

Analysis of Stratified Surveys

Stratification

- Why stratify?
- Stratification by:
 - Geographic area
 - Survey
 - Species / cluster size
- Limitations of Distance

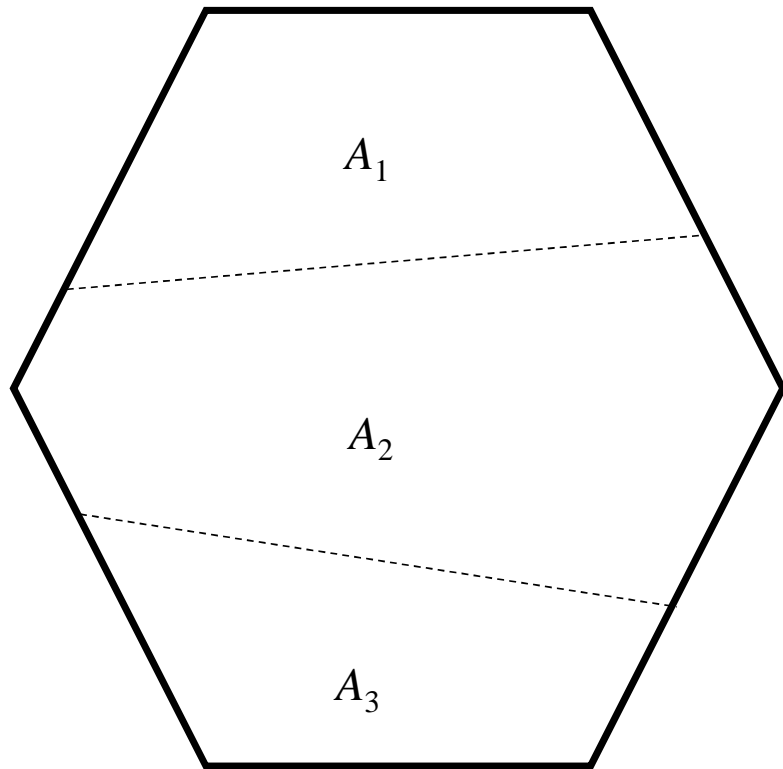
Stratification is used to:

- reduce variance and improve precision
- and for producing estimates in regions of interest

Stratify by:

- AREA or GEOGRAPHIC REGION
 - the study region is partitioned into smaller regions
- SURVEY
 - used when different surveys cover the same geographic area
- POPULATION/SPECIES/CLUSTER SIZE
 - same geographic region with different 'sub-stocks' in it

Area/Geographic stratification



Estimate density in each sub-region

$$\hat{D}_1, \hat{D}_2, \hat{D}_3$$

Abundance in each sub-region is given by

$$\hat{N}_1 = A_1 \hat{D}_1$$

$$\hat{N}_2 = A_2 \hat{D}_2$$

$$\hat{N}_3 = A_3 \hat{D}_3$$

Total size of study region

$$A = A_1 + A_2 + A_3$$

Total abundance is

$$\hat{N} = \hat{N}_1 + \hat{N}_2 + \hat{N}_3$$

$$= A_1 \hat{D}_1 + A_2 \hat{D}_2 + A_3 \hat{D}_3$$

Overall (Global in Distance) density is

$$\hat{D} = \frac{\hat{N}}{A} = \frac{A_1 \hat{D}_1 + A_2 \hat{D}_2 + A_3 \hat{D}_3}{A_1 + A_2 + A_3}$$

$$= \left(\frac{A_1}{A} \right) \hat{D}_1 + \left(\frac{A_2}{A} \right) \hat{D}_2 + \left(\frac{A_3}{A} \right) \hat{D}_3$$

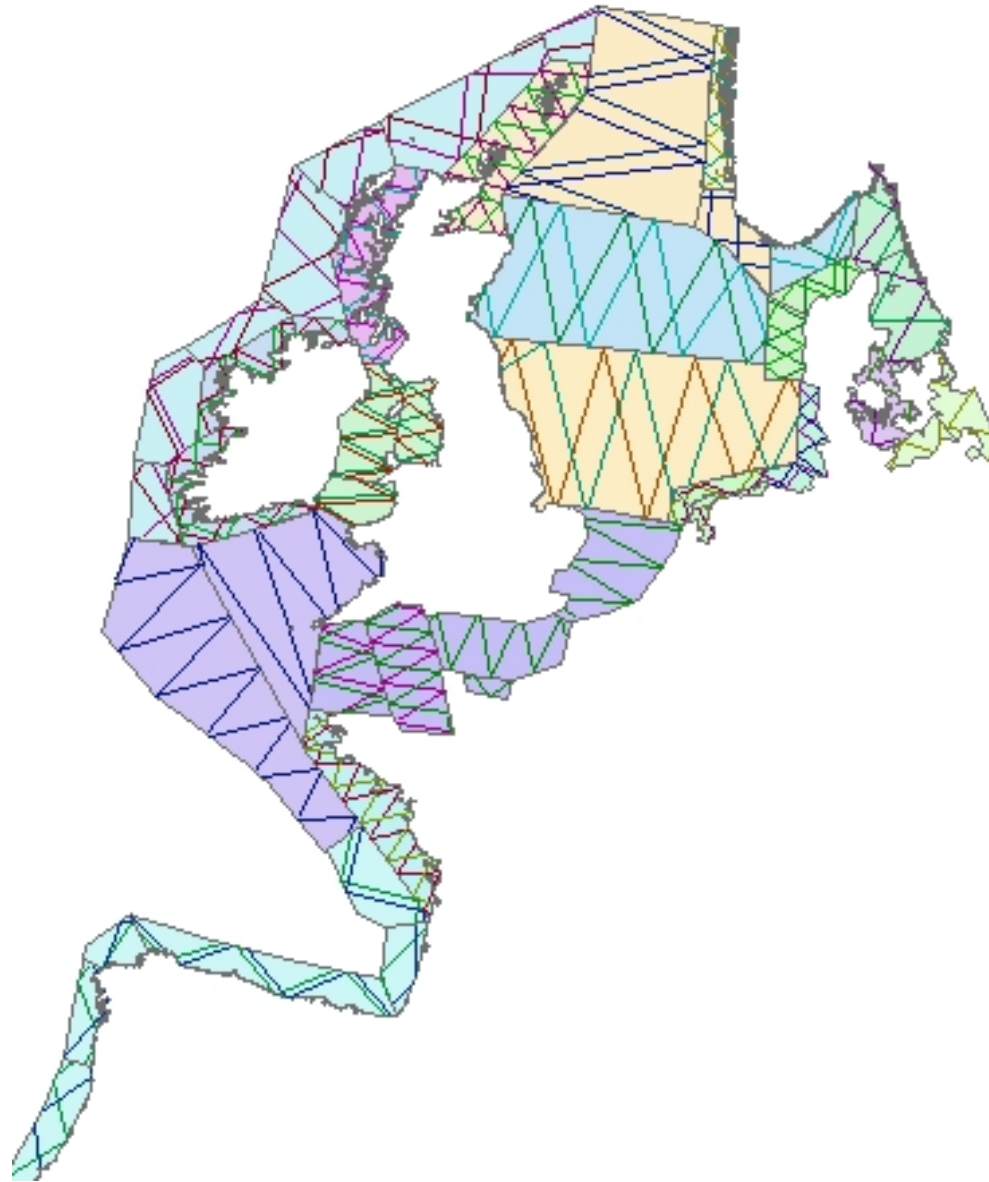
$$= \sum_{i=1}^3 \left(\frac{A_i}{A} \right) \hat{D}_i$$

Note form of equation

Example: SCANS II (2005)



SCANS II survey effort



Example of stratified data

Distance - Stratify example

File View Tools Data Window Help

Project Browser

Data Maps Designs Surveys Analyses Simulations

Data layers

Study Area

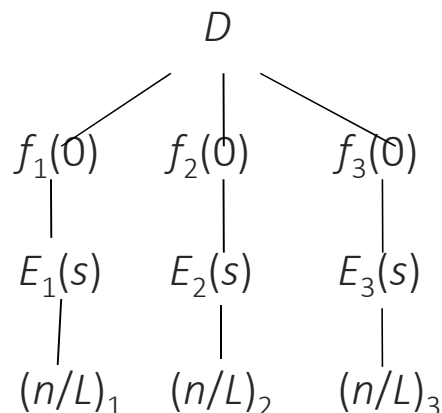
Region

Line

Contents of Observation layer 'Observation' and all fields from higher layers

Study Area				Region			Line transect			Observation			
ID	Label	G0	G0 SE	ID	Label	Area	ID	Label	Line length	ID	Perp distance	Cluster size	Cluster strat
ID	Label	Decimal	Decimal	ID	Label	Decimal	ID	Label	Decimal	ID	Decimal	Integer	Integer
n/a	n/a	[None]	[None]	n/a	n/a	nautmi2	n/a	n/a	nautmi	n/a	nautmi	[None]	[None]
Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int	Int
1	Stratify example	0.8367	0.1738	1	Ideal Habitat	85000	12	12	1	31	0.1	1	1
							13	13	80	32	0.68	1	1
										33	0.31	2	2
										34	0.58	2	2
										35	0.49	1	1
										36	0.46	2	2
										37	0.36	2	2
										38	0.09	2	2
										39	0.03	2	2
										40	0.49	1	1
										41	1.94	8	3
										42	1.1	10	3
										43	0.85	5	3
										44	0.63	7	3
										45	0.39	3	3
										46	0.65	1	1
										47	1.16	2	2
										48	0.91	2	2

Example: Full geographic stratification



Model Definition Properties: [Full Stratification]

Analysis Engine: CDS - Conventional distance sampling

Estimate | Detection function | Cluster size | Multipliers | Variance | Misc.

Stratum definition

☐ No stratification

☒ Use layer type: Stratum

☐ Post-stratify, using: Stratum Area

Sample definition (for encounter rate)

Use layer type: Sample

Quantities to estimate and level of resolution

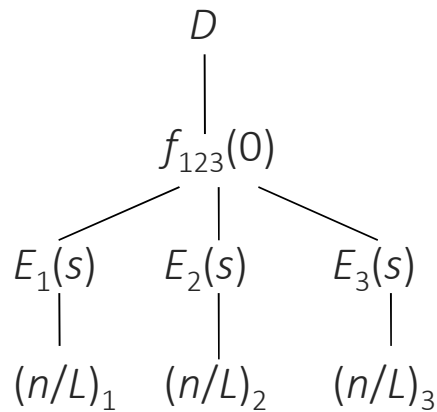
	Level of resolution of estimates		
	Global	Stratum	Sample
Density	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Encounter rate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Detection function	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cluster size (if required)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Global density estimate is Mean of stratum estimates

weighted by Stratum area ☐ Strata are replicates

Defaults Name: Full Stratification OK Cancel

Example: $f(0)$ pooled



Model Definition Properties: [f(0) pooled]

Analysis Engine: CDS - Conventional distance sampling

Estimate | Detection function | Cluster size | Multipliers | Variance | Misc.

Stratum definition

☐ No stratification

☒ Use layer type: Stratum

☐ Post-stratify, using: Stratum Area

Sample definition (for encounter rate)

Use layer type: Sample

Quantities to estimate and level of resolution

	Level of resolution of estimates		
	Global	Stratum	Sample
Density	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Encounter rate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Detection function	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cluster size (if required)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Global density estimate is: Mean of stratum estimates

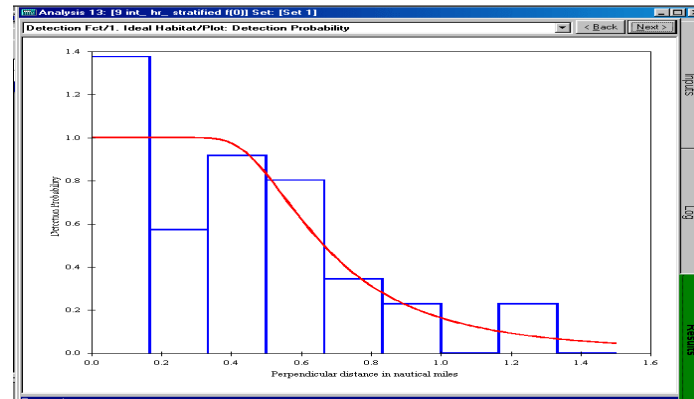
weighted by: Stratum area ☐ Strata are replicates

Defaults Name: f(0) pooled OK Cancel

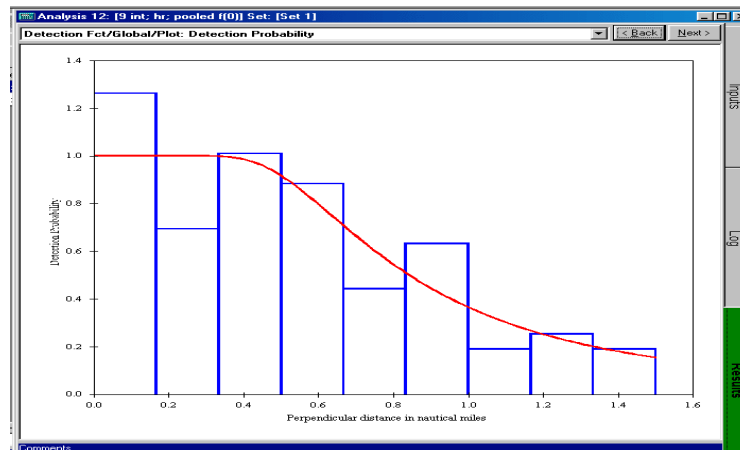
Pooled vs Stratified $f(0)$

Stratified

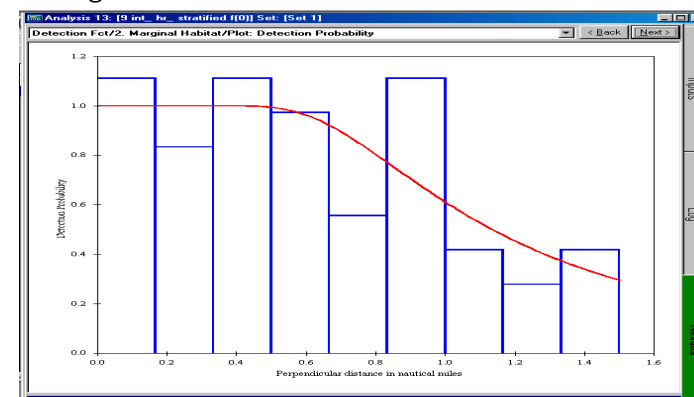
Ideal Habitat $n=39$



Pooled $n=88$



Marginal Habitat $n=49$



It is a Model Selection Problem

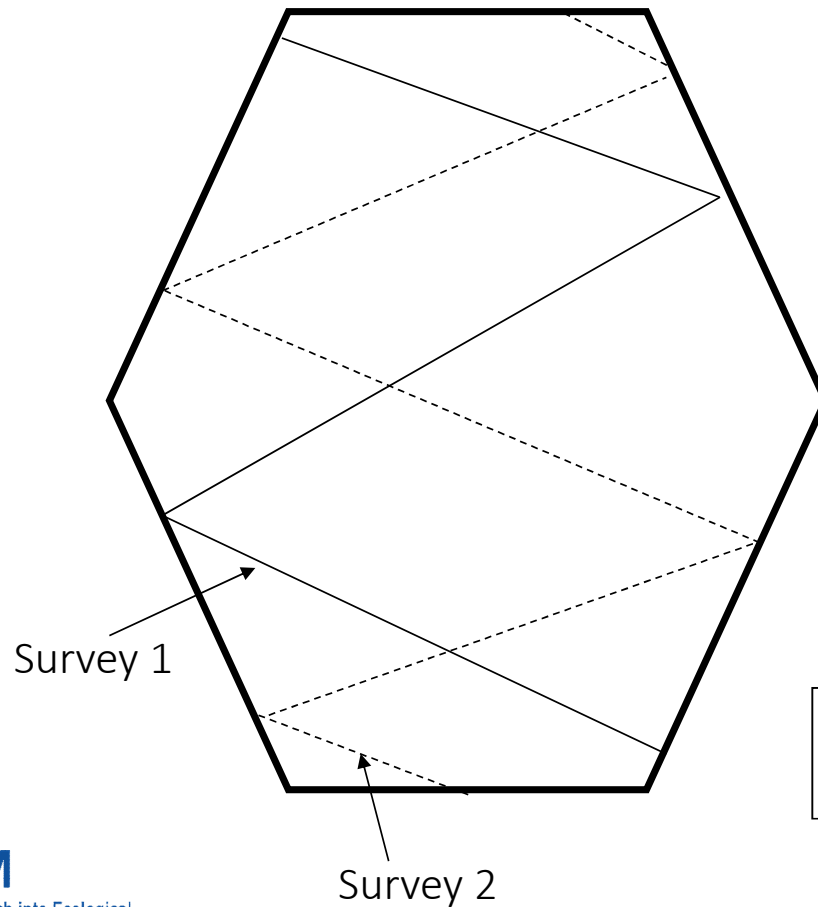
	Pooled	Stratum 1	Stratum 2	Stratum Sum
Log likelihood $\log_e(L)$	-180.490	-72.699	-104.676	-177.375
No. parameters (q)	2	2	2	4
AIC	364.980	149.398	213.352	362.75

Criterion for stratification of $f(0)$:

Fit separate $f(0)$ for each strata if

$$AIC_{pooled} > \sum_{strata} AIC_{stratum}$$

Non-geographic stratification -- Stratification by survey



Let L_i be effort for survey i

Global density is given by

$$\begin{aligned}\hat{D} &= \left(\frac{L_1}{L_1 + L_2} \right) \hat{D}_1 + \left(\frac{L_2}{L_1 + L_2} \right) \hat{D}_2 \\ &= \sum_{i=1}^2 \left(\frac{L_i}{L} \right) \hat{D}_i\end{aligned}$$

This is the same form as before, but weighting factor now depends on effort

Stratification by survey

Model Definition Properties: [Strata are surveys]

Analysis Engine: CDS - Conventional distance sampling

Estimate | Detection function | Cluster size | Multipliers | Variance | Misc.

Stratum definition

☐ No stratification

☒ Use layer type: Stratum

Layer type: Stratum

Field name:

☐ Post-stratify, using: Stratum Area

Sample definition (for encounter rate)

Use layer type: Sample

Quantities to estimate and level of resolution

	Level of resolution of estimates		
	Global	Stratum	Sample
Density	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Encounter rate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Detection function	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cluster size (if required)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

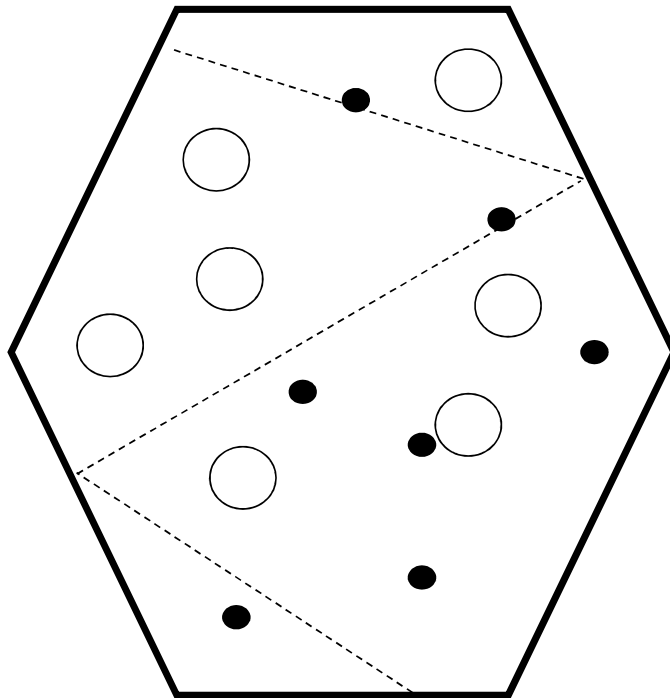
Global density estimate is Mean of stratum estimates

weighted by Total effort in stratum ☒ Strata are replicates

Defaults Name: Strata are surveys OK Cancel

Stratification by species

○ *Species 1*
● *Species 2*



$$\hat{D} = \hat{D}_{sp1} + \hat{D}_{sp2}$$

Model Definition Properties: [Post stratify by cluster]

Analysis Engine: CDS - Conventional distance sampling

Estimate | Detection function | Cluster size | Multipliers | Variance | Misc.

Stratum definition

☐ No stratification Layer type: Field name:

☐ Use layer type: Stratum

☒ Post-stratify, using: Observation Cluster strat

Sample definition (for encounter rate)

Use layer type: Sample

Quantities to estimate and level of resolution

	Level of resolution of estimates		
	Global	Stratum	Sample
Density	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Encounter rate	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Detection function	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Cluster size (if required)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

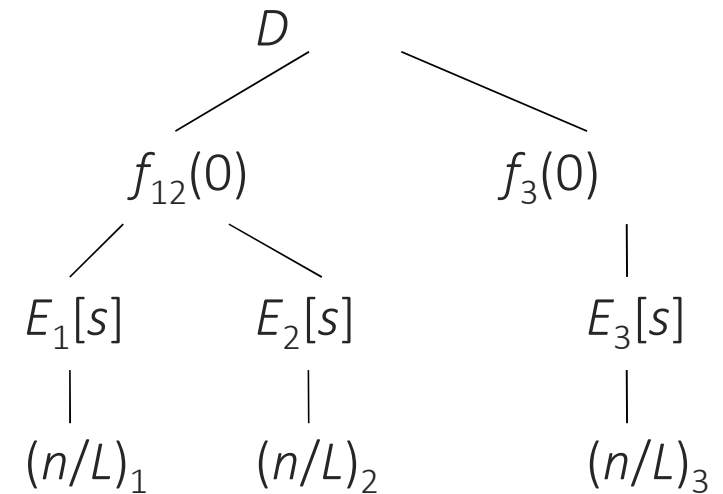
Global density estimate is Sum of stratum estimates

weighted by ☐ Strata are replicates

Defaults Name: Post stratify by cluster OK Cancel

Limitations in Distance

- Distance cannot currently do multilevel stratification in one run
- Two runs are necessary
 - Estimate $f(0)$, $E[s]$ and n/L by stratum
 - Combine strata 1 and 2 to estimate $f_{12}(0)$
- Care must be taken when calculating CVs because the density estimates for stratum 1 and 2 have an estimated $f(0)$ in common



Alternatives to stratification in Distance

- Small sample sizes can lead to low precision in stratum-specific estimates
- An alternative approach to reducing bias due to heterogeneity is Multiple Covariates Distance Sampling (MCDS)
 - Covariates, other than distance, are incorporated into the scale parameter of the detection function
 - MCDS can be used to fit the detection function at multiple levels e.g. stratum-specific density estimates can be obtained even if you don't have enough data to fit separate detection functions for each stratum
 - MCDS methods are covered in an upcoming lecture.