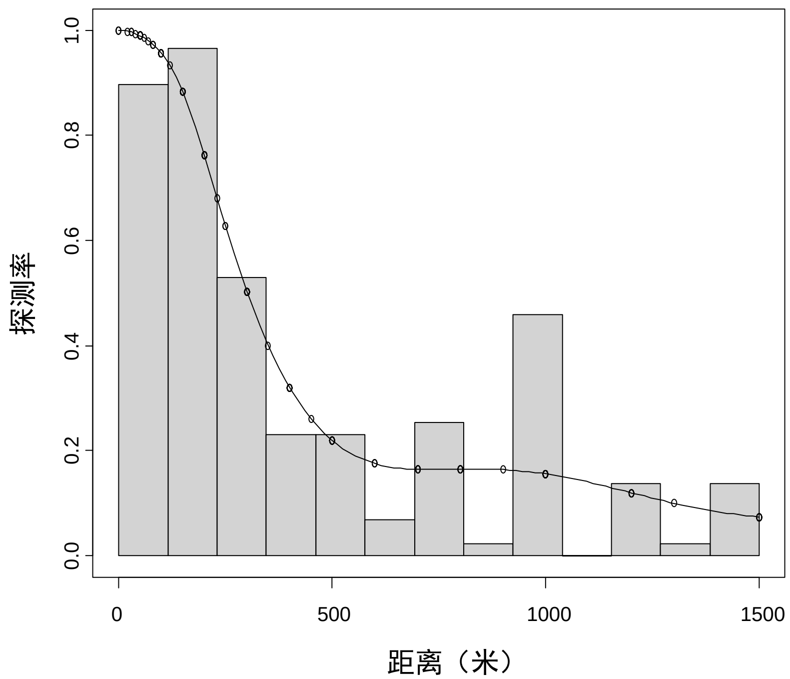
1. Introduction

Species abundance is fundamental information for ecological research, and various methods are used for its estimation [1–3]. Direct counting, including spotlight, track, and roadkill counts, can provide initial information on the abundance of several species, such as water birds and deer [4,5]. Catch-per-unit effort is often used for regularly captured species, such as commercial fish [6,7]. The mark-recapture technique has been widely used and it is suitable for populations in a small region when target species can be easily captured [8,9].

Li X, Li N, Li B, Sun Y, Gao E. 2022. abundanceR: A novel method for estimating wildlife abundance based on distance sampling and species distribution models. Land 11:660.

Distance sampling





R code for distance sampling

```
library(devtools)
install_github("Xinhai-Li/abundanceR", force = TRUE)
library(abundanceR)

kiang = kiang[kiang$distance<=500, ];
mean(kiang$size); sd(kiang$size)
sum(kiang$size) # 449
length(kiang$size) #103

# Distance sampling - fit detection function
library(Distance)
ds.fit <- ds(kiang, convert_units = 0.001)
# Convert.units adjusts for distance measured in metres and effort (transect length) in km
ds.kiang <- ds(kiang, key ="hn", adjustment ="cos",  convert_units = 0.001, truncation=1000)
```

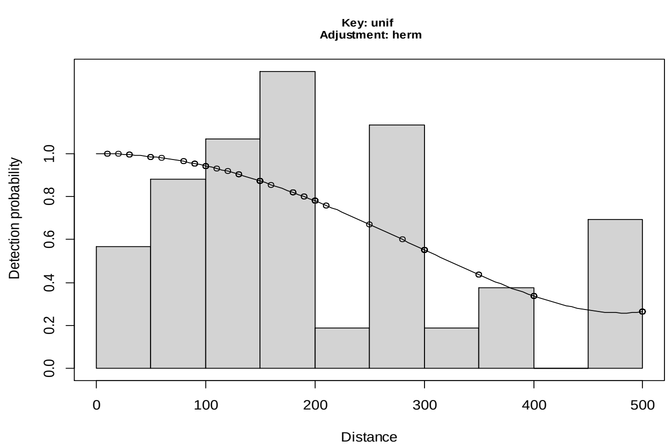
Optimizing parameters for distance sampling

```
set.seed(1)
AICs = distanceSampling(kiang[kiang$distance<=500, ]) # Take 0.1 - 5 mins

AICs = AICs[!is.na(AICs$AIC),] # sometimes no AIC value can be calculated
ds.kiang <- ds(kiang, key = AICs$Key[1], adjustment = AICs$Adjustment[1],
               convert_units = 0.001, truncation=500)

summary(ds.kiang)
SM = summary(ds.kiang)
```

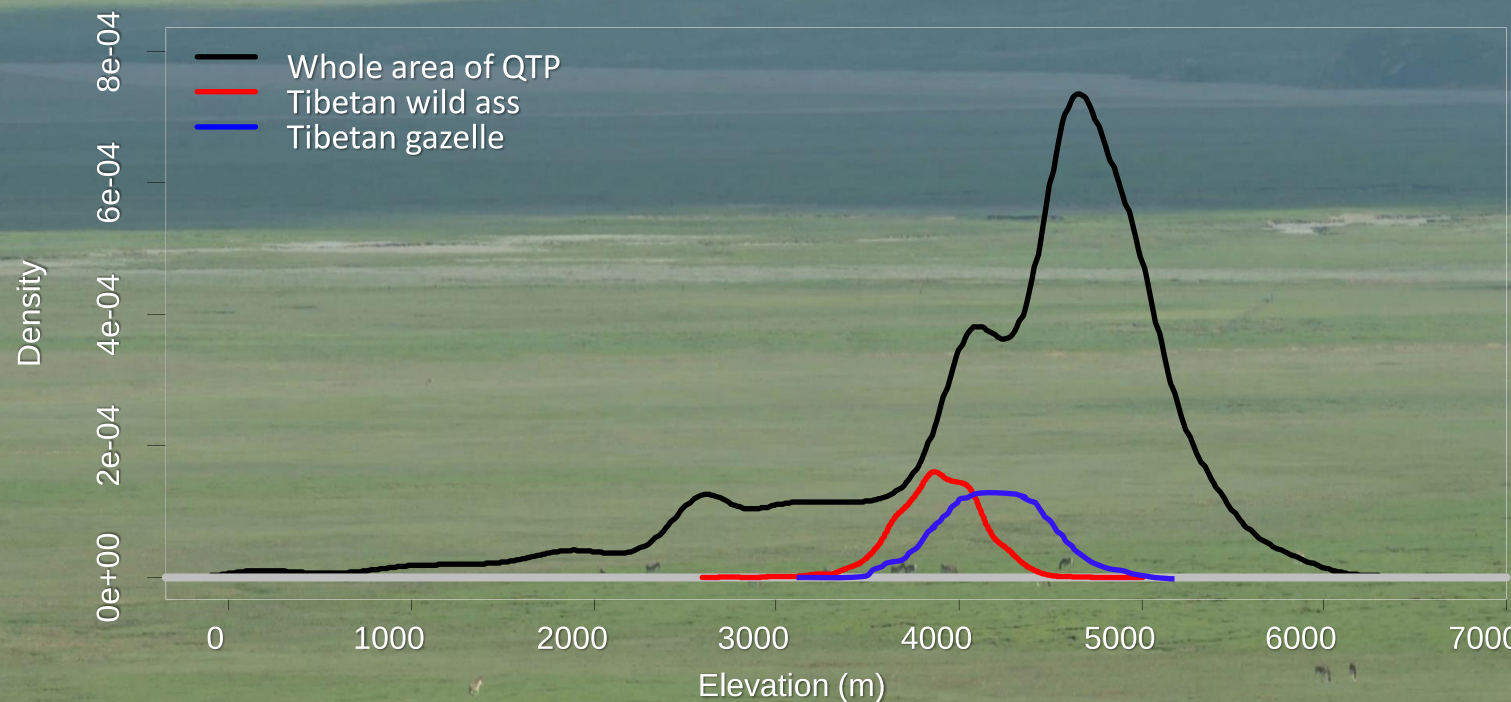
	Estimate	SE	CV
Average p	0.75	0.08	0.10
N in covered region	137.98	15.61	0.11

ID	Key	Adjustment	AIC
5	Unif	Herm	1261.143
6	Unif	Poly	1276.473
4	Unif	Cos	1276.605
1	Hn	Cos	1278.443
2	Hn	Herm	1278.443
3	Hn	Poly	1278.443
7	Hr	Cos	1278.659
8	Hr	Herm	1278.659
9	Hr	Poly	1278.659

```
Average.p = SM$ds$average.p; Average.p # average detection rate across the distance range
survey.uncertainty = 1 - Average.p # survey uncertainty
plot(ds.kiang, main=paste("Key:", AICs$Key[1], "\n", "Adjustment:", AICs$Adjustment[1], sep=" "))
```



Species-environment relationship



Environmental variables

```
library(raster)
BioClim <- brick('var29.grd')
BioClim <- brick('SJY.grd')

names(BioClim) = c("bio_1","bio_2","bio_3","bio_4","bio_5","bio_6","bio_7","bio_8","bio_9","bio_10","bio_11","bio_12","bio_13","bio_14","bio_15","bio_16","bio_17","bio_18","bio_19","elev","solar1","solar7","vapor1","vapor7","wind1","wind7","footprint", "wetland", "landcover")
```

数据引用：

气候、辐射、风速、水汽压 (Hijmans et al. 2005)
海拔 (Rabus et al. 2003)
人类足迹指数 (Sanderson et al. 2002)
土地利用 (Belward et al. 1999)
湿地 (Lehner et al. 2004)

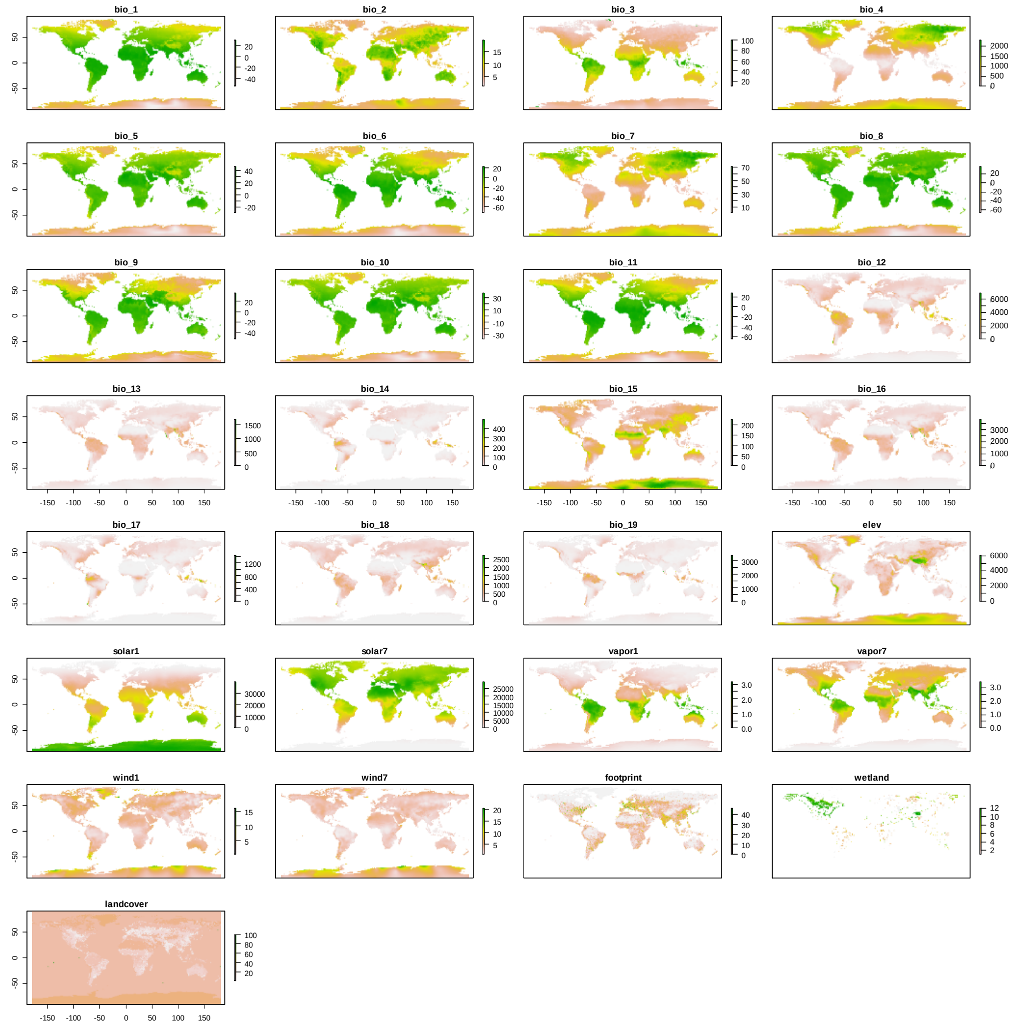
Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25, 1965-1978.

Rabus, B., M. Eineder, A. Roth, and R. Bamler. (2003) The shuttle radar topography mission - a new class of digital elevation models acquired by spaceborne radar. Isprs Journal of Photogrammetry and Remote Sensing, 57, 241-262.

Sanderson, E. W., M. Jaiteh, M. A. Levy, K. H. Redford, A. V. Wannebo, and G. Woolmer. (2002) The Human Footprint and the Last of the Wild. Bioscience, 52, 891-904.

Belward, A.S., Estes, J.E., and Kline, K.D., (1999) The IGBP-DIS 1-Km Land-Cover Data Set DISCover: A Project Overview: Photogrammetric Engineering and Remote Sensing , v. 65, no. 9, p. 1013-1020.

Lehner, B. and Döll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296/1-4: 1-22.



全球29个环境变量的数据，有105G，压缩后有8G，在百度云上分享。链接：https://pan.baidu.com/s/1noU8A7WcsuYx0MSiQq6CeQ 提取码：1234

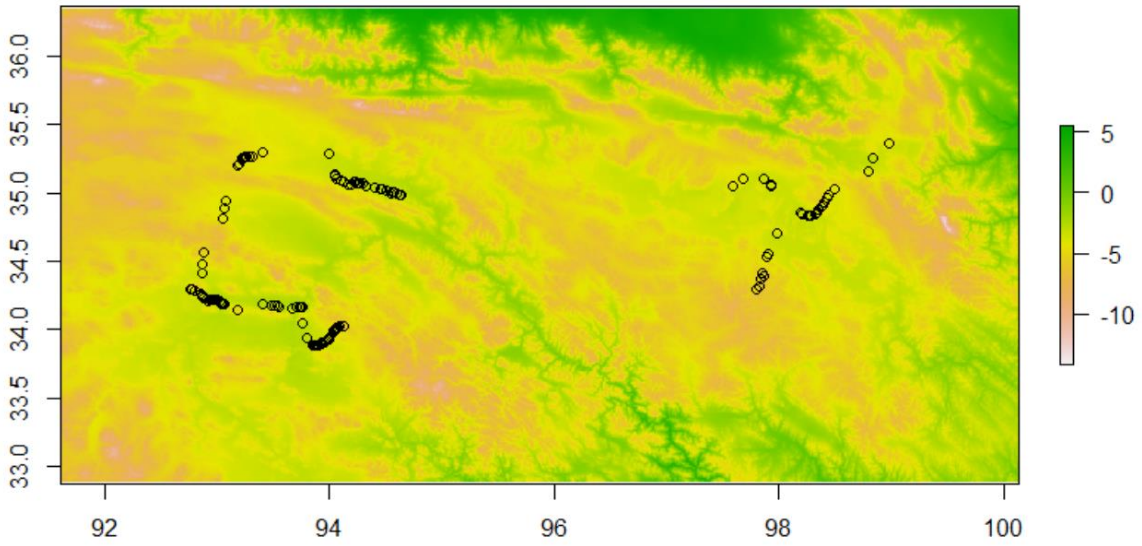
Occurrence data

data(kiang); head(kiang)

species	**size**	**distance**	Side	**Lat**	**Lon**	Elev	Date	Time
kiang	9	130	e	34.83078	98.37612	4217.56	2017/7/17	13:28:43
kiang	32	150	e	34.8462	98.4416	4223.56	2017/7/17	13:22:53
kiang	7	600	e	34.8508	98.2975	4225.55	2017/7/17	13:37:12
kiang	8	350	e	34.85908	98.45139	4232.56	2017/7/17	13:18:01
kiang	1	210	e	34.87584	98.47288	4236.59	2017/7/17	13:15:24
kiang	3	200	e	34.89577	98.49666	4244.47	2017/7/17	13:12:01

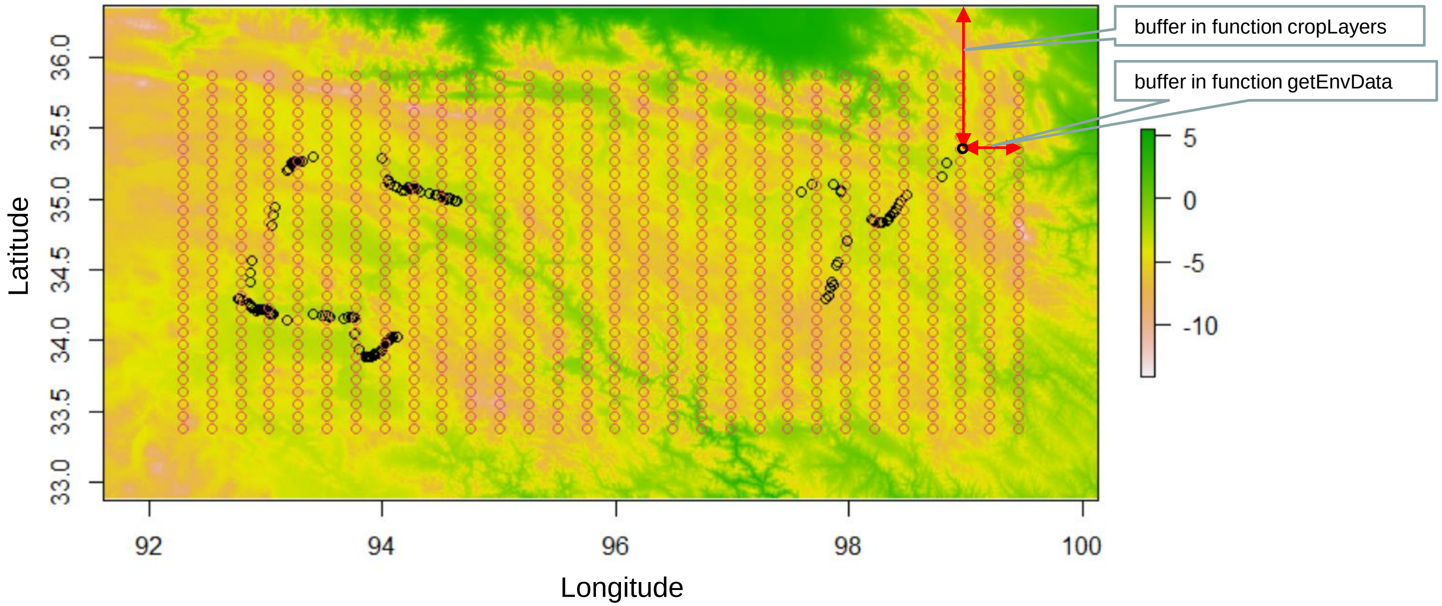
Crop environmental layers

```
data(kiang)
BioClim = cropLayers(kiang, buffer = 1.5, Envlayers = BioClim)
plot(BioClim[[1]])
points(kiang$Lon, kiang$Lat)
```



Derive data for species distribution modeling

```
# This buffer should be smaller than that in cropLayers
Data = getEnvData(kiang, buffer = 0.5, absence = 30, Envlayers = BioClim)
```



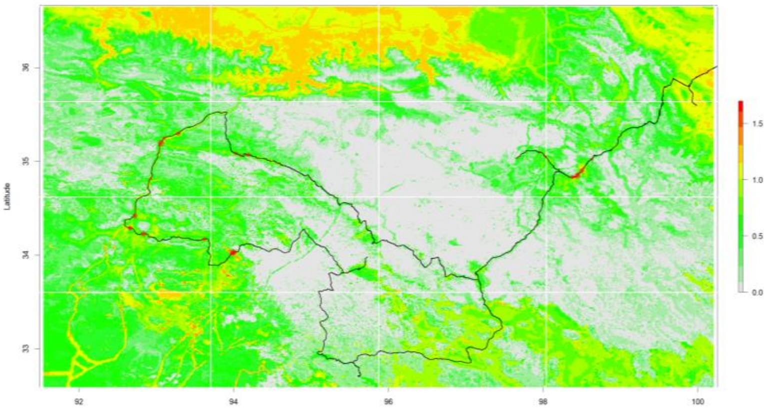
Name	Count	bio_1	bio_2	bio_3	bio_4	bio_5	bio_6	bio_7	bio_8	bio_9	bio_10	bio_11	bio_12	bio_13	bio_14	bio_15	bio_16	bio_17	bio_18	bio_19	elev	solar1	solar7	vapor1	vapor7	wind1	wind7	footprint	Lon	Lat
Kiang	3	-3.9	13.4	35.1	898.0	13.2	-25	38.2	6.57	-14.52	6.57	-15.35	367	84	3	94.58	220	11	220	12	4232	9490	19972	0.08	0.68	3.4	2.3	9.9	98.97605	35.25475
Kiang	1	-4.0	13.2	34.5	900.9	13.2	-25	38.2	6.53	-14.60	6.53	-15.42	364	82	3	94.03	217	11	217	12	4212	9630	20115	0.08	0.68	3.2	2.8	8.7	98.93184	35.15825
Kiang	5	-4.0	13.2	34.5	904.0	13	-25.3	38.3	6.45	-14.77	6.45	-15.57	351	79	4	93.35	208	12	208	13	4255	9655	20097	0.09	0.67	3.3	2.8	1.3	98.52607	34.92995
Kiang	1	-3.9	13.3	34.6	904.6	13	-25.3	38.3	6.50	-14.72	6.50	-15.52	350	79	3	93.36	207	11	207	12	4240	9664	20095	0.09	0.67	3.3	2.7	1.3	98.50558	34.90782
Kiang	17	-4.0	13.3	34.7	902.8	13	-25.3	38.3	6.42	-14.73	6.42	-15.53	352	78	3	92.96	208	11	208	12	4254	9713	20155	0.09	0.67	3.1	2.8	1.3	98.49941	34.89912
Kiang	3	-4.0	13.3	34.7	902.8	13	-25.3	38.3	6.42	-14.73	6.42	-15.53	352	78	3	92.96	208	11	208	12	4254	9713	20155	0.09	0.67	3.1	2.8	1.3	98.49666	34.89577

The species distribution model: random forest

```
library(randomForest)
# fill null values
no.col = ncol(Data)
Dat.fill <- na.roughfix(Data[,2:(no.col-4)]) # use 27 variables
# no.col-2: using 29 variables including landcover and wetland

set.seed(2)
RF <- randomForest(Dat.fill[, 2:ncol(Dat.fill)], Dat.fill[,1], ntree = 500,
                    importance = TRUE, na.action = na.roughfix)
RF
max(RF$rsq)
model.uncertainty = 1-max(RF$rsq)
```

Estimate animal abundance

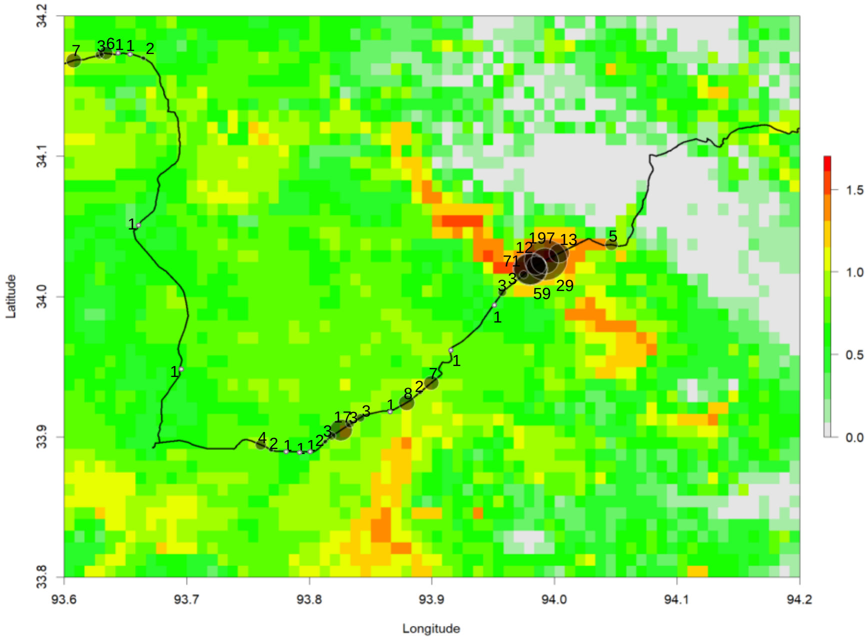
```
pred = popSize(BioClim, RF)
plot(pred)
cellStats(pred, stat='mean', na.rm=TRUE);
cellStats(pred, stat='range', na.rm=TRUE)
cellStats(pred, stat='sd', na.rm=TRUE)
cellStats(pred, stat='sum', na.rm=TRUE)
```



```
# Show the predicted animal density
species = kiang
grid=4
buffer = (range(species$Lon)[2]-range(species$Lon)[1])/5
lon.r = range(species$Lon) + c(buffer/1.2*(-1), buffer/1.2)
lat.r = range(species$Lat) + c(buffer*(-1), buffer)
lon = seq(lon.r[1], lon.r[2], length.out = grid+1)
lat = seq(lat.r[1], lat.r[2], length.out = grid+1)

plot(log(1+log(1+pred)), xlab = "Longitude", ylab = "Latitude", main='',
    col = colorRampPalette(c("grey90", "green", "yellow", "red"))(12),
    xlim=lon.r, ylim=lat.r)
for (i in 1:(grid+1)){
  abline(v = lon[i], col="white",lwd=2)
  abline(h = lat[i], col="white",lwd=2)
}
lines(shape, lwd=1.5)
points(species$Lon, species$Lat, pch=16,
        cex=log(species$size)/2, col=adjustcolor("red", 0.5))
```

Predicted and observed abundance

```
par(mfrow=c(1,1))
plot(log(1+log(1+pred)), xlab="Longitude", ylab="Latitude",
      main="", col=colorRampPalette(c("grey90", "green", "yellow",
"red"))(12),
      xlim=c(93.6, 94.2), ylim=c(33.8, 34.2)) #topo.colors(20)

lines(shape, lwd=2)
points(Data$Lon[Data$Name=="kiang"],
Data$Lat[Data$Name=="kiang"],
          cex=log(Data$size)/0.7, col=adjustcolor("black", 0.5), pch=16)
points(Data$Lon[Data$Name=="kiang"],
Data$Lat[Data$Name=="kiang"],
          cex=log(Data$size)/0.7, col="white", pch=1)

library(calibrate)
textxy(Data$Lon[Data$Name=="kiang"],
Data$Lat[Data$Name=="kiang"],
          Data$size[Data$Name=="kiang"], cex=0.8, col="red")
```



Adjust animal abundance

```
# Estimate animal abundacne
tracks = trackPoints(shape)
EST = estPopSize(pred, tracks, kiang, Average.p); EST
```

Original prediction	Predicted population on the routes	Observed population on the routes	Final estimation
1079714.8	5792.98	601.4797	112106

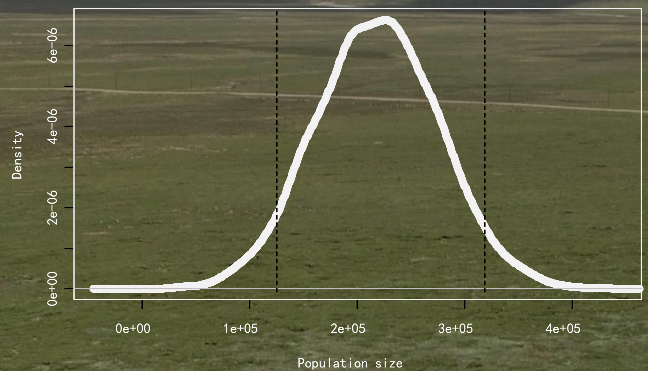
```
# Total number of individuals for the raster
pop_ori = cellStats(pred, stat='sum', na.rm=TRUE) # 1079715
pre <- extract(pred, tracks) # predicted animal abundance on survey routes
pop_pre = sum(pre, na.rm=T) # The predicted number of individuals on the route, 5792
kiang = kiang[kiang$distance<=500, ] # keep occ within 500m to match quadrat of 1 km
pop_obs = sum(kiang$size) # observation, 449
pop_obs_adj = pop_obs / Average.p # distance sampling adjustment, 601
adjust = pop_pre / pop_obs_adj # SDM adjustment, 9.6
# Adjusted animal abundance
pop_est = cellStats(pred, stat='sum', na.rm=TRUE) / adjust # 112105
```

Quantitively uncertainty

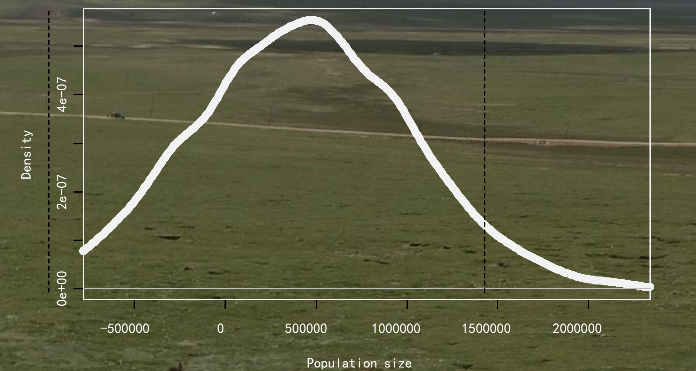
```
adjust.uncertainty = spatialMatch(kiang, tracks, pred, 4); adjust.uncertainty
survey.uncertainty = 1 - Average.p
model.uncertainty
adjust.uncertainty

# Confidence intervals
CI(EST[4], survey.uncertainty, model.uncertainty, adjust.uncertainty) # 41643.09 206444.71
CI(EST[4], 0.001, model.uncertainty, adjust.uncertainty) # 54478.44 174285.24
CI(EST[4], survey.uncertainty, 0.001, adjust.uncertainty) # 15452.1 158504.4
CI(EST[4], survey.uncertainty, model.uncertainty, 0.001) #  54909.34 173578.67
CI(EST[4], 0.001, 0.001, adjust.uncertainty) #  92537.83 131794.09
CI(EST[4], 0.001, model.uncertainty, 0.001) # 55481.11 168557.00
CI(EST[4], survey.uncertainty, 0.001, 0.001) # 55980.13 167354.87
```

Confidence intervals

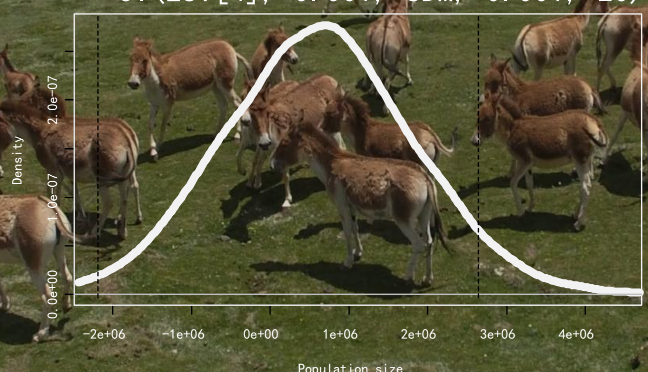
CI(EST[4], 0.001, 0.001, sd.Match, 2)



CI(EST[4], detection, 0.001, 0.001, 10)



CI(EST[4], 0.001, SDM, 0.001, 20)



cameratrapR

Ecological Informatics 69 (2022) 101597

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

cameratrapR: An R package for estimating animal density using camera trapping data

Xinhai Li [a,b,*], Huidong Tian [c], Zhengji Piao [d], Guiming Wang [e], Zhishu Xiao [a,b], Yuehua Sun [a,b,*], Erhu Gao [f,**], Marcel Holyoak [g]

[a] Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
[b] University of Chinese Academy of Sciences, Yuquan Road, Beijing 100049, China
[c] Cancer Registry of Norway, Ullernchausseen 64, 0379 Oslo, Norway
[d] Changbai Mountain Academy of Sciences, Yanbian 133613, China
[e] Department of Wildlife, Fisheries and Aquaculture, Mississippi State University, Mississippi State, MS 39762-9690, USA
[f] Academy of Inventory and Planning, National Forestry and Grassland Administration, Beijing 100714, China
[g] Department of Environmental Science and Policy, University of California, 1 Shields Avenue, Davis, CA 95616, USA

ARTICLE INFO

Keywords:
Camera trapping
Correlated random walk
Footprint chain
Movement pattern
Population density
R

ABSTRACT

1. Camera trapping plays an important role in wildlife surveys, and provides valuable information for estimation of population density. While mark-recapture techniques can estimate population density for species that can be individually recognized or marked, there are no robust methods to estimate density of species that cannot be individually identified.
2. We developed a new approach to estimate population density based on the simulation of individual movement within the camera grid. Simulated animals followed a correlated random walk with the movement parameters of segment length, angular deflection, movement distance and home-range size derived from empirical movement paths. Movement was simulated under a series of population densities. We used the Random Forest algorithm to determine the population density with the highest likelihood of matching the camera trap data. We developed an R package, cameratrapR, to conduct simulations and estimate population density.
3. Compared with line transect surveys and the random encounter model, cameratrapR provides more reliable estimates of wildlife density with narrower confidence intervals. Functions are provided to visualize movement paths, derive movement parameters, and plot camera trapping results.
4. The package allows researchers to estimate population sizes/densities of animals that cannot be individually identified and cameras are deployed in a grid pattern.

1. Introduction

In recent years, camera trapping has been widely used in wildlife surveys (Li et al., 2018; Rich et al., 2017), and a number of methods have been developed for estimating population density from such data. For animals that can be individually recognized, such as tigers and leopards, capture-recapture analysis of camera-trap data has become a common method for estimating population density (Gardner et al., 2010; Royle et al., 2009). For animals that cannot be individually recognized, several approaches are available for estimating population density (Gilbert et al., 2020). Distance sampling methods have been applied to analysis of camera trap data for animal density estimation (Bessone et al., 2020; Cappelle et al., 2019), and proved reliable results for estimating density of homogenously rather than patchily-distributed species (Bessone et al.,

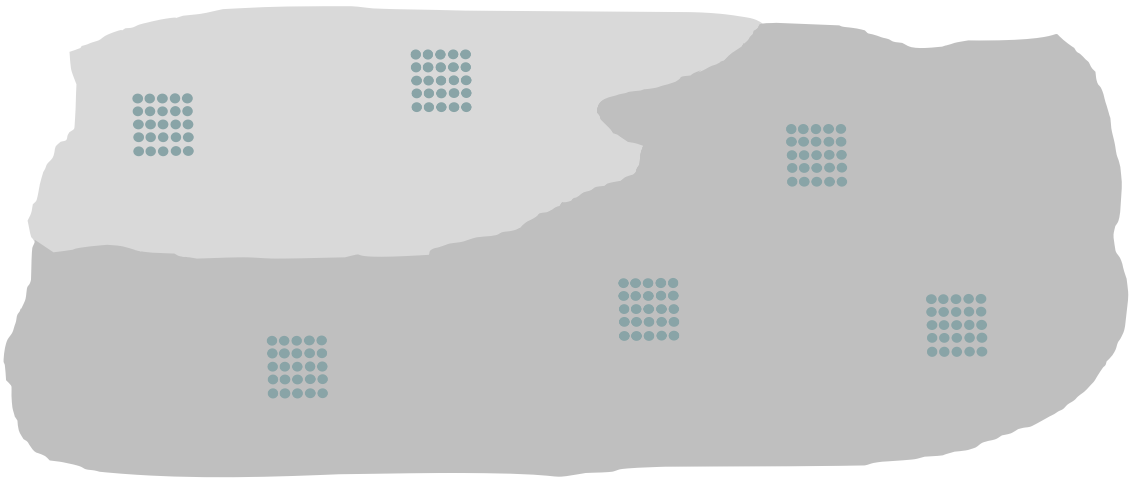
* Corresponding authors at: Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.
** Corresponding author.
E-mail addresses: lixh@ioz.ac.cn (X. Li), sunyh@ioz.ac.cn (Y. Sun), 805482272@qq.com (E. Gao).

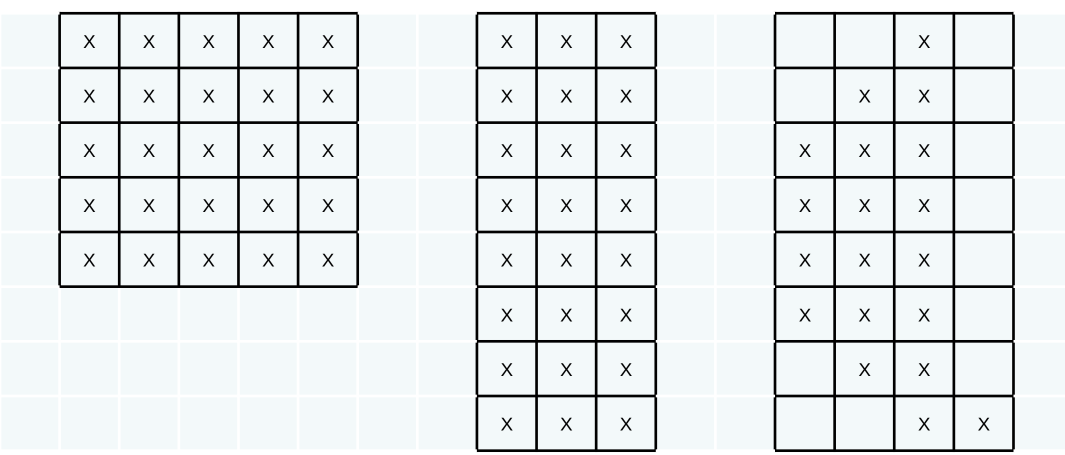
https://doi.org/10.1016/j.ecoinf.2022.101597
Received 24 June 2021; Received in revised form 8 February 2022; Accepted 8 February 2022
Available online 12 February 2022

install_github("Xinhai-Li/cameratrapR")

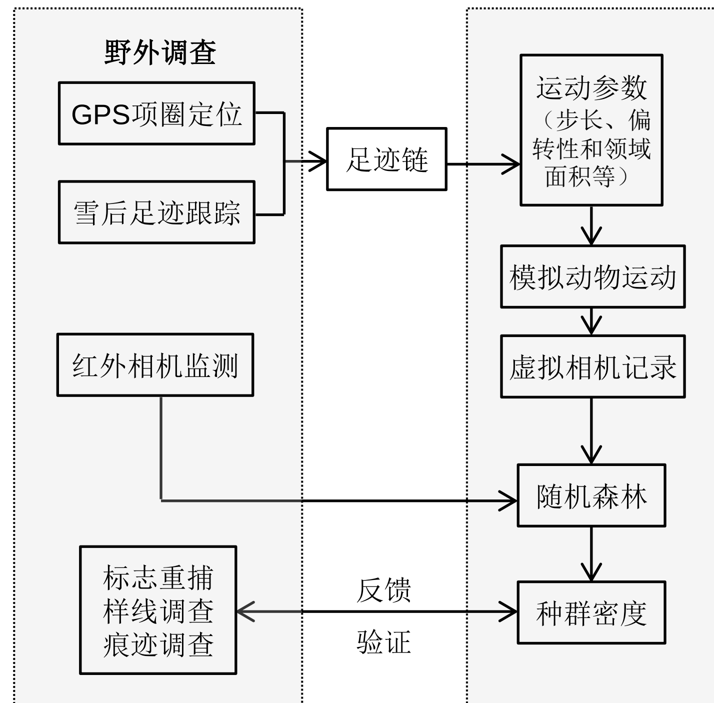
Li X, Tian H, Piao Z, Wang G, Xiao Z, Sun Y, Gao E, Holyoak M. 2022. cameratrapR: An R package for estimating animal density using camera trapping data. Ecological Informatics 69:101597.

物种	物种英文名	领域面积/校正值 (km^2)	领域面积的标准差	运动步长 (m)	偏转角度	足迹链个数
Species	Name	Area /Adjusted area (km^2)	Area_SD	Step length ± SD (m)	SD of Angular deflection	No. of footprint chains
草兔	Cape hare	0.13/0.25	0.11	13.49±13.27	37.09	3
狐	Corsac fox	0.61/1.5	0.09	14.50±19.02	21.89	3
猞猁	Eurasian lynx	8.81	13.21	21.08±13.40	50.15	6
斑羚	Goral	9.2	/	17.20±34.78	32.94	1
豹猫	Leopard cat	0.29/4	0.15	16.97±12.73	31.93	4
黄喉貂	Marten	0.54/2	0.68	13.98±7.02	52.27	4
貉	Raccoon dog	0.1/2	0.05	12.12±5.43	30.28	3
马鹿	Red deer	1.95	1.15	15.46±10.21	40.43	6
西伯利亚狍	Roe deer	0.66	0.69	13.82±6.77	43.08	25
紫貂	Sable	0.35/2	0.38	11.27±6.91	53.20	17
野猪	Wild boar	1.15	1.28	16.13±11.01	43.21	15





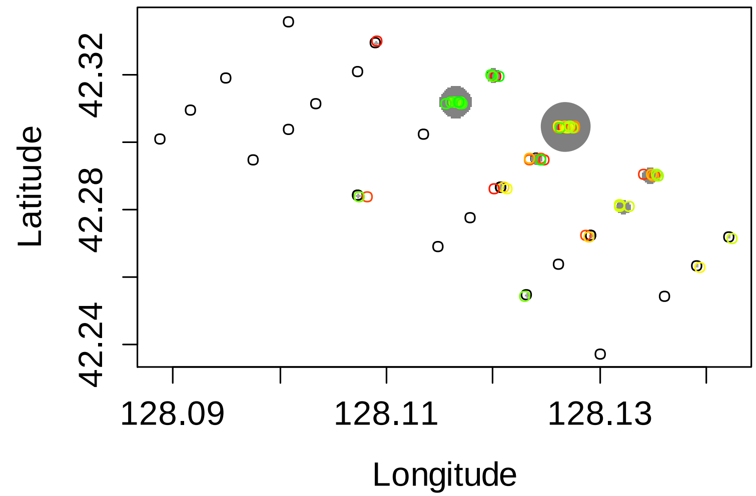
cameratrapR



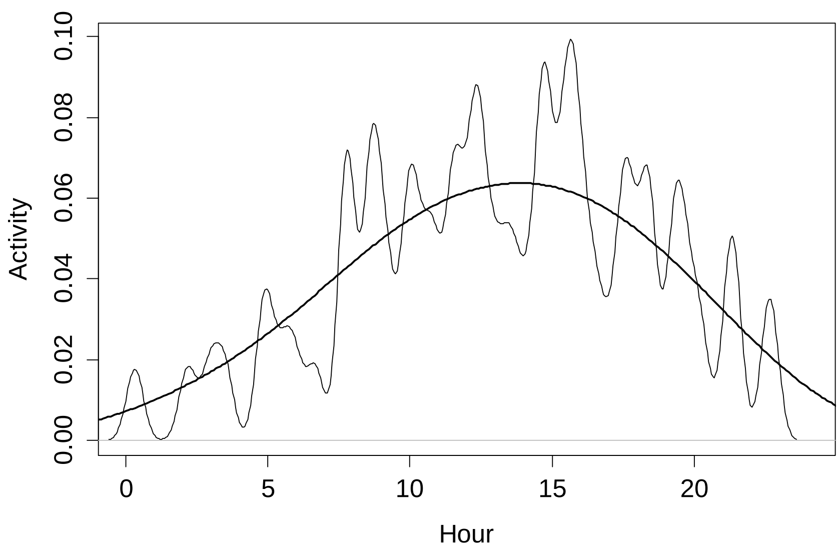
```
data(trapresult)
head(trapresult)
x = trapresult
```

Species	Lat	Lon	Group_size	Date	Time	Species_C
Wild boar	42.31898	128.095	1	2013/10/18	5:57:46	野猪
Wild boar	42.30374	128.1008	1	2013/10/18	13:29:22	野猪
Wild boar	42.31126	128.1033	2	2013/10/18	16:15:34	野猪
Wild boar	42.30224	128.1135	1	2013/10/18	10:10:07	野猪
Wild boar	42.26907	128.1149	3	2013/10/18	11:25:15	野猪
Wild boar	42.27751	128.1179	1	2013/10/18	12:08:01	野猪

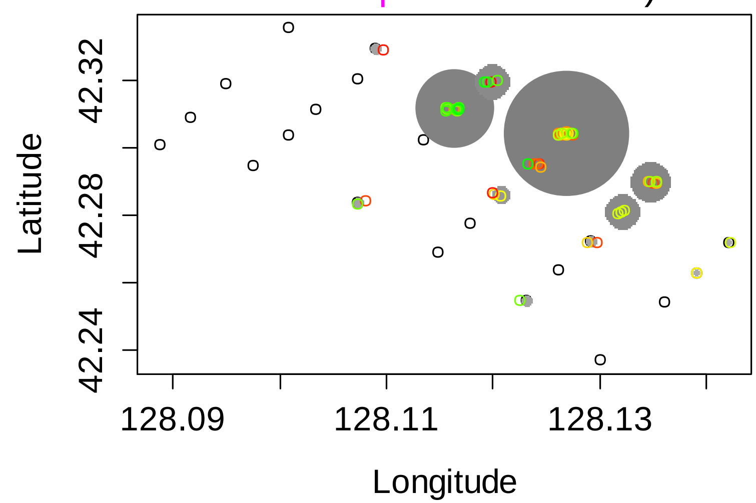
plotCamtrap(x)



dailyRhythm(x)



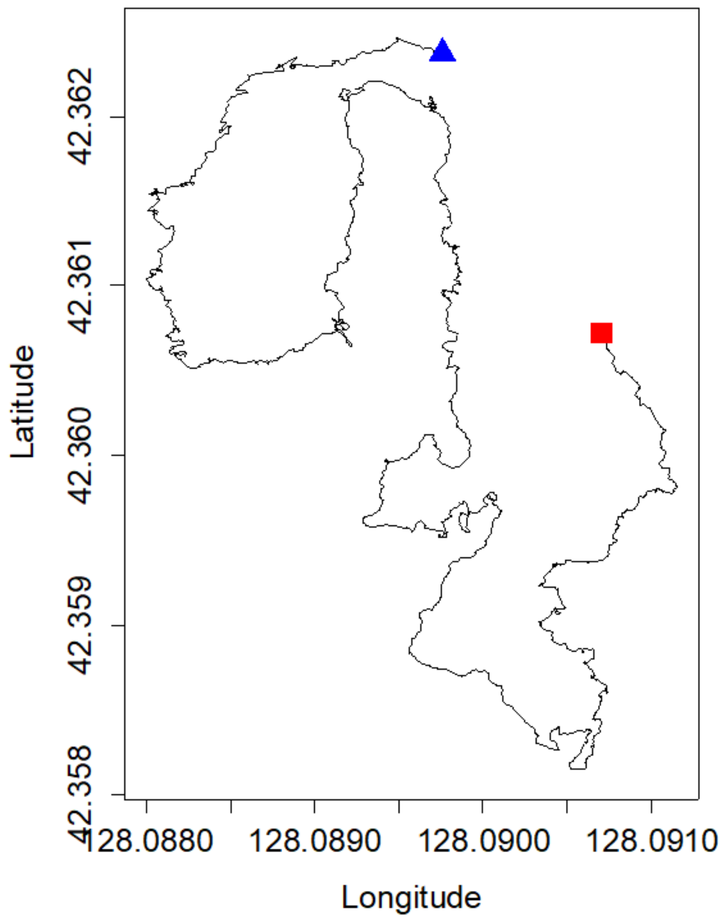
plotCamtrap(x, circle.size = 0.5, point.scatter = 5)



Wild boar movement on Dec. 8, 2013

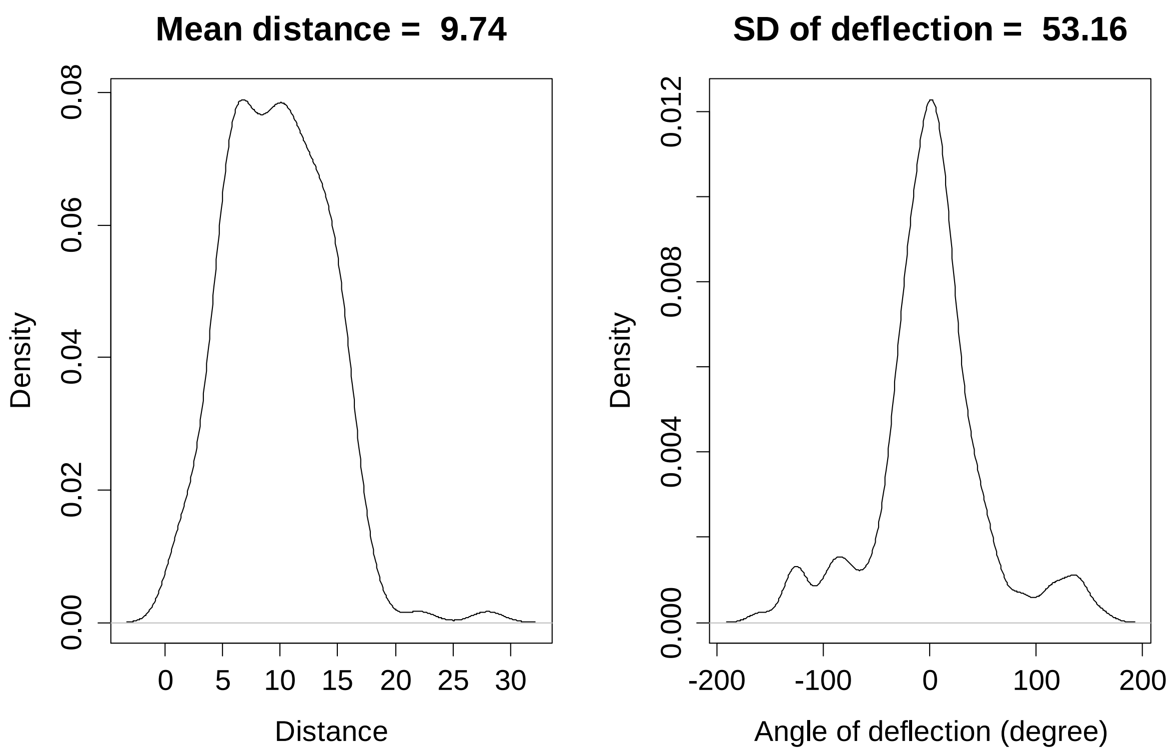
```
plotFootprint(chain)
head(chain)
```

Species	Lat	Lon	Date
wild boar	42.36237	128.0898	2013/12/8
wild boar	42.36237	128.0896	2013/12/8
wild boar	42.36236	128.0895	2013/12/8
wild boar	42.36235	128.0893	2013/12/8
wild boar	42.36231	128.0892	2013/12/8
wild boar	42.36231	128.0890	2013/12/8



Moving distance and angle of deflection

moveBias(chain, 11)



Simulate animal movement

- The movement is simulated by giving a random starting location (X_0, Y_0), defining a step length S and step direction θ, and running for N times. The animal location at time t+1 is:

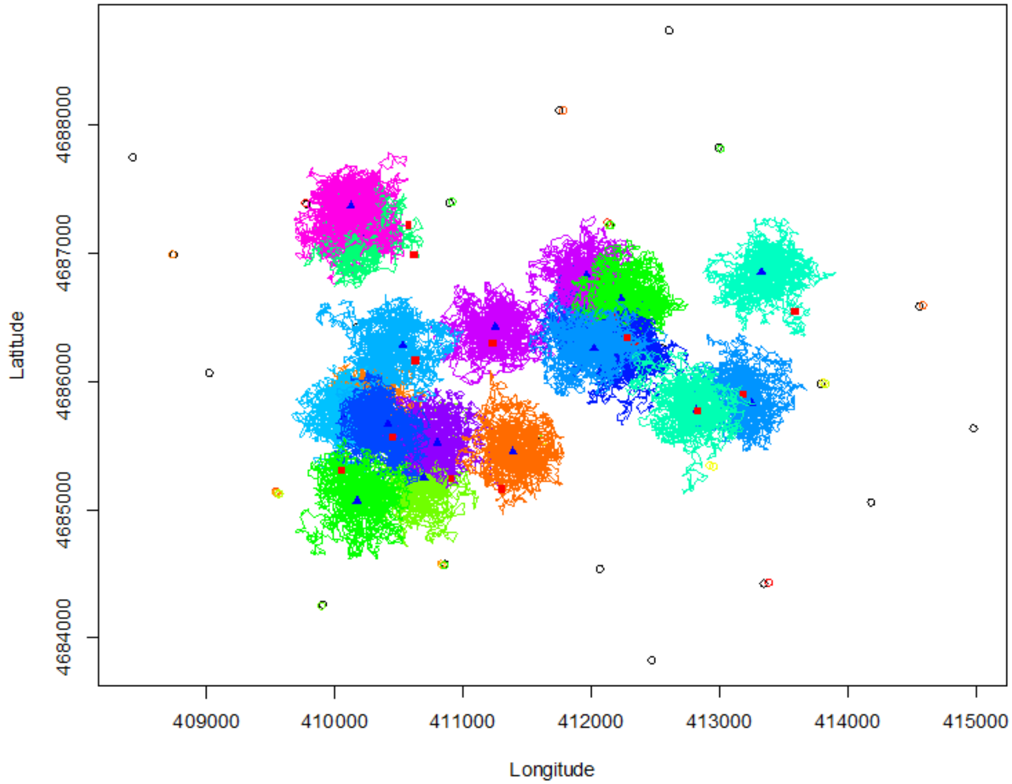
$$\mathrm{X}_{t+1} = \mathrm{X}_t + S_{(t+1)} \times \cos(\theta_t + b)$$

$$\mathrm{Y}_{t+1} = \mathrm{Y}_t + S_{(t+1)} \times \sin(\theta_t + b)$$

- where $S_{(t+1)}$ is step length, which is normal distributed number with a mean M and standard variance σ_1; $\theta_{(t)}$ is the movement direction at time t. b is the variable defining how much the direction changes (it is normal distributed number with a mean 0 and standard variance σ_2). The number of steps is associated with the duration of camera trapping and animal activity level.

Simulation the animal movement and camera trapping

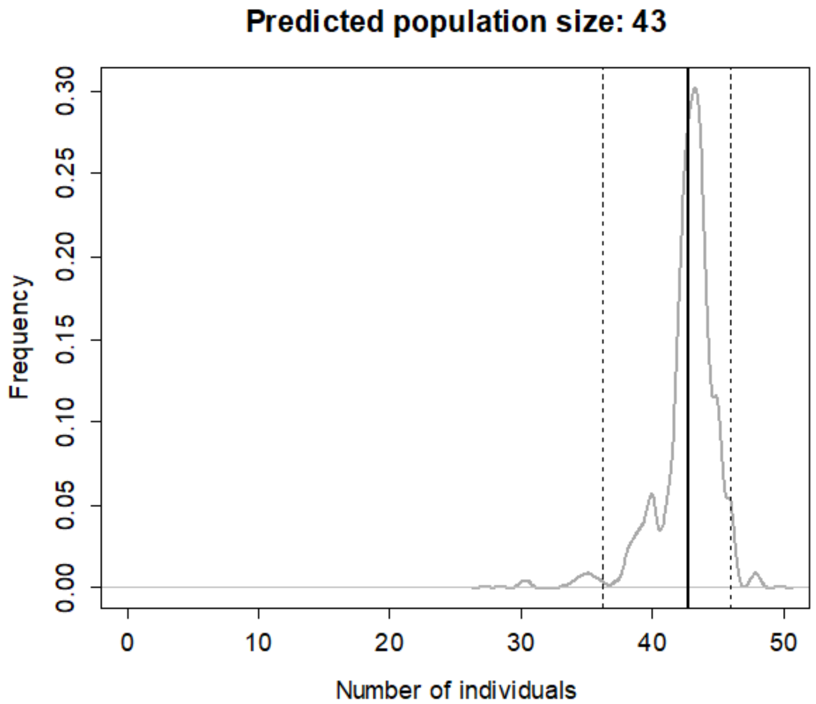
simuCamtrap(x, ind = 20, iteration = 2)



Estimate population size

```
attach(sim.out)
predictCamtrap(sim.out, x, plot=TRUE)
```

2.5%	36.2
50%	43.0
97.5%	46.0



Assignment

- For a given population (e.g. N=1000), try simple random sampling (n=100), systematic sampling (n=100), stratified sampling (you need to make groups for this) (n=100), and cluster sampling (n=100).