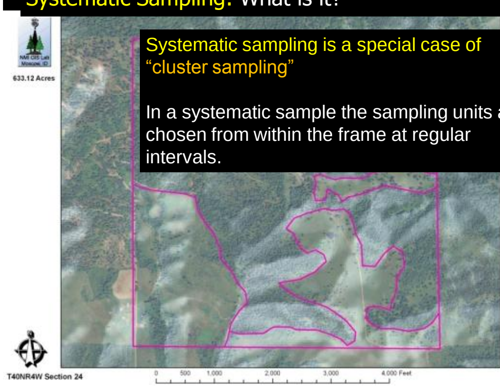
FOR 474: Forest Inventory Techniques

Systematic Sampling

- What is Systematic Sampling?
- Why do we use it?
- How do we use it?
- Distance between samples
- How effective is this method?

Systematic Sampling: What is it?

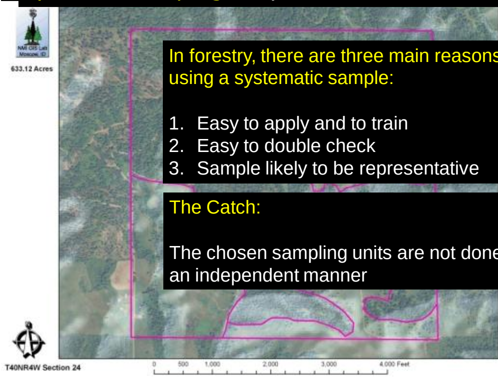


Systematic sampling is a special case of "cluster sampling"

In a systematic sample the sampling units are chosen from within the frame at regular intervals.

Systematic Sampling: What is it?

Systematic Sampling: Why do we use it?



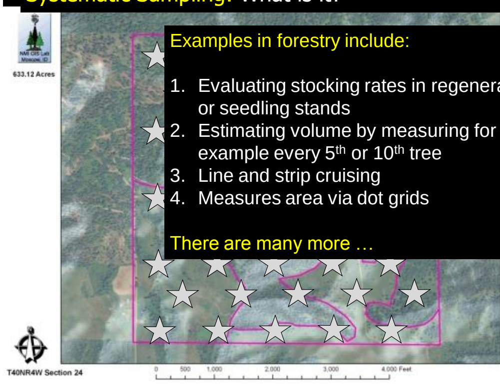
In forestry, there are three main reasons for using a systematic sample:

1. Easy to apply and to train
2. Easy to double check
3. Sample likely to be representative

The Catch:

The chosen sampling units are not done in an independent manner

Systematic Sampling: What is it?



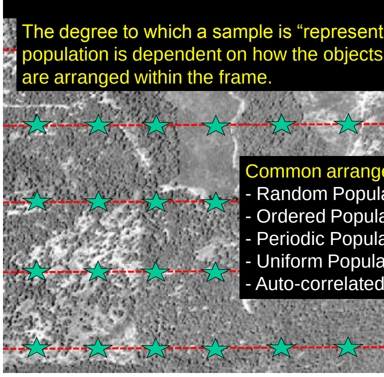
Examples in forestry include:

1. Evaluating stocking rates in regeneration or seedling stands
2. Estimating volume by measuring for example every 5th or 10th tree
3. Line and strip cruising
4. Measures area via dot grids

There are many more …

Systematic Sampling: Population Types

The degree to which a sample is "representative" of the population is dependent on how the objects in the population are arranged within the frame.

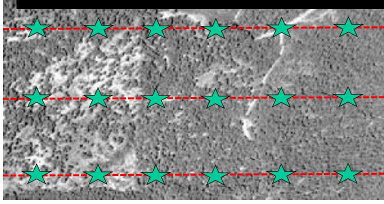


Common arrangements include:
- Random Populations
- Ordered Populations
- Periodic Populations
- Uniform Populations
- Auto-correlated Populations

Systematic Sampling: Populations

The probability of occurrence in a random population at any location is equal. In random populations systematic (and random) samples will be representative.

In non-random populations systematic samples will tend to be more representative than random samples as the sample size increases.



Systematic Sampling: Populations

Ordered populations exhibit some form of rank, such as age.

Periodic populations follow undulates patterns like a sine or cosine wave. Periodic populations occur naturally in forestry due to aspect, slope position, and disturbances.

Auto correlated populations occur when individuals are more likely to be similar to their closer neighbors than those further away AND when a relationship can be inferred as a function of the position. In these populations, the value of any component can be inferred from the value of one that occurred earlier in the sequence.

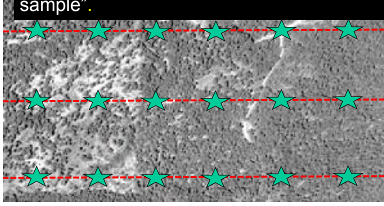
Stratified populations are a type of auto correlated population, which can occur in forests due to water availability, topography, disturbance, and management history.

Systematic Sampling: What is it?

Consider the following plot list:

{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}

We call this the “basic” sampling frame. If we decide that we will measure every L^{th} plot, then we have a “1 in L systematic sample”.



Systematic Sampling: What is it

If L = 5 then we would collect the following sets of data:

{1,6,11} {2,7,12} {3,8,13} {4,9,14} {5,10,15}

Typically only one of these subsets are selected for sampling. This means the subset is a sample. Each of these samples contains n=3 elements.

Sampling intensity = n/N = 3/15 = 20% = 1/L

Systematic Sampling: Scenario 1

Each subset is a cluster and only one cluster is selected

{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}

Cluster 1: 1, 6, 11
Cluster 2: 2, 7, 12
Cluster 3: 3, 8, 13
Cluster 4: 4, 9, 14
Cluster 5: 5, 10, 15

In this scenario each cluster spans the range of data (i.e. not compact). The samples are interlocking. No sub-sampling – each cluster will likely be representative.

Systematic Sampling: Scenario 2

Effectively 2 stage cluster sampling:

{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15}

	Cluster 1	Cluster 2	Cluster 3
Sample 1:	1,	6,	11
Sample 2:	2,	7,	12
Sample 3:	3,	8,	13
Sample 4:	4,	9,	14
Sample 5:	5,	10,	15

Each cluster is compact and contains L =5 elements. A subsample is selected from EACH cluster (n=1). In this case the sample is likely to be representative while the cluster is not.

Systematic Sampling: Population Parameters

Since systematic sampling is a special case of cluster sampling, the mean is the same.

Under Scenario 1 if we use the simple random sample equation for variance of the mean the answer will be larger than that we would get if we had a simple random sample.

$$V(\bar{y}) = \frac{1}{M}\sum_{i}^{M}(\mu_i - \mu_o)^2$$

Under Scenario 2 we must use the following equation:

$$V(\bar{y}) = \frac{1}{M^2}\sum_{i}^{M}\sigma^2$$

Systematic Sampling: Effects of Trends

If a linear trend is present (with no random component) the variance of the mean of Scenario 1 is "n x greater" than Scenario 2:

$$V(\bar{y})_{Scen2} = \frac{1}{n}V(\bar{y})_{Scen1}$$

In other words, "the efficiency of Scenario 2 is n x greater than Scenario 1".

If the trend is a known relationship, it can be removed to produce a new population.

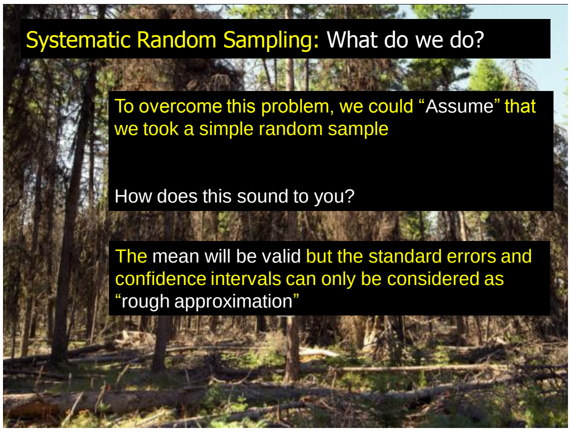
[n = sample size]

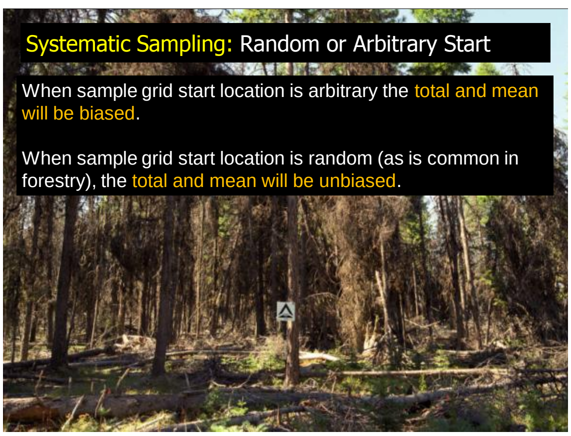
Systematic Sampling: Statistics

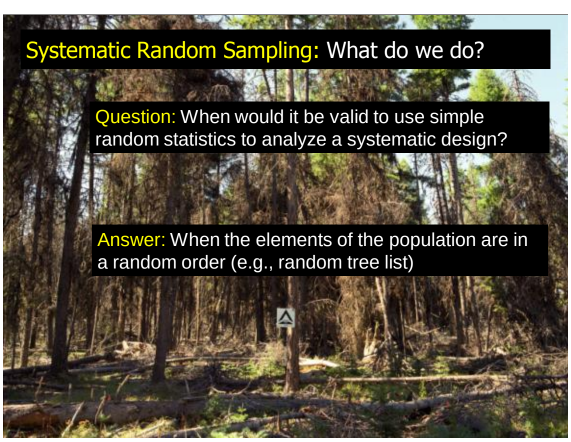
Scenario 1: The sample total, mean, and variance are based on all values within a single cluster.

Scenario 2: The sample consists of 1 elements from each of the M clusters. Therefore, sample size n = M.

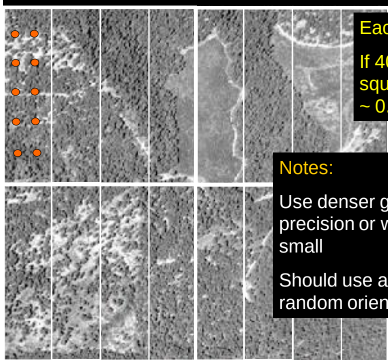
By either scenario, the variance of the total and the variance of the mean can not be calculated (due to divide by 0).


Systematic Random Sampling: What do we do?
To overcome this problem, we could "Assume" that we took a simple random sample
How does this sound to you?
The mean will be valid but the standard errors and confidence intervals can only be considered as "rough approximation"


Systematic Sampling: Random or Arbitrary Start
When sample grid start location is arbitrary the total and mean will be biased.
When sample grid start location is random (as is common in forestry), the total and mean will be unbiased.


Systematic Random Sampling: What do we do?
Question: When would it be valid to use simple random statistics to analyze a systematic design?
Answer: When the elements of the population are in a random order (e.g., random tree list)

Other Systematic Samples: Dot Grids to Measure Area



Each block = 8 ac

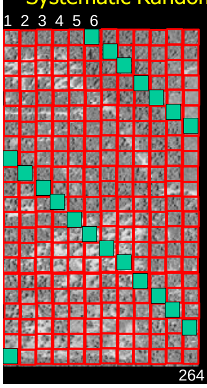
If 40 dots per block square then each dot ~ 0.2 acres

Notes:

Use denser grids to increase precision or when the region is small

Should use average of several random orientations

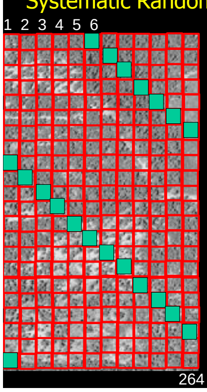
Systematic Random Sampling: How to do it?



Systematic Random Sampling:

1. We have a wanted sample size of n (20) and a population of N (264)
2. Calculate N/n = ~13
3. Select a random number between 1 and N, say = 6
4. Sample every k^{th} value

Systematic Random Sampling: How to do it

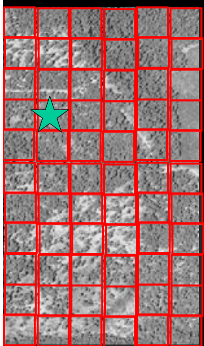


By acreage:

1. Divide the total Area (100 acres) by the number of samples (20) = 5
2. This 5 is the “area” represented by each point
3. For square plots, the spacing between plots is then the square root of the acreage:

   = √(43,560 *5) = ~ 467 feet

Systematic Random Sampling: How to do it



By a geographic map:

1. Randomly select a sample point from a whole population – this could be located anywhere
2. Then overlay grid and follow as steps a before

Notes: Default method, not easy to implement, more convenient than random sample

Systematic Random Sampling: How good is it?

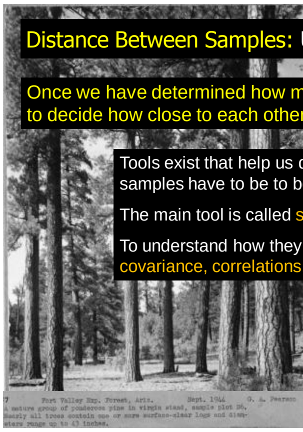
The sample will be inter-dispersed throughout the population and so may be representative if the stand is either totally heterogeneous or homogenous

Even though we selected a "random starting location", this is a controversial method because <u>you can't estimate the variance or standard error properly</u>

FOR 373: Forest Sampling Methods

Distance Between Samples

- Covariance
- Correlation
- Autocorrelation
- Semi-variograms
- Extracting Data: MAUP



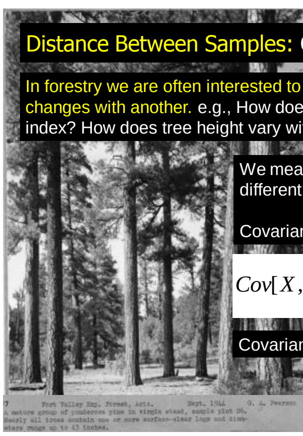
Distance Between Samples: Useful Tools

Once we have determined how many samples, we then have to decide how close to each other they can be.

Tools exist that help us determine how far apart samples have to be to be considered independent.

The main tool is called semi-variograms.

To understand how they work we first must understand covariance, correlations, and autocorrelation.



Distance Between Samples: Covariance

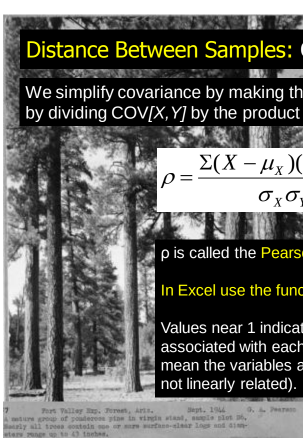
In forestry we are often interested to know how one variable changes with another. e.g., How does tree volume vary with site index? How does tree height vary with stand age?

We measure this variability between different metrics using covariance

Covariance is defined by:

$$Cov[X,Y] = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance can be + or -



Distance Between Samples: Correlation

We simplify covariance by making the values go between -1 and +1, by dividing COV*[X,Y]* by the product of the standard deviations:

$$\rho = \frac{\Sigma(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}$$

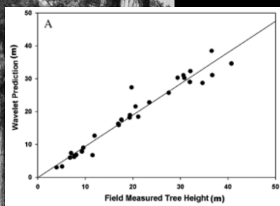
ρ is called the Pearson *Correlation Coefficient* (r)

In Excel use the function =PEARSON (data1,data2).

Values near 1 indicate that the 2 variables are associated with each other; but values near 0 does not mean the variables are independent (just that they are not linearly related).

Distance Between Samples: Correlation

Example: Comparison of tree heights: r = 0.97



IMPORTANT NOTE: This is only a measure of the strength of *linear* association between the two variables.

Distance Between Samples: Auto correlation

Spatial auto correlation is the correlation of a variable with itself through space.

If we imagine a forest where the location, age, species, and quality of the trees grew randomly over the unit: this would be an example of a stand exhibiting no spatial autocorrelation.

This never occurs in a managed site and usually trees of similar characteristics (volume, health, species, etc) grow near each other.

What does this mean?

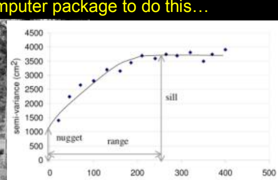
Problem: If our plots are too close together we will likely measure essentially the same measurements (as they will be too alike) – these measures will not be independent.

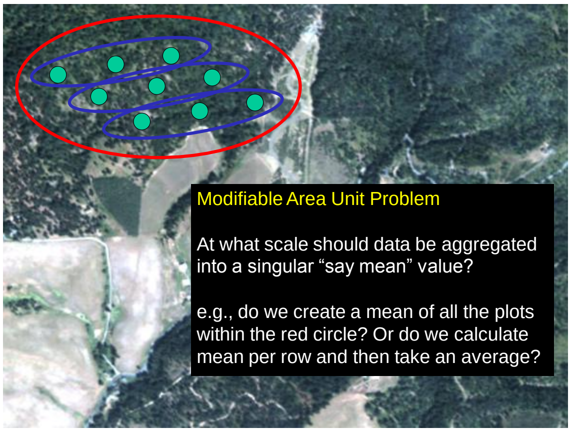
Distance Between Samples: Semi-variograms

To overcome this problem we use semi-variograms to tell us how far apart to place plots so they are independent.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (x_{(\alpha+h)} - x_{\alpha})^2$$

N is number of pairs: For every n observations there are n*(n-1)/2 pairs… clearly use a computer package to do this…





Modifiable Area Unit Problem

At what scale should data be aggregated into a singular "say mean" value?

e.g., do we create a mean of all the plots within the red circle? Or do we calculate mean per row and then take an average?

Distance Between Samples: MAUP

Data can be aggregated in different ways

10	15	5
5	10	15
5	10	5

n = 9 Mean = 8.88

Mean of the means		
6.6	11.6	6.6

n = 3 Mean = 8.26

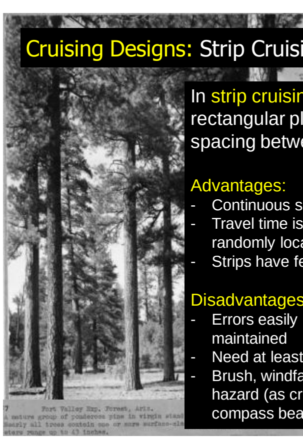
10	10	
5	12.5	
	10	5

n = 6 Mean = 10.83

Importantly, the mean of all samples does not necessarily equal the mean of the means.

It is very important that data is collected at scales that capture the variability within the data.

Cruising Designs: Strip Cruising



In strip cruising, parallel strips (long thin rectangular plots) are used where the spacing between the steps are constant

Advantages:
- Continuous sample
- Travel time is low (as compared to visiting randomly located plots)
- Strips have fewer boundary trees

Disadvantages:
- Errors easily introduced if correct strip width is not maintained
- Need at least 2 people
- Brush, windfalls, and surface debris are more of a hazard (as cruisers must cruise along a fixed compass bearing)

Cruising Designs: Strip Cruising

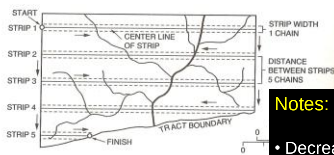


FIGURE 10-1
Diagrammatic plan for a 20 percent systematic strip cruise. Sample strips spaced at regular intervals of 5 chains.

TABLE 10-1
EXAMPLE OF CRUISING INTENSITIES FOR 1-CHAIN SAMPLE STRIP WIDTHS

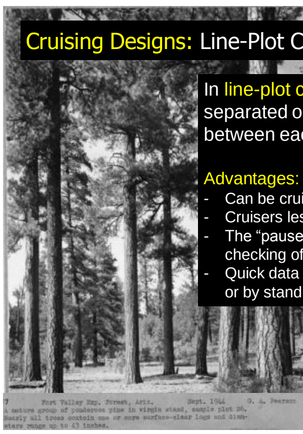
Distance between strip centerlines			
ft	chains	No. of strips per "forty"	Nominal cruise percent
1,320	20	1	5
660	10	2	10
330	5	4	20
165	2½	8	40

Notes:

- Decrease width in young stands with high stem count
- Increase width in more scattered high value timber
- Cross drainages at right angles
- In theory all timber conditions are samples and a representative sample taken
- Sampling intensity = (W/D)*100

Source: Avery and Burkhart Chapter 10

Cruising Designs: Line-Plot Cruising



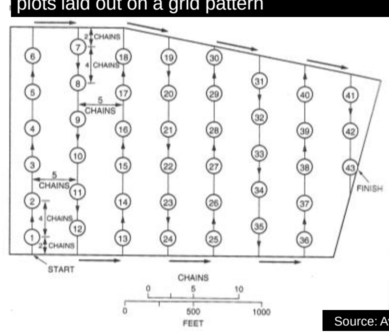
In line-plot cruising the plots are equally separated on each line, with equal spacing between each line – i.e. a grid

Advantages:

- Can be cruised by 1 person
- Cruisers less hindered by brush and windfall
- The "pause" at plot center enables better checking of borderline trees
- Quick data summaries can be obtained per plot or by stand / condition classes.

Cruising Designs: Line-Plot Cruising

In line-plot cruising a systematic tally of timber is taken from plots laid out on a grid pattern



Source: Avery and Burkhart Chapter 10

Cruising Designs: Fixed Area Plots

Notes:

- Circular 1/10 ac plots commonly used for timber tallies
- 1/25 ac for pulpwood trees
- 1/100 ac for regeneration counts

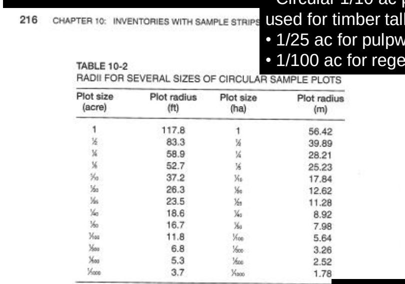

216 CHAPTER 10: INVENTORIES WITH SAMPLE STRIPS

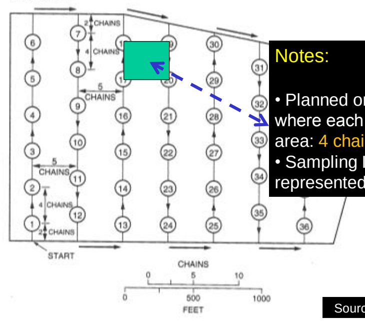
TABLE 10-2
RADII FOR SEVERAL SIZES OF CIRCULAR SAMPLE PLOTS

Plot size (acre)	Plot radius (ft)	Plot size (ha)	Plot radius (m)
1	117.8	1	56.42
1/2	83.3	1/2	39.89
1/4	58.9	1/4	28.21
1/5	52.7	1/5	25.23
1/10	37.2	1/10	17.84
1/20	26.3	1/20	12.62
1/25	23.5	1/25	11.28
1/40	18.6	1/40	8.92
1/50	16.7	1/50	7.98
1/100	11.8	1/100	5.64
1/300	6.8	1/300	3.26
1/500	5.3	1/500	2.52
1/1000	3.7	1/1000	1.78

Source: Avery and Burkhart Chapter 10

Cruising Designs: Plot Spacing

FIGURE 10-2
Diagrammatic plan for a 10 percent systematic line-plot cruise utilizing ⅕-acre circular sampling units.

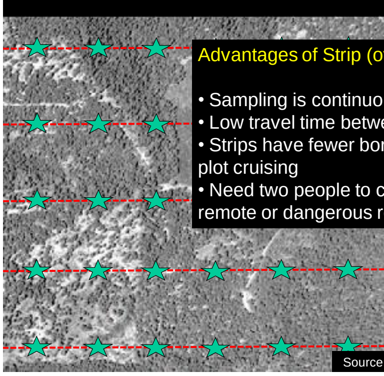


Notes:

- Planned on a % cruise basis, where each plot "represents" an area: 4 chains x 5 chains = 2 ac
- Sampling Intensity = plot size / represented area = 0.2/2 = 10%

Source: Avery and Burkhart Chapter 10

Cruising Designs: Fixed Area Plots

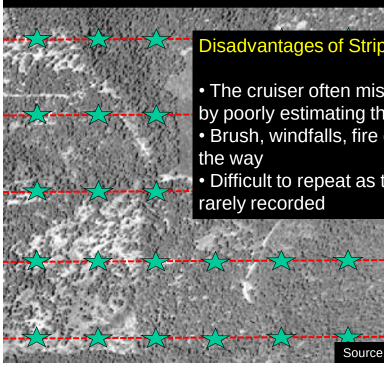


Advantages of Strip (over Plot) Cruising:

- Sampling is continuous
- Low travel time between plots
- Strips have fewer borderline trees than plot cruising
- Need two people to cruise – so safer in remote or dangerous regions

Source: Avery and Burkhart Chapter 10

Cruising Designs: Fixed Area Plots

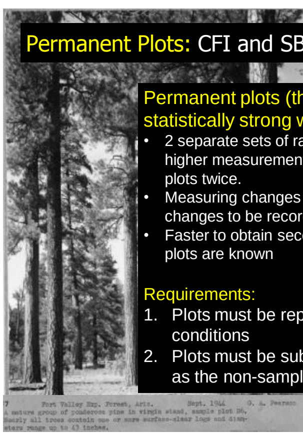


Disadvantages of Strip (over Plot) Cruising:

- The cruiser often misses borderline trees by poorly estimating the width of the strip
- Brush, windfalls, fire damage etc get in the way
- Difficult to repeat as the center line is rarely recorded

Source: Avery and Burkhart Chapter 10

Permanent Plots: CFI and SBI



Permanent plots (that are re-measured) provide statistically strong ways to evaluate changes

- 2 separate sets of random samples in a stand will have higher measurement errors that measuring the same plots twice.
- Measuring changes in the same place allows actual changes to be recorded
- Faster to obtain second inventory as general location of plots are known

Requirements:

1. Plots must be representative of stand / forest conditions
2. Plots must be subjected to the same treatments as the non-sampled parts of the forest

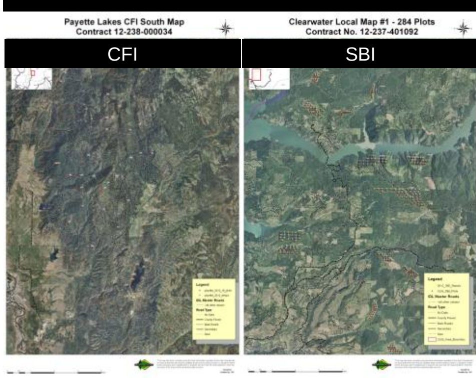
Permanent Plots: CFI and SBI

Continuous Forest Inventory (CFI):
Has been around since the 1950s. Involves permanent fixed plot centers – the same trees are measured over time to obtain growth and yield estimates. CFI is good at estimating volumes at the Area and State-wide endowment (IDL) level.

Stand Based Inventory (SBI):
Managers recognized the need for an inventory that provides reliable estimates of volume at the stand level. SBI plots are installed by stand at a level of 1 plot per 5 acres. This provides resource managers with better volume estimates and provides many of our models and simulators better data to model.

Permanent Plots: CFI and SBI

Permanent Plots: CFI and SBI

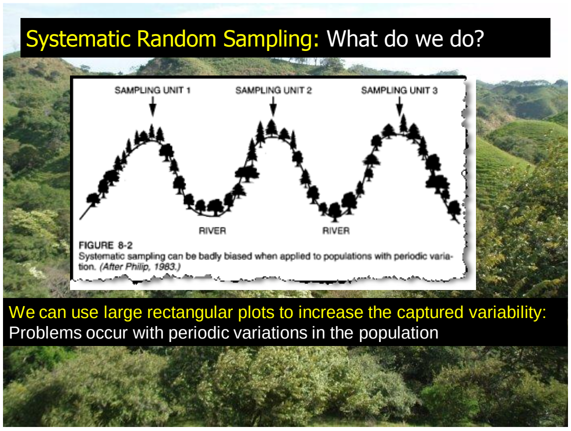


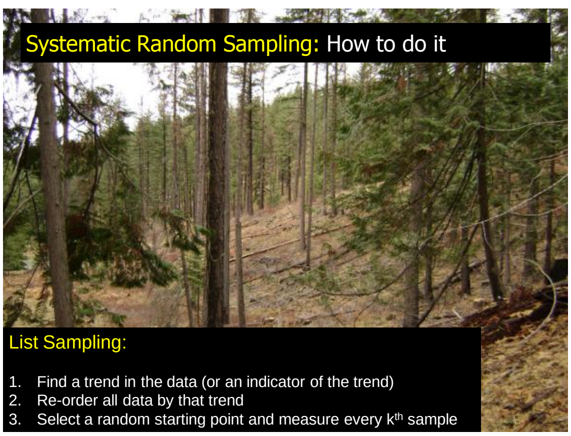
Permanent Plots: CFI and SBI

Notes on CFI:

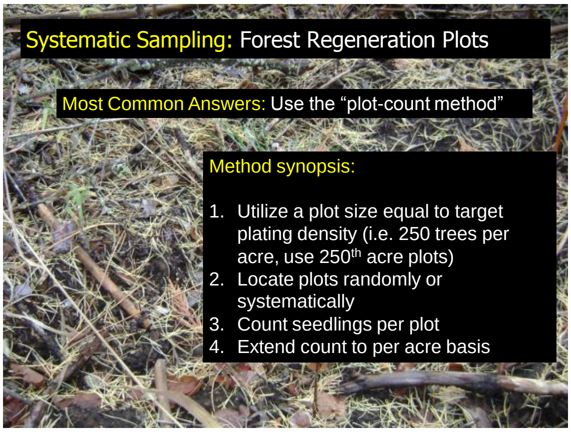
Often CFI is applied to a rigid systematic grid, where each plot represents equal proportion of the total forest area.

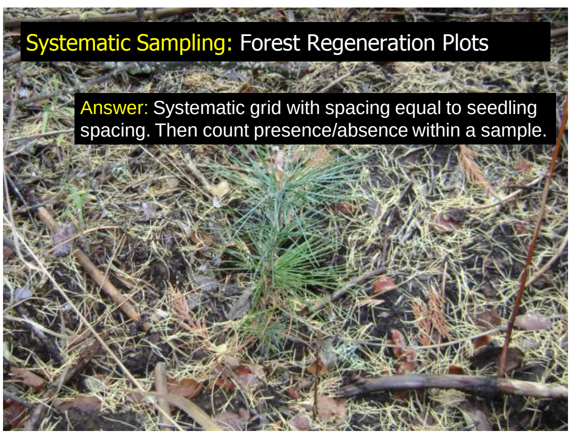
BUT systematic grids are inflexible to changing management priorities.

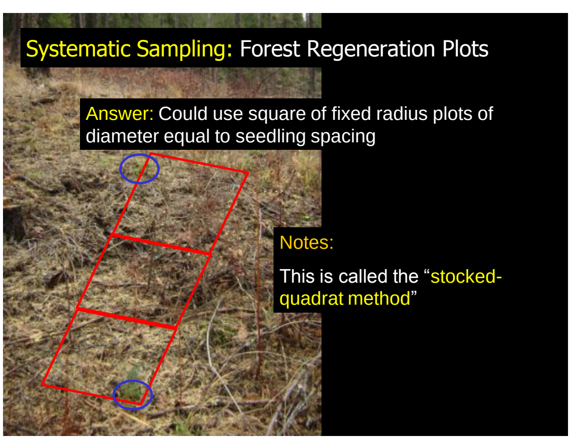

Systematic Random Sampling: What do we do?
SAMPLING UNIT 1
SAMPLING UNIT 2
SAMPLING UNIT 3
RIVER
RIVER
FIGURE 8-2
Systematic sampling can be badly biased when applied to populations with periodic variation. (After Philip, 1983.)
We can use large rectangular plots to increase the captured variability:
Problems occur with periodic variations in the population

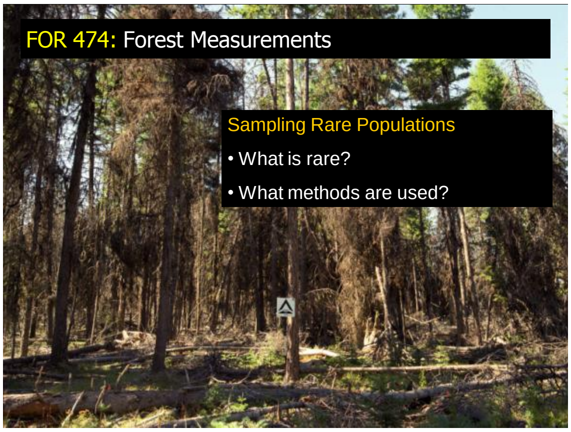

Systematic Random Sampling: How to do it
List Sampling:
1. Find a trend in the data (or an indicator of the trend)
2. Re-order all data by that trend
3. Select a random starting point and measure every k^th sample

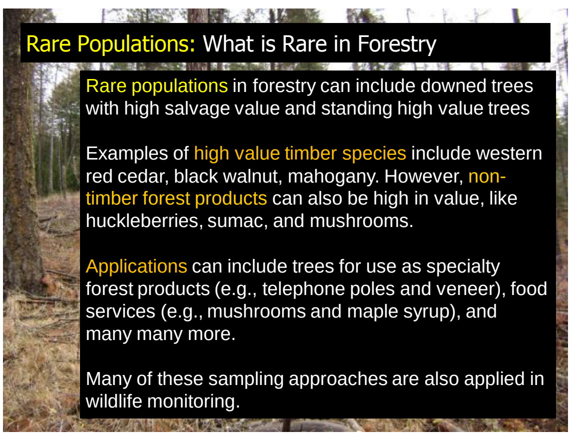

Systematic Sampling: Forest Regeneration Plots
Question: How would you sample the stocking rate in this regeneration stand?

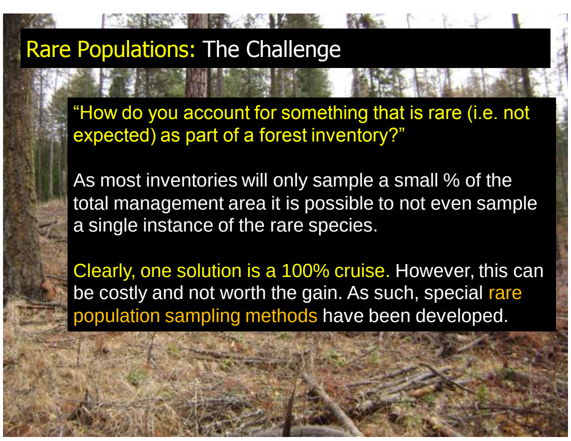

Systematic Sampling: Forest Regeneration Plots
Most Common Answers: Use the "plot-count method"
Method synopsis:
1. Utilize a plot size equal to target plating density (i.e. 250 trees per acre, use 250th acre plots)
2. Locate plots randomly or systematically
3. Count seedlings per plot
4. Extend count to per acre basis


Systematic Sampling: Forest Regeneration Plots
Answer: Systematic grid with spacing equal to seedling spacing. Then count presence/absence within a sample.


Systematic Sampling: Forest Regeneration Plots
Answer: Could use square of fixed radius plots of diameter equal to seedling spacing
Notes:
This is called the "stocked-quadrat method"


FOR 474: Forest Measurements
Sampling Rare Populations
• What is rare?
• What methods are used?


Rare Populations: What is Rare in Forestry
Rare populations in forestry can include downed trees with high salvage value and standing high value trees
Examples of high value timber species include western red cedar, black walnut, mahogany. However, non-timber forest products can also be high in value, like huckleberries, sumac, and mushrooms.
Applications can include trees for use as specialty forest products (e.g., telephone poles and veneer), food services (e.g., mushrooms and maple syrup), and many many more.
Many of these sampling approaches are also applied in wildlife monitoring.


Rare Populations: The Challenge
"How do you account for something that is rare (i.e. not expected) as part of a forest inventory?"
As most inventories will only sample a small % of the total management area it is possible to not even sample a single instance of the rare species.
Clearly, one solution is a 100% cruise. However, this can be costly and not worth the gain. As such, special rare population sampling methods have been developed.

Rare Populations: Use Systematic Strip Cruising

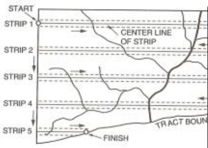


FIGURE 10-1
Diagrammatic plan for a 20 percent systematic strip cruis
spaced at regular intervals of 5 chains.

TABLE 10-1
EXAMPLE OF CRUISING INTENSIT
1-CHAIN SAMPLE STRIP WIDTHS

Distance between strip centerlines			
ft	chains	No. of strips per "forty"	Nominal cruise percent
1,320	20	1	5
660	10	2	10
330	5	4	20
165	2½	8	40

Approach covers very large areas but is inefficient when considering live trees.

When used for rare populations, auto correlation issues are minimized as samples are typically very far apart.

Measure all rare species (or objects) in the strip. If measuring downed logs, most samples include if center intersects the strip

Source: Avery and Burkhart Chapter 10

Rare Populations: Use Systematic Strip Cruising

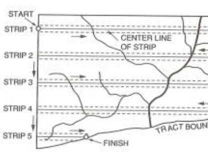


FIGURE 10-1
Diagrammatic plan for a 20 percent systematic strip cruis
spaced at regular intervals of 5 chains.

TABLE 10-1
EXAMPLE OF CRUISING INTENSITIES FOR
1-CHAIN SAMPLE STRIP WIDTHS

Distance between strip centerlines			
ft	chains	No. of strips per "forty"	Nominal cruise percent
1,320	20	1	5
660	10	2	10
330	5	4	20
165	2½	8	40

Problem occurs for rare populations - strips have different areas and lengths – need to correct for true areas

$$Volume = \frac{\sum_{i=1}^{n} V_i}{\sum_{i=1}^{n} A_i} A_T$$

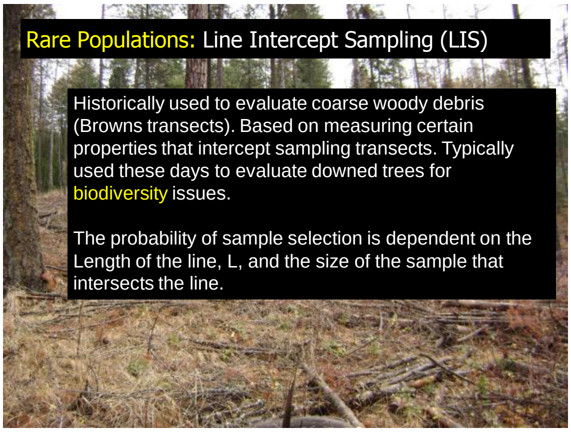
Source: Avery and Burkhart Chapter 10

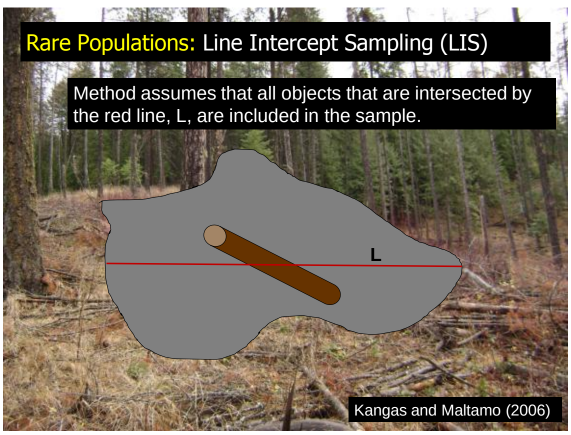
Rare Populations: Line Intercept Sampling (LIS)

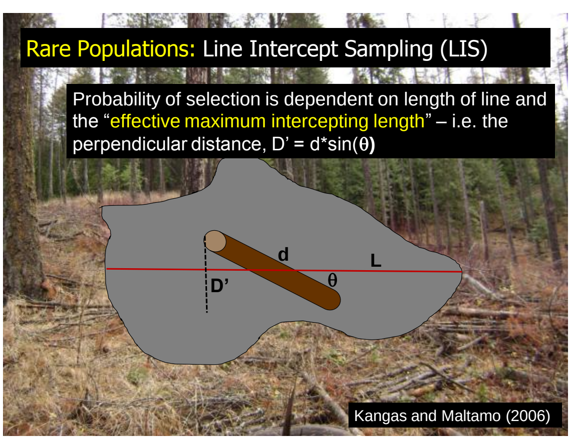
Developed by Cranfield (1941) and has been applied to numerous applications including:

- Estimating slash and fuel following logging
- Fuels
- Biodiversity
- Coarse woody debris
- Forest canopy cover
- Estimating wildlife populations

Generally unbiased and considered an improvement over quadrat sampling


Rare Populations: Line Intercept Sampling (LIS)
Historically used to evaluate coarse woody debris (Browns transects). Based on measuring certain properties that intercept sampling transects. Typically used these days to evaluate downed trees for biodiversity issues.
The probability of sample selection is dependent on the Length of the line, L, and the size of the sample that intersects the line.


Rare Populations: Line Intercept Sampling (LIS)
Method assumes that all objects that are intersected by the red line, L, are included in the sample.
L
Kangas and Maltamo (2006)


Rare Populations: Line Intercept Sampling (LIS)
Probability of selection is dependent on length of line and the "effective maximum intercepting length" – i.e. the perpendicular distance, D' = d*sin(θ)
d
L
θ
D'
Kangas and Maltamo (2006)

Rare Populations: Line Intercept Sampling (LIS)

If we use the Huber's equation for tree volume *(logs assumed to be cylinders of diameter at midpoint)* the total volume of coarse woody debris (m^3/ha) is given by:

$$Volume = (T) = \frac{\pi^2}{8L}\sum_{i=1}^{m} d_i^2$$

$$Var(T) = \frac{\sum_{i=1}^{n} L_j (T_j - T)^2}{(n-1)\sum_{j=1}^{n} L_j}$$

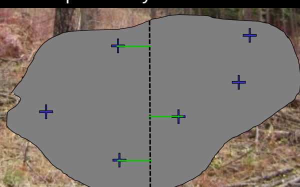
The variance is determined from the variability between sampling lines. T_j = m^3/ha of line j & T = total volume per hectare. These equations assume that lines are in random directions or trees fell in random directions.

Kangas and Maltamo (2006)

Rare Populations: Line Transect Sampling

Developed for estimating the density of wildlife.

Assumes randomly placed lines and randomly located wildlife, where as the distance of wildlife from the line increases the probability of observation decreases.



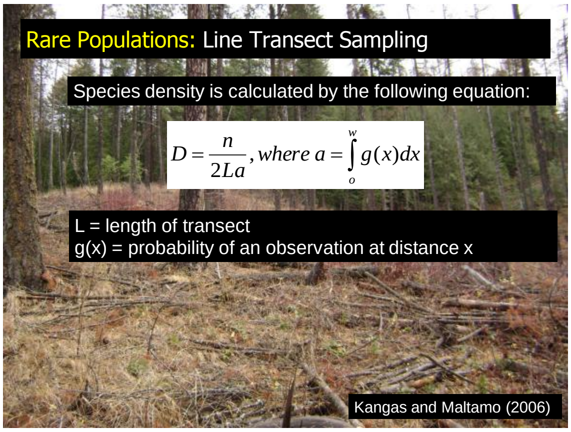
Kangas and Maltamo (2006)

Rare Populations: Line Transect Sampling

The method makes three assumptions:

- The probability of observation on the line is 100%
- You can't view the objects more than once (i.e. they don't move once they are spotted)
- Observations are independent events (likely invalid for herding species)

Kangas and Maltamo (2006)


Rare Populations: Line Transect Sampling
Species density is calculated by the following equation:
$D = \frac{n}{2La}, where\ a = \int_o^w g(x)dx$
L = length of transect
g(x) = probability of an observation at distance x
Kangas and Maltamo (2006)