Chapter 5

ANALYSIS

The purpose of this chapter is to review the main steps in analyzing basic line transect surveys using the Wyoming Technique with the computer program, DISTANCE (Buckland et al. 1993, Laake et al. 1996). DISTANCE is a comprehensive analysis program for analyzing distance sampling surveys, of which line transects are one type. It represents a substantial evolution from its predecessor, TRANSECT (Laake et al. 1979, Burnham et al. 1980). As discussed below, I do not recommend using the TRANSECT program (Laake et al. 1979) that was described in the *Guidelines* by Johnson and Lindzey (1990). Some modifications to the Wyoming Technique are required in order to analyze these aerial survey data using other programs like LINETRAN (Gates 1979 and 1980), SIZETRAN (Drummer 1991) or TRANSAN (Routledge and Fyfe 1992). In my opinion, DISTANCE is the most comprehensive analysis package available for analyzing distance sampling data and is most amenable to analyzing data collected using the Wyoming Technique.

I only provide an overview of the basic procedures for analyzing survey data and interpreting results here. Most of the emphasis is on organizing, screening, and preparing data for analysis with DISTANCE. This chapter also provides some basic instruction on running an analysis, examining output and interpreting results of basic surveys. Users should view this section as an introduction to analysis and not as the final word or an analysis cookbook. The DISTANCE User's Guide (Laake et al. 1996) contains more complete instruction on running the program and what various options will do. **The comprehensive text on distance sampling by Buckland et al. (1993) provides more detail on the rationale and appropriateness of using various procedures implemented in DISTANCE.** I strongly recommend that analysts review Chapters 2 and 4 of Buckland et al. (1993). Advanced users may find the underlying statistical theory described in Chapter 3 of that book to be helpful in explaining the estimation process, criteria used, and other concepts related to analysis.

Novice analysts are strongly encouraged to seek the advice of biologists and statisticians who have considerable experience analyzing line transect data using the Wyoming Technique. Peer review is most helpful in performing objective analyses. The successful estimation of population abundance using line transects depends on the proper design and conduct of surveys. As with all population estimation techniques, the ability to compensate for poor survey design and implementation is limited at the analysis phase.

The analysis of line transect data is an iterative process. In most cases, a preliminary analysis is performed to explore patterns in the data, identify problems, and screen various estimators. Based on initial runs, modifications are made to the analysis to improve estimates. Additional options such as bootstrapping may be performed in subsequent analyses depending upon initial results, and the requirements of the survey. Rarely will an initial, single analysis be adequate.
This chapter focuses on selected options in DISTANCE that are of interest to Wyoming's aerial line transect surveys for pronghorn. The basic elements for the analysis of line transect data using the Wyoming Technique involve the following:

- perpendicular distances grouped by distance intervals (adjusted for height AGL),
- clustered populations corrected for size-bias,
- objective model selection criteria, and
- calculation of population totals as well as densities with their associated measures of precision.

Analysts should **use a great deal of caution and objectivity throughout the analysis process**. As much potential exists during the analysis as at any other phase of the survey for subjective decisions to influence results. Therefore, it is important to document the rationale used in selecting one procedure or estimator over another, or for deviating from standard procedures. The analysis should be driven on the quality of the estimates rather than getting the "right" (i.e., preconceived) answer. However, seemingly unreasonable estimates should be scrutinized along with other data on which the reasonableness of the line transect estimates are compared (Guenzel 1994). In some cases, other data (e.g., trend counts, POP-II simulations, etc.) have been found to contain previously unrecognized errors.

PRELIMINARY SCREENING AND USING THE FLIGHT REPORT

Data from the survey should be screened and organized prior to running the analysis with DISTANCE. This review helps alert users to problems and patterns in the data and also helps in selecting candidate models for estimation. The following steps can help guide the analysis.

Review the Survey Procedures and Conditions

Prior to the analysis, go through the following steps:

- Review the survey conditions and how these might influence the analysis,
- Review the performance and experience of the pilot and observers (be sure to ask how they thought they did on the survey and any problems or factors they think might influence the outcome),
- Note any deviations from standard survey procedures (e.g., using one observer, recording data manually, navigational problems),
- Review the desired sample size for which the survey was designed,
- Note which criterion was used for estimating cluster sizes (*ad hoc* vs. letting DISTANCE handle cluster size bias), and
- Review how height AGL will be calculated (e.g., was it recorded for each observation, or at random or systematically spaced points along lines).

Obtain Areas for Herd Units and Strata

Estimate the area for which the population estimate is to apply. This should have already been performed during the design phase. The area may be the total in square miles for the herd unit, total occupied habitat (TOH), or some other region. Be sure to estimate areas to be excluded and subtract them from the appropriate totals. For stratified surveys, estimate the appropriate areas for each stratum. Areas can be estimated using a compensating planimeter, dot grid, planimetric map (counting sections), or GIS.

Screen the Flight Report and Summarize Data for Entry

Providers of Class 1 line transect services for the Wyoming Game and Fish Department (Appendix I) provide a flight report summarizing line lengths, number of clusters observed by distance band, total number of pronghorn observed by distance band, and other data. The flight report should be screened and then summarized in a format conducive to entering data into DISTANCE.

Each cluster needs to be tallied by line and the corresponding midpoint of the perpendicular distance band (corrected for height AGL) in which it was observed.

- **Summary Data:** From the flight report, obtain the estimated total line length surveyed (in miles), and the total number of clusters and individuals observed by distance band.
- **Examine the Sample Size:** The tally of clusters observed during the survey is the sample size. Ideally, this should meet or exceed the desired sample size (number of clusters) for which the survey was designed. Hopefully, this is at least 200 clusters. If the sample size is less than 100, consider the effects of the small sample on results as the analysis proceeds.
- **Estimate the Average Survey Height AGL:** Do not use the average height AGL reported on the flight report. Those are usually calculated differently from what is needed for our analysis. Calculate the average height AGL from the positions where pronghorn were observed along the transects. A position is where one or more clusters were observed at the same location. These are usually reported on the same row on the flight report. Do not calculate the average height AGL from each cluster (if you saw two clusters at the same location, you didn't really survey twice as much at that height). Do not include radar altimeter readings from transect end-points. Often these coincide with prominent terrain features and may not be representative for the survey. Random or systematic radar altimeter readings taken along lines can be used for this average. The midpoint of the distance band for each observation is normally adjusted for the actual height AGL. Those are used as perpendicular distances in the analysis.
- **Calculate the Adjusted Average Distance Band Cut-points:** Adjust each distance band cut-point (on the ground) for the average height AGL from the previous step. Appendix III contains tables of adjusted cut-points for heights AGL ranging from 100 to 650 feet or these may be calculated by hand. First, calculate the ratio of the average height AGL (in feet) to the nominal height AGL of 300 feet (what the struts were marked for) for the survey. Multiply this ratio by each of the outer limits for each distance band (e.g., 25, 50, 100 and 200 meters). This gives an estimate for the mean of the actual distance bands surveyed (these will also be used later in the analysis).

  The general equation for adjusting distance band cut-points for actual mean survey height AGL is as follows:

$$CP_{ai} = CP_{ni} * (HT_a/HT_n) \qquad \text{(Eq. 5.1)}$$

where CP_i	=	cut-point for the ith distance band
HT	=	height AGL
a	=	actual (corrected for AGL)
n	=	nominal (if the survey were flown at 300' AGL)

For example, suppose that the survey averaged 330 feet AGL (i.e., HT_a/HT_n = 330'/300' = 1.1). The corrected cut-points would be calculated as follows:

CP_{a0}	(i.e., the line)	=	0 m * 1.1	=	0.0 m
CP_{aA}	(i.e., outer limit of "A")	=	25 m * 1.1	=	27.5 m
CP_{aB}	(i.e., outer limit of "B")	=	50 m * 1.1	=	55.0 m
CP_{aC}	(i.e., outer limit of "C")	=	100 m * 1.1	=	110.0 m
CP_{aD}	(i.e., outer limit of "D")	=	200 m * 1.1	=	220.0 m

- **Check Line Lengths and Straightness:** Transect lengths presented on most flight reports assume that the transect is straight. The lengths are calculated as the distance from end-point to end-point. This may not always be accurate. Ideally, line lengths should be calculated from a series of locations along the flight path. To approximate this, the locations along the transect where pronghorn were seen along with the end-points can be used to estimate the actual line length (see Equation 5.2). For the following example, let the UTM Easting and Northing coordinates be designated X and Y, respectively. The distance from each point to the next along the transect can be calculated using the Pythagorean Theorem. Let the distance between the ith and the ith+1 points along the line be designated as d. You can calculate d as follows:

$$d = \sqrt{[(X_i - X_{i+1})^2 + (Y_i - Y_{i+1})^2]} \qquad \text{(Eq. 5.2)}$$

  The total (corrected) line length is the sum of the distances between points.

  One measure of how straight transect lines are is a straightness index (Batschelet 1981). This is the ratio of the length calculated from the end-points to the length estimated by summing all the individual observation-to-observation (or end-point) distances. From a practical sense, the inverse of this ratio is easier to understand since it is impossible for the actual line flown to be shorter. The closer the ratio is to 1.0, the straighter the line. Higher ratios indicate substantial departure from a straight line. Table 5.1 shows an example of the calculation of actual line length and the straightness index. Line lengths should be checked, particularly when density or total estimates are outside their expected range of values. A listing of accurate line lengths should be prepared. In some cases, this can be generated directly from the flight report. Where questions arise, actual line lengths should be approximated using the procedure described above.
- **Plot the Histogram of the Observed Number of Clusters by Adjusted Distance Band:** Make sure you calculate a histogram and not a bar graph. In histograms, the area under each bar or column is proportional to the number it represents. In bar graphs, the height of the bar,

Table 5.1. Example analysis of straightness and other diagnostics for a transect.

POSITION ALONG TRANSECT	UTM Easting	UTM Northing	DISTANCE FROM PREVIOUS POSITION (km)	CUMULATIVE DISTANCE FROM START OF LINE, l (km)	NUMBER OF CLUSTERS (n)	CUMULATIVE ENCOUNTER RATE (n/l)
Start of Transect	4574591.606	452429.912	0.000	0.000	0	0.000
1st Observation	4579944.866	452585.111	4.456	4.456	1	0.224
2nd Observation	4580505.576	452597.532	0.561	5.017	1	0.399
3rd Observation	4586376.742	452672.903	5.872	10.889	1	0.276
4th Observation	4588963.308	452659.344	2.587	13.476	2	0.371
5th Observation	4589690.243	452671.077	0.727	14.203	1	0.426
6th Observation	4590424.279	452669.418	0.734	14.937	1	0.469
7th Observation	4592659.245	452693.198	2.235	17.172	3	0.582
8th Observation	4593316.552	452690.574	0.657	17.829	1	0.617
End of Transect	4594464.383	452684.702	1.148	18.977	0	0.580
				TOTAL:	11	**0.580**
		TOTAL	**(segments)**	**18,977**		
		LENGTH	**(start-stop)**	**18,974**		
		STRAIGHTNESS	**INDEX**	**1.0001**		

not its area, is proportional to the value it represents. This is an important distinction for line transect sampling. Figure 5.1 illustrates the difference. Assume for simplicity that the average survey height AGL was 300 feet so there is no adjustment needed for height. In this example, suppose we counted 20 clusters in the A band, 18 in the B, 24 in the C and 24 in the D band. A bar graph of these data would resemble Figure 5.1 A. Line transect interpretations based on that bar graph might mislead us into thinking that we did not meet line transect assumptions since it appears that we saw relatively more animals away from the line. The histogram shown in Figure 5.1 B correctly takes into account the area for each distance band in which the clusters were seen. There is twice the area in the C band as in A or B, and twice the area of C in the D band (four times the area of A or B). Detectability declines with distance in the example shown in Figure 5.1.

Use the following steps to calculate a histogram for data obtained using the Wyoming Technique:

1. Plot the distance band cut-points along the x-axis (e.g., 0, 25, 50, 100 and 200 m) adjusted for actual height AGL.
2. Calculate the relative height (observed number of clusters) for each distance band. The heights for the A and B bands are the actual counts. The relative height of the C band is half the number of clusters observed in C because it has twice the area of A or B. The relative height of D is one-fourth of the number of clusters observed since it has four times the area of A or B.
3. Plot the heights for each distance band (see Fig. 5.1).

- **Identify Potential Estimators:** Interpretations based on the slope of this histogram should be reflective of the actual survey. The slope of the histogram may suggest potential estimation models to try in DISTANCE, and/or may suggest that assumptions were violated (see Fig. 5.2). For histograms such as in Figures 5.2 A and B, the **uniform** (relatively flat shoulder) or **half normal** (right half of a bell-shaped curve) keys with cosine adjustment terms would be reasonable. Figure 5.2 C suggests that an assumption was not met (see next section). An estimator based on the **negative exponential** (rapidly declining) or **hazard** (flat shoulder near the line then rapidly declining) key would be reasonable for a histogram such as Figure 5.2 D where detectability declines rapidly with distance.
- **Check Assumptions:** One of the critical assumptions of line transect sampling is that all pronghorn on the line are seen. Examine the histogram from the above step to see if the graph indicates that detectability declines with distance. If it does not (e.g., Fig. 5.2 C), then you may have a problem. Many problems are not insurmountable, but may require additional assumptions or correction factors. If the A band is substantially lower than in B or C, then this indicates that an assumption was violated. Two possibilities are that (1) observers missed animals on the line, or (2) animals moved away before being detected. You will have to decide what you think happened in order to compensate for the problem. It is common for new or poorly trained observers to miss animals near the line because they have not developed the discipline to search on and near the line intently. That is why training and practice are so critical. With small sample sizes, assigning one or two animals to a particular distance band may have a stronger influence on the shape of the histogram. Some of this variation may be due to chance, however (see Fig. 5.2 B). That is one reason to strive for larger sample sizes.

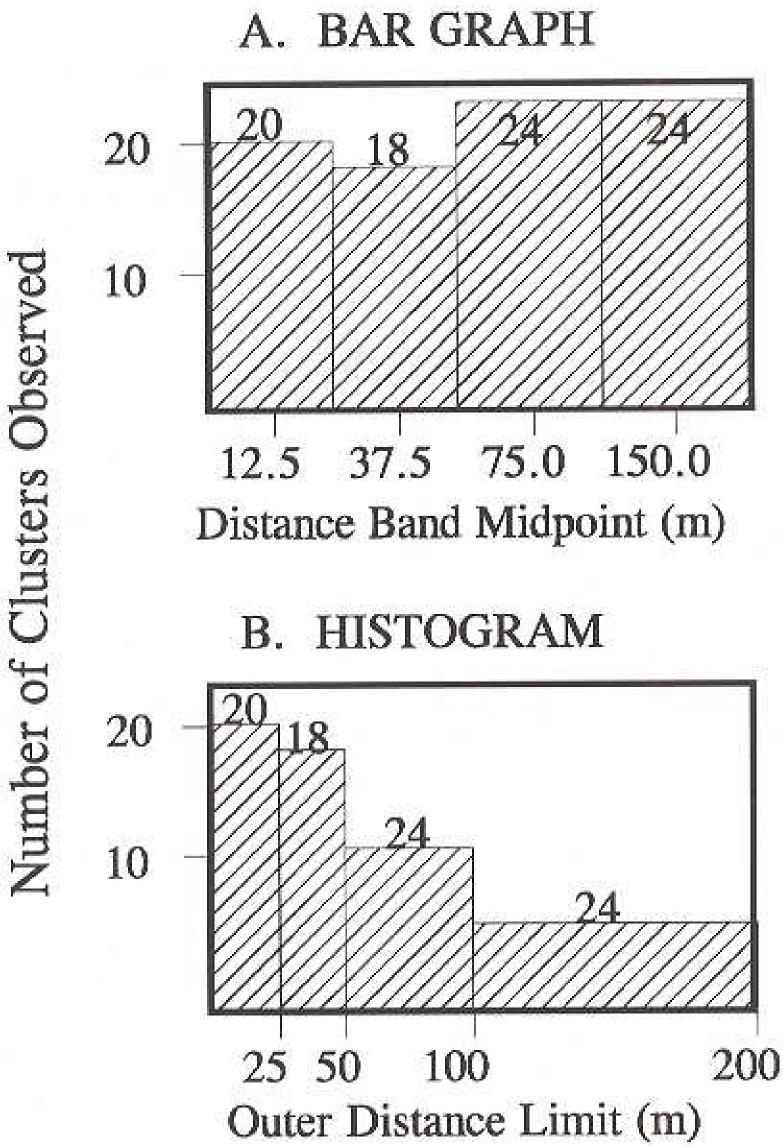


Figure 5.1. Comparison of line transect data displayed as a bar graph (A) and as a histogram (B). Proper analysis and interpretation of line transect surveys is based upon the histogram of observed distances.

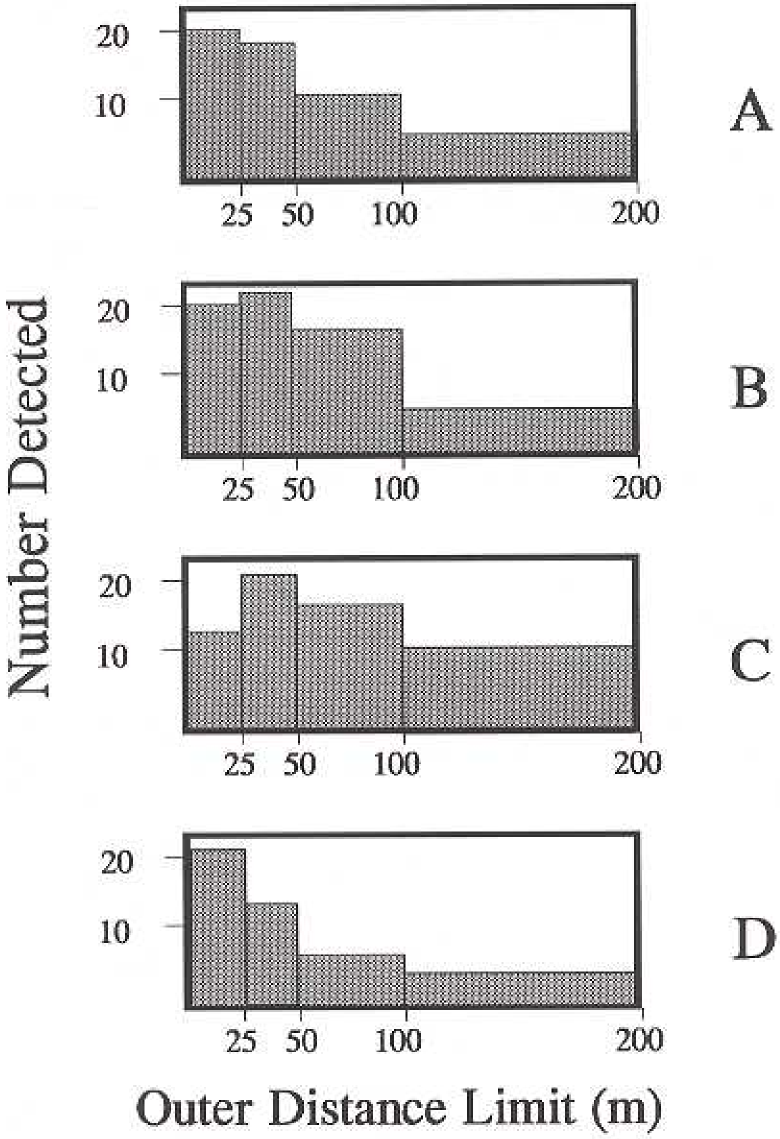


Figure 5.2. Sample line transect histograms showing a range of shapes. Histograms (A) and (B) indicate detectability is relatively flat near the line and variously declines with distance. These are preferable to shapes like (C) and (D). A slightly lower inner band as in (B) may be due to chance or minor bias. More serious bias such as missing animals in the inner band or animals moving before being detected may appear as in (C). Additional assumptions are often required to correct such problems. Histogram (D) indicates detection declines rapidly with distance. These are often difficult to model (*see text for more explanation*).

- **<u>Calculate Quick & Dirty Estimate:</u>** The Quick & Dirty method is a way to obtain a ballpark estimate from the survey prior to actually running the analysis with DISTANCE. This approach assumes that all animals in the first interval are seen. To obtain this estimate, you calculate density as a very narrow strip transect. Divide the number of pronghorn (individuals) seen in the A band for the entire survey (from both observers if using two) by the area surveyed. The area surveyed in the A band is the total line length times two times the adjusted width (for height AGL) of the A band. Convert the width from meters to miles by dividing meters by 1,609. Multiply the density by the area to which the estimate applies to get a ballpark estimate of the total. This may help identify problems if estimates from DISTANCE are grossly different since the Quick & Dirty estimate should be close to DISTANCE's estimates. In rare cases where the A band is lower than the B band, you might calculate the density for both the A and B bands. The Quick & Dirty method is helpful in gauging the acceptability of different line transect estimates using various models and in selecting a final model. The following example illustrates the steps in the procedure:

<u>Centennial Pronghorn - Example:</u>

A line transect survey was conducted in the Centennial Pronghorn Herd in 1993. The herd occupies an area (A) of 1,153 mi^2 of habitat The following results were obtained:

Total length of transect surveyed, L	=	334 mi.
Total number of pronghorn counted in the "A" band, n	=	104
Average height AGL for the survey, HT_a	=	343.2'

<u>*Step 1.*</u> *Calculate the area of the strip, a_s, surveyed within the "A" band (adjusted for altitude):*

$$a_s = L * w_A * (HT_a/HT_n) * 2 \qquad \text{(Eq. 5.3)}$$

where

L	=	length of transects
w_A	=	width of "A" band
HT	=	height AGL
a	=	actual
n	=	nominal

$$a_s = 334 \text{ mi} * (25 \text{ m} * 1 \text{ mi}/1609 \text{ m}) * (342.4'/300') * 2$$
$$= 11.87 \text{ mi}^2$$

<u>*Step 2.*</u> *Calculate observed density, d:*

$$d = n/a_s \qquad \text{(Eq. 5.4)}$$

$$d = 104 \text{ pronghorn}/11.87 \text{ mi}^2$$
$$= 8.76 \text{ pronghorn/mi}^2$$

Step 3. *Calculate the population total, N:*

$$N = A * d \qquad \text{(Eq. 5.5)}$$

$$N = 1{,}153 \text{ mi}^2 * 8.76 \text{ pronghorn/mi}^2$$
$$= 10{,}099 \text{ pronghorn}$$

The actual line transect estimate (using a negative exponential estimator) for this survey was 9,932 pronghorn. The Quick & Dirty estimate is within 2% of the line transect estimate for the survey.

- **Evaluate Cluster Size Bias:** It is often helpful to know prior to running the analysis if the data suggest cluster size bias (i.e., observers tend to see larger clusters farther away than smaller clusters or single animals). If cluster size bias is not accounted for, the population may be overestimated. The magnitude of the apparent bias can suggest options for analysis, depending upon how the larger clusters were handled during the survey (i.e., counted or truncated distance limits for cluster size estimation; see Chapters 3 and 4 of this manual). To check for size bias, calculate the mean cluster size for each distance band (i.e., number of pronghorn observed/number of clusters observed in that distance band). These can be plotted against the distance band mid-points to see if there is evidence of mean cluster size increasing with distance (see Fig. 5.3). If the mean cluster size increases with distance, your data probably have cluster size bias since the arithmetic mean cluster size will be weighted by seeing more clusters in the outer bands. If there is evidence of cluster size bias, your analysis with DISTANCE should probably use one of the corrections. This approach doesn't work as well where the *ad hoc* procedures are used.

Prepare Data for Entry into DISTANCE

After the preliminary diagnostics have been performed, the data set needs to be converted to a format that can be analyzed in DISTANCE. Although data collected with the Wyoming Technique are "grouped" by distance band, the observations will be entered into the program as if they were ungrouped data and then later analyzed as grouped data using adjusted cut-points. This allows for the correction of cluster size bias and other analytical improvements.

Work through the flight report by transect and observation, listing the data in a format similar to how it will be entered in DISTANCE. For each transect (and stratum if more than one), compile the line length in miles, a label if desired (e.g., "Line 1 - W104° 47' N"), and then list the observations pairing each perpendicular distance (adjusted midpoint of distance band) with the corresponding cluster size. If more than one cluster was observed at a given location, list it separately. For each transect, there should be two columns of data: distance and cluster size. If the data are from a stratified survey, be sure to partition the lines and observations by stratum and provide an estimate of the area of that stratum (in mi^2).

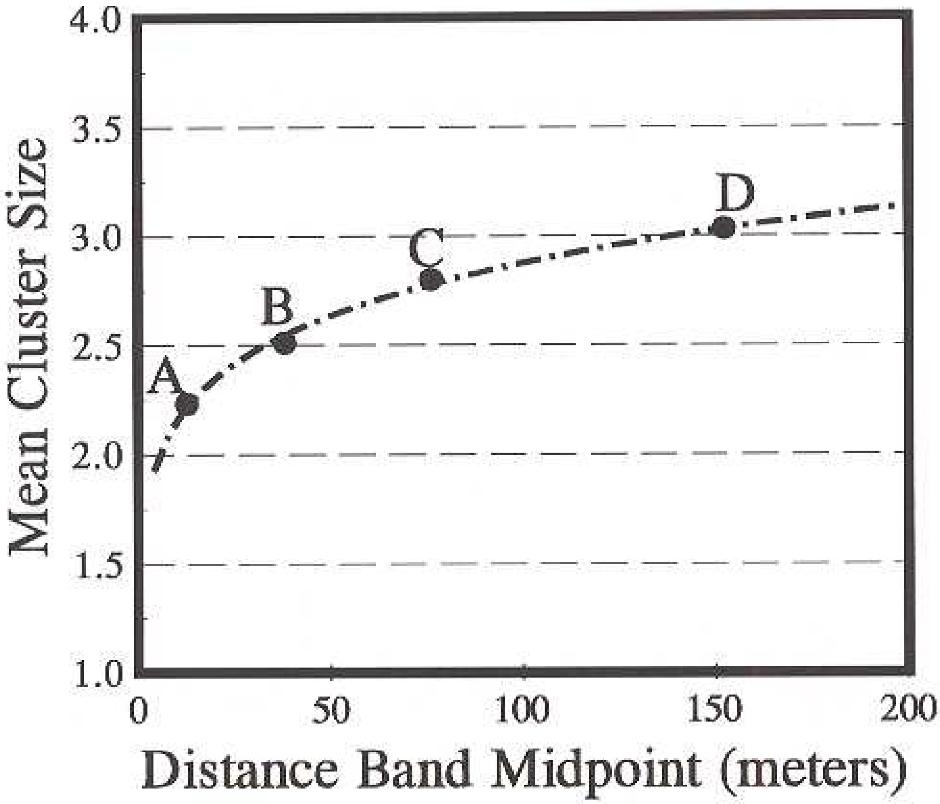


Figure 5.3. Increasing mean cluster size as a function of distance. Data are from a survey in the Hawk Springs Pronghorn Herd (Olson 1995). The overall average of 2.7 pronghorn observed per cluster is inflated by the larger number of clusters observed at outer distances (number of clusters: A=31, B=37, C=60, D=69). These data suggest a nonlinear size bias. A cluster size correction using the log-based regression in DISTANCE would be appropriate for such cases. In many situations, cluster sizes are more variable.

The following steps will lead you through the process:

1. **Line Lengths:** List the total length in miles for each transect.

2. **Adjust Midpoints of Distance Bands for Height AGL:** For each cluster observed, adjust the midpoint of the distance band for the actual height AGL using the procedures outlined above (ratio of actual height AGL to nominal height AGL). This will be the "distance" (in meters) to be input with each cluster. At the nominal height of 300 feet AGL, the distance band midpoints are 12.5, 37.5, 75.0 and 150.0 meters for the A, B, C and D bands, respectively. Adjusted midpoints can be looked up from the tables in the Appendix III or calculated by hand. The general equation follows:

$$MP_{ai} = MP_{ni} * (HT_a / HT_n) \qquad \text{(Eq. 5.6)}$$

where:

MP_i	=	midpoint for the ith distance band
HT	=	height AGL
a	=	actual
n	=	nominal

For the same example used for cut-points where the survey averaged 330' AGL (i.e., HT_a/HT_n = 330'/300' = 1.1), the corrected midpoints would be calculated as follows:

MP_{aA}	(i.e., midpoint of "A")	=	12.5 m * 1.1 =	13.75 m
MP_{aB}	(i.e., midpoint of "B")	=	37.5 m * 1.1 =	41.25 m
MP_{aC}	(i.e., midpoint of "C")	=	75.0 m * 1.1 =	82.50 m
MP_{aD}	(i.e., midpoint of "D")	=	150.0 m * 1.1 =	165.00 m

The adjusted midpoints will be entered into DISTANCE as the perpendicular distance in meters for each observation. DISTANCE automatically converts the distance data so that the density estimate comes out in the desired units for area.

3. **Observed Cluster Size:** List the observed cluster size after each distance (adjusted midpoint) for each transect.

4. **Adjust Distance Cut-points:** If you did not already calculate the adjusted distance cut-points for the overall survey (or stratum, if desired) as described in the section on Screening the Flight Report and Summarizing Data for Entry, you need to do so. Average the actual heights AGL (radar altimeter readings) for all positions where one or more clusters were seen (not observations). Exclude end-points. Adjust the outer distance limits for the average height AGL as described previously (see Equation 5.1 above) or look these up in the tables in Appendix III.

5. **Check Your Data:** Inevitably, some errors creep into the analysis with the large number of paired observations and hand calculations required. Be sure to check over your data for transcription and other errors before proceeding to build your input file for DISTANCE.

ANALYZING DATA WITH DISTANCE

Persons analyzing line transect data using the Wyoming Technique should have some background on the program DISTANCE (Buckland et al. 1993, Laake et al. 1996) since the technique has been developed to use specific options of this software. This manual describes the analysis of aerial line transect surveys similarly to features in Versions 2.1 (Laake et al. 1994) and 2.2 (Laake et al. 1996) of DISTANCE. To avoid confusion due to different default options, **the Wyoming Game and Fish Department no longer supports using earlier versions of DISTANCE**. Version 2.2 contains corrections, improved estimation procedures and different default options from earlier releases. As of January 1997, it was the "official" version of the program. Version 2.2 is intended to be the last "MS-DOS" release of DISTANCE. Subsequent versions are being designed for 32 bit platforms like Windows 95 and Windows NT. A test version, DISTANCE 3.0, was made available for evaluation at the end of 1996. It allows for flat file input, user specification of the desired level for confidence intervals, improved bootstrap estimates, and the addition of the Bayesian Information Criterion (BIC) for model selection. This "beta" version is still command driven. Future releases will have a graphical user-interface.

DISTANCE vs. TRANSECT-II

Some discussion of the advantages of DISTANCE (Laake et al. 1996) over TRANSECT (Laake et al. 1979) is warranted because the original guidelines (Johnson and Lindzey 1990) described the use of the program TRANSECT-II (the microcomputer version of TRANSECT) to analyze line transect data from aerial pronghorn surveys. **TRANSECT is now obsolete and should not be used** (*see* Buckland et al. 1993, Laake et al.1996). The use of TRANSECT-II to analyze aerial line transect surveys using the Wyoming Technique is not described in this manual. Although some users may prefer the ease of running the menu-driven version of TRANSECT to the "command language" of DISTANCE, the advantages are well worth the inconvenience to switch to DISTANCE. DISTANCE and TRANSECT may give different estimates for the same data (Buckland et al. 1993:73). DISTANCE estimates include many corrections and refinements that were not implemented in TRANSECT.

DISTANCE performs a more thorough analysis of line transect data than TRANSECT. Advantages of DISTANCE over TRANSECT include the following (Laake et al. 1996):

- a wider range of analysis options including better behaved (e.g., hazard rate) and more adaptable estimators using the key + adjustment functions (Buckland et al. 1993);
- estimates clustered populations directly including log-based regressions to correct for size bias;
- has better numerical procedures with improved constraints for monotonicity;
- allows several ways of adjustment selection including forward, sequential and all possible combinations;
- accommodates stratified survey designs;
- permits left and right truncation;
- supports bootstrapping;
- allows direct estimation of population totals, and
- provides more objective model selection criteria (e.g., AIC).

DISTANCE is also capable of handling a wider range of distance sampling applications (e.g., point transects or variable circular plots) than TRANSECT. Its advantages for analyzing aerial line transect data are numerous.

Obtaining DISTANCE

The most direct way to obtain the latest version of the DISTANCE software and User's Guide is to download them via the Internet. The software and manual were originally available from the Colorado Cooperative Fish and Wildlife Research Unit at Colorado State University. Dr. Jeffrey Laake of the National Marine Mammal Laboratory (NMML) in Seattle continues to update the program and manual. NMML maintains a DISTANCE home page on the Internet where you can find information on the program and how to download the latest version. Persons wanting a copy of DISTANCE Version 2.2 must complete a User Agreement (on-line) regarding the use of some proprietary software before downloading the program. The User's Guide can be downloaded in a variety of formats. A "printer ready" file containing the User's Guide is available for Hewlett-Packard Laser Jet printers which is easy to print. Instructions on how to contact the DISTANCE home page for information on upgrades, obtaining a copy of the manual, and downloading the program are described in Appendix IV.

The installation of DISTANCE 2.2 and associated files is relatively easy. Once you've downloaded the program to the appropriate subdirectory where you plan to use it, run the self-extracting file "DISTZIP.EXE" to expand the files.

Analysis Strategy Using DISTANCE

Buckland et al. (1993:50-51) consider the analysis of line transect surveys as a three-phase process: (1) exploration, (2) model selection, and (3) final analysis and inference. These authors provide guidelines for analysis (Buckland et al. 1993:41-51). I recommend following their strategy for analyzing line transect data using the Wyoming Technique.

The **Exploratory Phase** consists of preparing histograms and examining the data for possible violations and other complications. The Wyoming Technique imposes constraints on the number of distance groupings and the potential for truncating data. Additional truncation, if needed, must be done prior to the model selection phase. The Exploratory Phase for the Wyoming Technique is primarily limited to constructing and examining the observed histogram as it was collected. This process provides insight into potential estimation models and was described above under the Preliminary Screening section.

The second phase of the analysis involves **Model Selection**. This begins with the actual analysis of data once the data have been prepared for input into DISTANCE (Buckland et al. 1993:41-51). The program provides various measures of how well models fit the data (Buckland et al. 1993:73-77). The most important of these criteria for our purposes is Akaike's Information Criterion, **AIC** (Buckland et al. 1993:75-76; Burnham and Anderson 1992). DISTANCE also provides Likelihood Ratio and Chi-Square Goodness of Fit Tests. As Buckland et al. (1993:50-51) note, **the fit of the model to the distance data near the line is what is most important**. DISTANCE can be instructed to select the model with the lowest AIC among those requested. The sections below on running DISTANCE review this process with regard to the Wyoming Technique.

The third phase is **Final Analysis and Inference** (Buckland et al. 1993). At this stage, a final model is selected and the analyst may examine other considerations like bootstrapping. The final estimates of abundance and associated measures of precision are selected, limitations or other qualifications are noted, and the results are interpreted. The implementation of this phase to the analysis of the Wyoming Technique is also outlined below.

Running *DISTANCE*

Once the data have been summarized as described above, you can begin analyzing your data with DISTANCE. **Be sure to review the DISTANCE User's Guide** (Laake et al. 1996) and the referenced chapters in Buckland et al. (1993) before proceeding. The Quick Reference card provided with the DISTANCE User's Guide up through Version 2.1 (Appendix V) can be helpful. Although many elements of analyzing line transect data obtained using the Wyoming Technique will be similar from one survey to the next, **users are cautioned not to become complacent and "cookbook" analyses**. The analysis of line transect data should be tailored for each survey due to the complexity of factors that influence surveys and variations in survey design (Buckland et al. 1993:49).

DISTANCE can be run "interactively" or in "batch" mode. Batch mode allows the user to build an input file and then submit it to DISTANCE where the output is routed (usually) to an output file to be printed or viewed. This is the most common way to use the program and is the recommended approach for analyzing data using the Wyoming Technique. Alternately, you could run the program interactively by entering the data as the program runs and seeing the output displayed as it is calculated (this can also be done with batch operation). The analysis described in this chapter is based on batch operation since it is easier to correct errors when using that mode of operation.

File Management

Some remarks about file management and conventions with DISTANCE are in order before getting too far into the discussion of running the program. There are three basic types of files commonly used in line transect analyses (particularly in batch operation): input files, output files, and log files. I recommend that all three file types be specified when analyzing data with DISTANCE. DISTANCE also uses other files for more advanced analyses but those are not discussed here. It is helpful to name files informatively to distinguish one analysis from another. For example, a name consisting of a herd number and year such as "A527_96.INP" could be used for the input file for a survey conducted in Herd 527 in 1996. I recommend that the same filename be assigned to the corresponding input, output and log files (the file extension will differ). You can add a letter to distinguish subsequent runs (e.g., A527_96b.INP). The user can also specify whether or not to append or replace log and output files for subsequent analyses.

<u>Input:</u> Input files are constructed by the user and submitted to DISTANCE. The input file usually specifies the names of the output and log files, the title of the analysis, the options, the data, the estimators requested and the instructions for controlling the analysis. Data can be read from a separate file but that option is not described here. The input file contains the commands just as if they were entered interactively. Input files should have a ".INP" file extension. The easiest way to build and modify input files is through the MS-DOS text editor, EDIT. Input files can also be composed on standard word processors but must be saved as ASCII (i.e., text) files starting in first column. Appendix VI contains "templates" for constructing input files to analyze surveys using the Wyoming Technique. Appendix VII has an example input file for comparison.

Output: Output files are generated by DISTANCE. If no output filename is assigned, DISTANCE writes output to a scratch file, DIST.OUT. Output will be appended to this file unless otherwise specified. I recommend that an output file be assigned in the input file to a specific filename having an ".OUT" file extension. You can add a "/REPLACE" or "/APPEND" switch as desired. DISTANCE commands allow the user to control some of the output written to this file. The output file can be viewed using the BROWSE utility described below or routed to a printer to generate a hard copy. An example output file can be reviewed in Appendix VII.

Log: DISTANCE writes error messages to the log file and is the easiest way to debug problems with the analysis. I also recommend that the log file be assigned in the input file using a ".LOG" file extension. The log can also be viewed using BROWSE or printed. The User's Guide provides more advice on using log files.

Starting DISTANCE in Interactive Mode

To start DISTANCE in interactive mode, go to the subdirectory in DOS where the DISTANCE program is stored and enter:

DIST

The program will begin, giving you a DISTANCE> prompt where you can start entering commands. The program can also be run under Windows using Program Manager. The interactive mode of operation is not described in detail in this manual.

Running DISTANCE in Batch Mode

To analyze an input file with DISTANCE in batch mode, go to the subdirectory in DOS where the DISTANCE program is stored. Enter the following command from the DOS prompt:

DIST I=*filename*.INP

If the input file is located in another subdirectory or drive, you will need to include the path for the input file. You can also specify the name of the output and log files on this command line. However, I recommend that you ASSIGN these file names from within the input file. See Laake et al. (1996) for more explanation on running the program.

DISTANCE Command Language

DISTANCE uses a simple command language that instructs the program as to how to interpret and analyze input data (Laake et al. 1996). A Quick Reference Sheet (Appendix V) outlines the commands. Although DISTANCE will usually be run by batch operation, it is helpful to discuss the command language as if the program were being run interactively. DISTANCE is particular about the order in which most commands are given and requires specific syntax. DISTANCE expects a ";" at the end of each command line except when entering distance and cluster size data for a particular transect (a semicolon in that case denotes the end of data entry for a particular line). If you forget to enter a ";" DISTANCE responds with the following message:

[waiting for input]

The main program prompt is:

DISTANCE>

To stop the program, enter "EXIT;" or "END;" at the DISTANCE> prompt. The program is somewhat hierarchical. There are three primary commands issued from the DISTANCE> prompt:

OPTIONS;
DATA;
ESTIMATE;

When you enter one of these commands interactively, the program prompt changes accordingly:

OPTIONS>
DATA>
ESTIMATE>

Subcommands are issued under each of these primary commands. The order in which most commands are given will often make a difference. To return to the DISTANCE> prompt, enter "END;" at the primary command prompt.

The following section lists some common commands in the order they generally occur in a typical input file. See the DISTANCE User's Guide, the Quick Reference (Appendix V), and examples (Appendices VI and VII) and for more information. **Italicized items are particular values or words you provide.**

ASSIGN: Another important command issued at the DISTANCE> prompt is ASSIGN (Laake et al. 1996:17). This is normally one of the first commands entered in a DISTANCE input file and is used to assign names to input, output and log files. It is important to assign output and log files for all batch runs. The first two commands in most input files should assign the output and log files:

ASSIGN LOG='*filename*.LOG';
ASSIGN OUTPUT='*filename*.OUT';

You can also add a switch to append or replace these files from subsequent runs (see the Quick Reference sheet). Use single quotes to assign filenames.

OPTIONS: The OPTIONS command allows the user to specify various options and information about the type of analysis (Laake et al. 1996:22). This is where you can give the analysis a title, define the type of analysis (perpendicular distances, clustered populations, etc.), and list the units in which line lengths, distances and areas will be handled. Specify "OBJECT=CLUSTERS/EXACT" since cluster sizes were counted and not estimated. DISTANCE will interpret and apply the proper conversions. At the end of this section of input, add the "END;" command to return the program to the DISTANCE> prompt. For typical line transect analyses using the Wyoming Technique, the OPTION section will resemble the following:

OPTIONS;
TITLE='*Your title goes here*';
DIST=PERP/MEASURE='METERS';

```
LENGTH/MEASURE='MILES';
AREA/UNITS='SQ. MILES';
OBJECT=CLUSTER/EXACT;
END;
```

Be sure to specify these units.

If only one observer participated in the survey, you can compensate using the sampling fraction (SF) command under OPTIONS> by specifying "SF=0.5;" By default, SF=1.0.

DATA: The next section of the input file lists the data (Laake et al. 1996:39). If you want DISTANCE to estimate a population total, you provide an estimate of the area using the "STRATUM" command even if the survey is not stratified (there will be just one stratum in that case). If you are analyzing a stratified survey, you would list the area of the first stratum here. You can also give a label for the stratum.

The next command line identifies the first transect and its length using the "SAMPLE" command. You can also give that a label. After you enter the ";" for the "SAMPLE" line, list matched pairs of perpendicular distances (adjusted distance band midpoints) and cluster sizes (from the summary you prepared). When you finish the data for a line, enter ";" to tell the program that all the data for that line have been entered. It is easier to check when the data are listed as two columns separated by spaces.

To start with the data for the next line, enter the "SAMPLE" command and repeat the process. To enter data for another stratum, enter the "STRATUM" command and related instructions and proceed to list the data for each transect as before. When you've entered all your data, enter "END;" to return to the DISTANCE> prompt.

The following gives an example of the commands listing the data for two lines where *d* represents a perpendicular distance band midpoint and *c* represents the corresponding cluster size for that observation:

```
DATA;
STRATUM/AREA=1234.5/LABEL='Occupied habitat';
SAMPLE=22.2/LABEL='Line 1';
d11   c11
d12   c12
d13   c13;
SAMPLE=33.3/LABEL='Line 2';
d21   c21
d22   c22
d23   c23
d24   c24;
END;
```

You are now back to the DISTANCE> prompt. Here is where you can request DISTANCE to print out certain information such as the options and data. The appropriate commands are as follows:

PRINT OPTIONS;
PRINT DATA;

ESTIMATE: The program is now ready to begin the estimation process. I will discuss specific estimators later. The following describes some basic commands for running the analysis (Laake et al. 1996:43-46). Buckland et al (1993:62-64, 69-71) describe the merits of particular models using the key + adjustment format. I provide an overview of model fitting using this approach later. Consult with the DISTANCE User's Guide for requesting specific estimators. In the following example, estimation options will be printed. Data are to be analyzed using the distance intervals listed. The *cp*s are the distance band cut-points adjusted for height AGL and summarized earlier. Be sure to include the cut-point "0.0" (i.e., the line) in the list. Three estimators are requested in this example. The first is a half-normal with cosine adjustments where DISTANCE sequentially fits a 1 parameter model, then a 2-parameter model and so on until a criterion is reached. The second estimator is a uniform model where I've forced DISTANCE to use 2 cosine adjustment terms (NAP=number of additional parameters). The last estimator is a hazard model with a polynomial adjustment. The "CLUSTER/TEST" command requests that DISTANCE test for significant cluster size bias and use a log-based regression to correct for size bias if it is significant. "PICK=AIC" specifies that DISTANCE should select the "best" among these models using the AIC. DISTANCE will list only the detailed output for that model. "END;" returns to the DISTANCE> prompt.

ESTIMATE;
PRINT OPTIONS;
DISTANCE/INTERVALS=0.0,*cp1,cp2,cp3,cp4*;
ESTIMATOR/KEY=HNORMAL/ADJUST=COSINE/SELECT=SEQUENTIAL;
ESTIMATOR/KEY=UNIFORM/ADJUST=COSINE/SELECT=SPECIFY/NAP=2;
ESTIMATOR/KEY=HAZARD/ADJUST=POLY;
CLUSTER/TEST;
PICK=AIC;
END;

Entering another "END;" will stop the program and return to the DOS prompt. At that point, you can review the log file for errors and examine the results listed in the output. Following the review of the output, the input file could be revised and the analysis resubmitted.

Special Considerations for Analyzing Aerial Line Transect Surveys

The above DISTANCE commands are fairly general. Additional commands and options should be considered for analyzing aerial survey data using the Wyoming Technique. Review the example templates (Appendix VI) and analysis (Appendix VII) for aerial pronghorn surveys. The following discussion highlights certain aspects in the analysis of data using DISTANCE. These may be helpful when considering refinements to the estimation following a preliminary analysis.

Key + Adjustment Functions: Ideally, line transect estimators should satisfy the shape criterion by having a shoulder (region of relatively flat detection) near the y-intercept where virtually all of the clusters are detected (Buckland et al. 1993:42-44). To some extent, the detection curve models the manner in which observers search, but also reflects the observability of clusters.

Readers should review pages 41-51 of Buckland et al. (1993) for an overview of modeling detection using the key+adjustment approach. Intuitively, the uniform key with cosine adjustments is a good model for cases where the data exhibit a shoulder (Buckland et al. 1993:46-48). This is essentially the same as the Fourier Series Estimator from TRANSECT, but with improved numerical properties. The uniform key can be used to estimate strip transect data as well. The half-normal key with cosine adjustments (Buckland et al. 1993:46-49) also tends to "behave well" for many line transect surveys having a shoulder on the detection curve near the line.

During some aerial surveys, detectability may decline more rapidly. This is common where mean cluster sizes are small, densities are low, observers are focusing on a narrow region and/or visibility conditions are poor. In some of these cases, the negative exponential key may provide an acceptable estimate. However, the potential also exists to grossly overestimate density using the negative exponential because of the lack of a defined shoulder (Buckland et al. 1993:45) because they fit the region near the line better (Buckland et al. 1993:50-51). Although the negative exponential may have the lowest AIC, other models may be better. In those situations, the hazard rate key with polynomial adjustments (Buckland et al. 1993:43-49) may be acceptable because it forces a shoulder within at least a small region near the line.

Problems occur when observers fail to detect a significant number of clusters in the A band. To salvage these estimates, one must assume whether the animals were missed or moved to an outer band before being seen. In most cases, the "inertia" of the detection curve tends to compensate for missing animals near the line because the A band is relatively narrow..

Some data sets are difficult to model, particularly given the limited number of distance bands and numerical constraints imposed using the Wyoming Technique. In some instances, a more reasonable point estimate may be obtained by forcing DISTANCE to use a model with extra parameters (using "/SELECT=SPECIFY/NAP=*m*" where *m* is the number of parameters you want to use). Additional parameters allow the detection curve to be more flexible, thereby fitting the histogram near the line better. This approach often takes a toll on the precision of estimates. With the Wyoming Technique, the Goodness of Fit test cannot be calculated with models having more than two adjustment terms. Analysts should document why forcing a particular model is appropriate. In many cases, the improvement over more parsimonious (simpler) models is marginal. Biologically, that estimate may be no better than a simpler model.

As a rule of thumb, let the AIC objectively determine the initial "best" model (i.e., the one with the lowest AIC value) from among those you specified. While "PICK=AIC;" is the default, you can manually examine the individual AICs in the Estimation Summary - Detection Probability table when you specify "PICK=NONE;" under ESTIMATE>. In many cases other viable models will have AIC values close to the one for the "best" model. The best model should fit the histogram near the line (i.e., the A band) well.

<u>Correction for Cluster Size Bias:</u> The proper procedure for handling cluster size bias depends upon which criterion was used during the survey. The preferred method is to let DISTANCE perform a size-bias regression to test for cluster size bias, and if significant bias is present, to use the regression of the natural log of cluster size on detection probability to correct the bias. To perform this test, use the "CLUSTER/BIAS=GXLOG/TEST;" command under ESTIMATE> (Laake et al. 1996:48). Where significant bias is not evident, precision can be improved by using the average cluster size instead of the expected cluster size from the regression. When significant bias is present, the GXLOG regression seems to perform adequately.

For those who use the *ad hoc* method of estimating mean cluster size by truncating within the A and B bands, the analysis will need to reflect that procedure. The proper command is "CLUSTER/WIDTH=*cp2*/MEAN;" where *cp2* is the outer distance limit for the B band adjusted for mean survey height AGL. The *ad hoc* method assumes there is no size bias within the A and B bands.

Stratified Surveys: Stratified survey designs have practical advantages although they currently receive limited use in Wyoming pronghorn surveys. These can be analyzed using DISTANCE. An accurate estimate of the size of each stratum is required. Options exist to either estimate a density correction factor (e.g., f(0)) for each stratum, or to estimate a pooled correction factor across strata. The sample sizes within strata will determine which approach is suitable. See Buckland et al (1993:99-102) and Laake et al. (1996:49-51) for options.

Bootstrapped Confidence Intervals: Bootstrapping is a technique to provide robust, nonparametric variance estimates and confidence intervals using computer-intensive resampling of the data (Buckland et al. 1993:94-96). Bootstrapped confidence intervals can be obtained with DISTANCE (Laake et al. 1996:23). This option may take a long time to calculate on some computers so potential users should plan ahead before requesting this analysis. Bootstrapping is helpful in evaluating uncertainty in model selection.

Controlling Output: DISTANCE allows the user to control some of the output. Defaults are set to provide results for the "best" model unless "PICK=NONE;" is specified. Adding the command "SQUEEZE=ON;" to OPTIONS> eliminates form feeds, which has the effect of compressing the output when it is printed (Laake et al. 1996:33).

DISTANCE Output

The DISTANCE output lists the type of analysis and options requested, the results, and numerous diagnostic information on which to assess the adequacy of the data and the estimation process. The following sections describe the DISTANCE output, which features on the output are most important for evaluating results of surveys using the Wyoming Technique, and the basics of interpreting analyses. The DISTANCE User's Guide (Laake et al. 1996) discusses example output and Buckland et al. (1993) provide background on the analysis procedures and examples from actual surveys.

Once you have run a preliminary analysis, you should inspect the results to make sure that data were entered properly, to determine whether any analytical problems occurred, and to see if modifications to the analysis are warranted. I recommend analysts review preliminary results on the computer before printing a hard copy of the output. DISTANCE output files can exceed 30 pages in many analyses. The output (*filename*.OUT) and log (*filename*.LOG) files can be easily viewed on the computer using the utility, BROWSE.COM, provided with the DISTANCE program. To run BROWSE, type "BROWSE *filename*.OUT" (or *filename*.LOG as the case may be) from the DOS prompt at the subdirectory where DISTANCE was installed (you may need to specify a path if the input and output are located in another subdirectory). You can scroll through the file with the cursor keys. To exit BROWSE, hit the ESC key. Alternately, you can print the output from the DOS prompt by entering "PRINT *filename*.OUT" (or *filename*.LOG as needed). If the computer prompts for a "[PRN]" device, hit the ENTER key. BROWSE can also be used to review input files. However, you may prefer to use the DOS editor, EDIT, to view and modify input files for reanalysis.

Overview of Output and Guidelines for Interpretation

The following section provides an overview of typical DISTANCE output based on the User's Guide (Laake et al. 1996) and highlights selected features that are of particular interest to analyses of surveys using the Wyoming Technique. A brief discussion of each of the sections in the DISTANCE output follows. Those marked with an "*" are particularly relevant for our purposes. Appendix VII contains sample input and output files from an actual survey which can be reviewed in conjunction with this discussion. Output from individual surveys may differ, depending upon the options specified in the analysis. The limited number and specified widths of distance bands using the Wyoming Technique constrain the analysis. In some situations, the analyst may need to "force" certain types of models to obtain acceptable estimates. You may wish to look at the output in Appendix VII as you read this section.

Program Options Listing: This is the first part of the output and lists the title (if specified), type of analysis, basic options, and program defaults. For the Wyoming Technique, this should indicate the Type as "line," Object as "clusters," Distance as "perp," Units (for distance) as "meters," Length in "miles," and Area as "sq. miles." If you specified other units for the analysis, those should be indicated here. Conversion factors are also listed. The conversion factor "1609.3" should be listed under "Area" to convert perpendicular distances from meters to miles.

Data Listing*: This optional section of output is frequently helpful in finding errors. Quite often, mistakes in input can be located by comparing the raw data to this listing. The area for the stratum or entire study area should be listed first if a population total is to be estimated. Data for each line within a stratum are listed next. This includes the line length (effort), the perpendicular distance (adjusted midpoint of distance band), and the corresponding cluster size.

Estimation Options Listing: This section of output lists the analysis options including how cluster sizes are handled and which estimators (key + adjustment terms) were requested. A glossary of abbreviations can be found at the end of this section of the output which is helpful when reviewing the tables of results generated by the program.

Probability Function Estimation - Model Selection/Fitting: The total line length ("Effort"), number of transects ("# samples"), outer distance "Width" (i.e., the outer cut-point specified) and the number of clusters ("# observations") are listed at the top of this section. Compare these with the raw data to be sure all lines were entered properly. If the number of clusters is different from the known total, check to see whether all data were correctly entered or if some observations were truncated. Occasionally, some clusters may be observed in the outer distance bands when the plane is exceptionally higher than the nominal height of 300'. In those instances, the adjusted distance band midpoint may occur outside the distance limit specified by "Width." This results in those observations being censored from the analysis. This section of output also shows the sequential results of fitting all requested models.

The next three subsections under "Probability Function Estimation" are only presented for the "best" model when "PICK=AIC" is specified. If "PICK=NONE" is specified, then data for all requested models are displayed for these subsections.

Probability Function Estimation - Parameter Estimates: The selected model and its parameter estimates are listed in this part of the output if the analysis specifies DISTANCE to select the "best" model based on assigned criteria (e.g., AIC). If the command "PICK=NONE" was entered under "ESTIMATE>", then results for all requested estimators are shown. A correlation matrix for parameter estimates is printed for models having two or more parameters (i.e., adjustment terms).

Probability Function Estimation - Detection Probability Plot*: DISTANCE generates a histogram of the observed data showing the estimated detection function for the selected model (or models if several are requested to be output). Figures 5.4 A-D show examples from actual Wyoming pronghorn surveys of the four basic key functions: uniform, half-normal, negative exponential, and hazard rate. The detection probability plot, when considered with other diagnostic measures, is extremely helpful in evaluating the adequacy of the estimator for modeling the observed data. This histogram should be grossly similar to the one calculated during data screening.

One of the main considerations in modeling detection is how well the estimated function models the histogram near the line (Buckland et al. 1993:50-51). The first histogram and estimator in the output of the example analysis in Appendix VII shows a model that poorly fits the A band. Ideally, the selected estimator should have a "shoulder" near the y-intercept of the detection curve as shown in the examples in Figure 5.4 A, B, and D. The Wyoming Technique is more likely to underestimate rather than overestimate herds due to the nature of aerial line transect sampling. However, detection curves that intercept the y-axis well above the "A" band should be scrutinized closely (this is not uncommon with models like the negative exponential). The narrowness of the "A" band makes it less likely that detectability declines sharply within that band. If observers follow proper procedures (e.g., do not try to place extra clusters in the "A" band), then the top of the "A" band is an appropriate "target" for the detection curve to intercept the Y-axis. Figure 5.5 A shows a real example where the estimated detection curve (negative exponential) "overshoots" the "A" band. This detection model does not meet the desired "shape criterion" (Buckland et al. 1993:42). A better-behaved model (see Buckland et al. 1993:42-45) would be a hazard-rate or half-normal (Fig. 5.5 B) estimator.

Occasionally, the detection probability plot indicates problems in the data. Check to see that the "A" band is the tallest (or is at least close to the top of the "B" band). If not, some additional

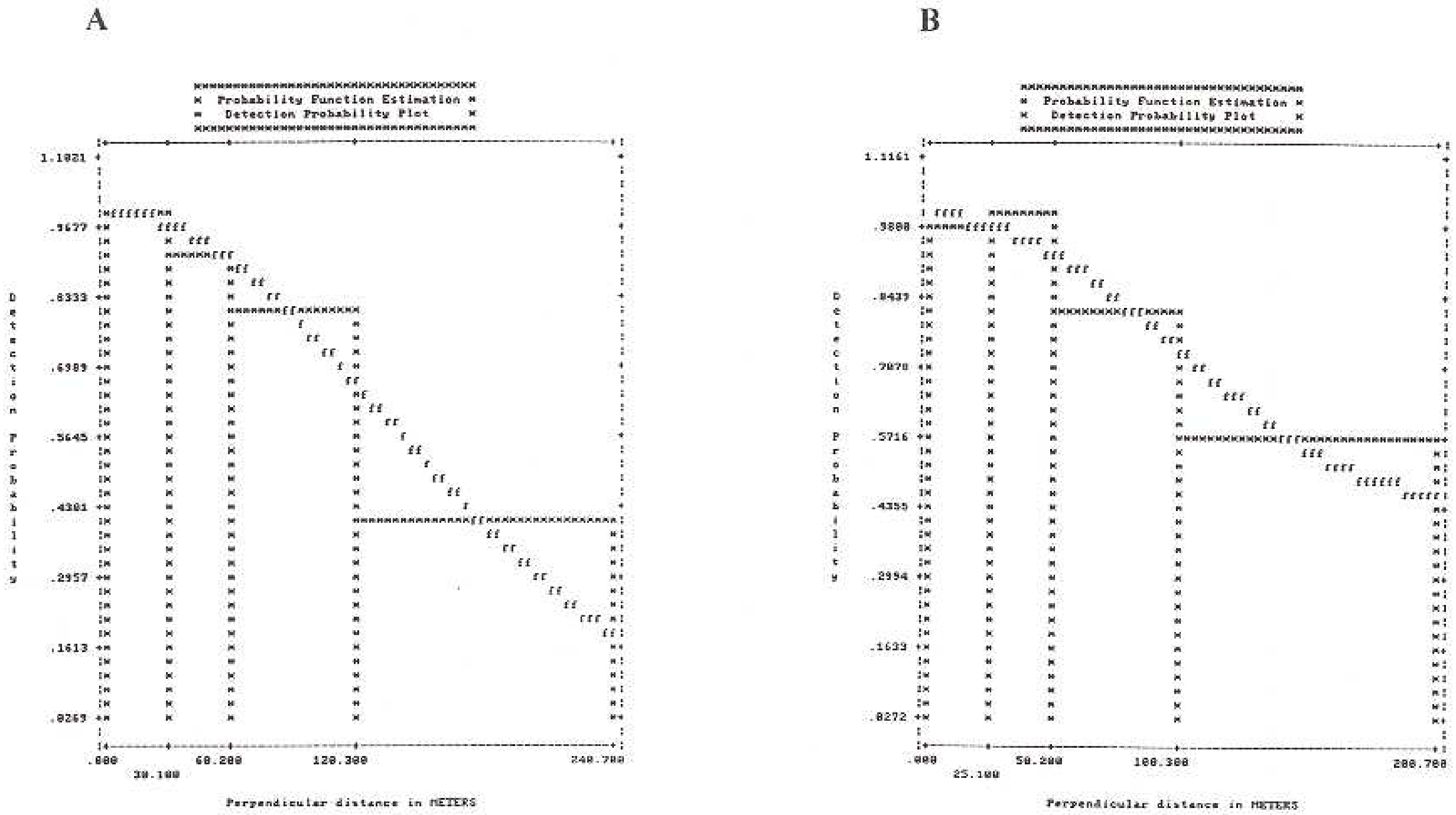
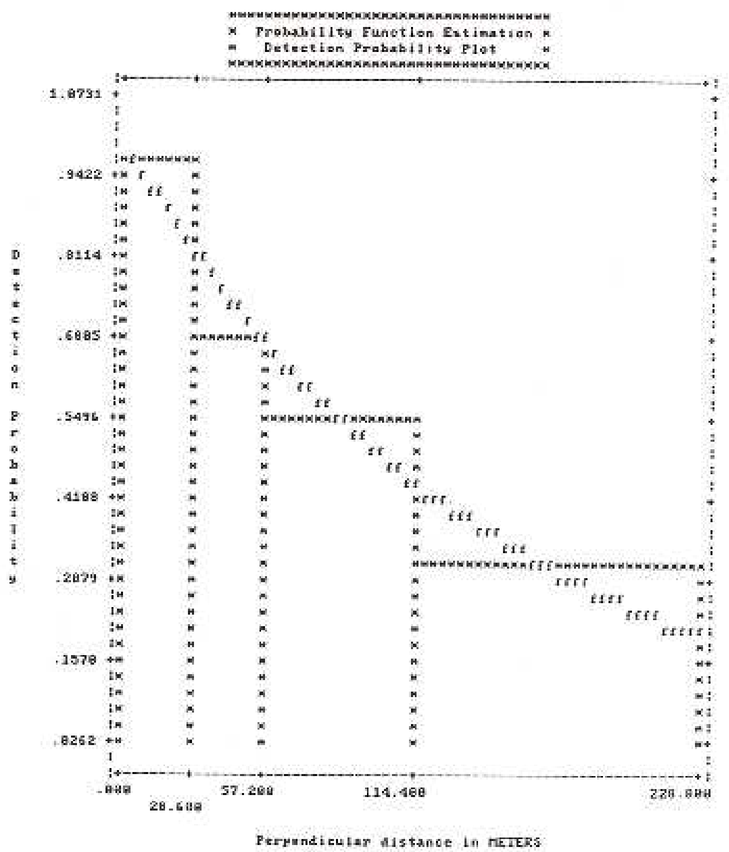


Figure 5.4. Representative histograms and detection curves for aerial line transect surveys of pronghorn using the Wyoming Technique (actual survey data): (A) half-normal and (B) uniform/cosine (continued).

C



D

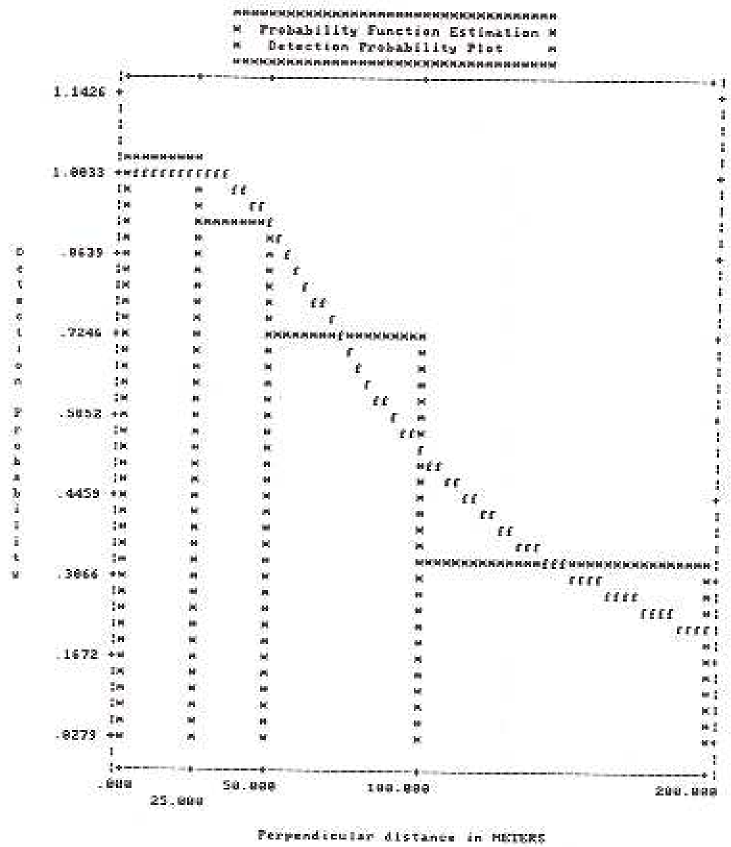
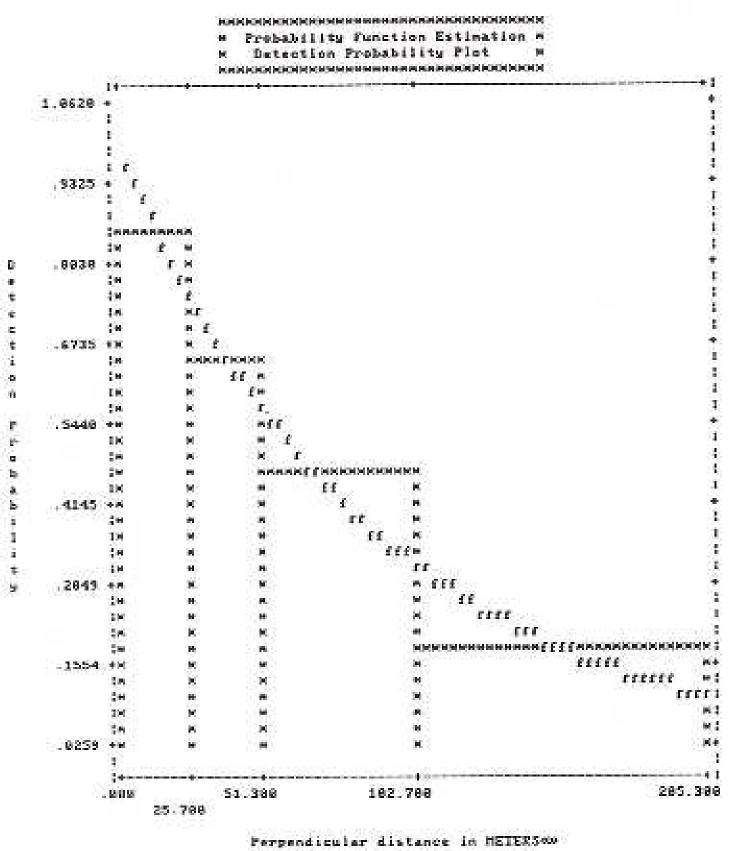


Figure 5.4 (continued). **Representative histograms and detection curve for aerial line transect surveys of pronghorn using the Wyoming Technique (actual survey data): (C) negative exponential and (D) hazard.**

A



B

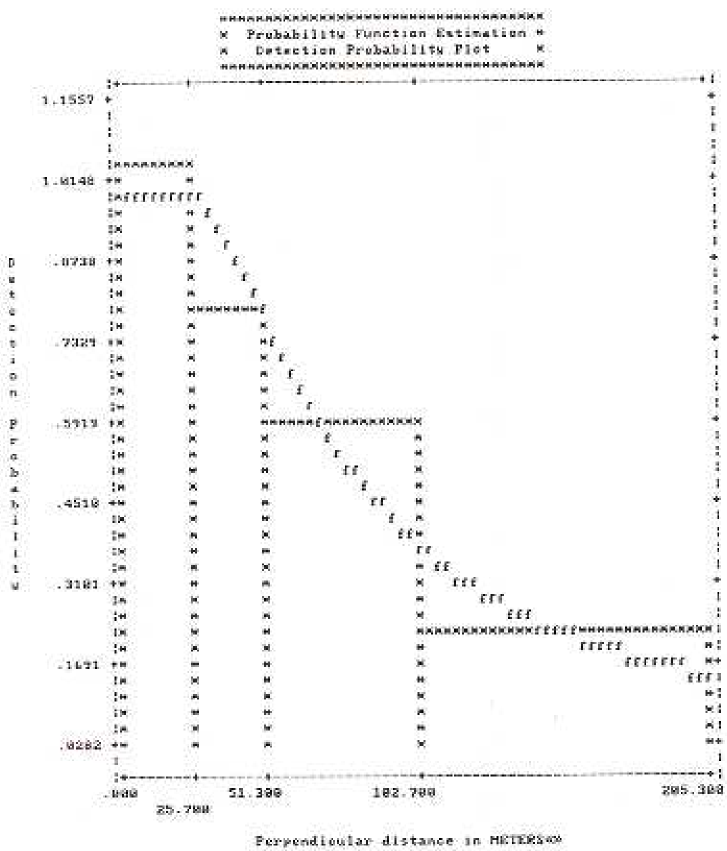


Figure 5.5. Example histograms and detection curves for aerial line transect surveys of pronghorn using the Wyoming Technique (actual survey data) where detectability declines rapidly: (A) negative exponential model overestimating density by not meeting the shape criterion, and (B) hazard rate model where a shoulder is imposed to satisfy the shape criterion.

assumptions may be required to select the "proper" model. Although the height of the "A" band is subject to some sampling variability, the detection curve from most aerial pronghorn surveys using the Wyoming Technique should be close to the top of the histogram for the "A" band. Figure 5.6 A is an actual example where the relative number of clusters seen in the "A" band was much lower than for the "B" or "C" bands. Histograms such as this suggest that pronghorn on and near the line were missed, violating a major assumption for the technique. An alternate explanation could be that pronghorn in "A" moved into the "B" and "C" bands before being observed. The latter scenario seems less likely based upon experience (Johnson et al. 1991). A histogram like Figure 5.6 A may occur where observers fail to detect all animals on and near the line. That can result from inexperienced and airsick-prone observers, or where airspeeds are too fast to accurately count close to the plane. Fortunately, the impacts of such problems tend to be minimized because of the narrow width of the "A" band using the Wyoming Technique (see Fig. 5.6 A). In rare instances, left truncation (i.e., excluding the A band from the analysis) may be appropriate (Buckland et al. 1993:15).

The opposite problem may occur when observers try to place clusters in the "A" band which were just outside. This might occur when pronghorn are either in the blind area, the inner region of the "B" band, or when observations are assigned to distance bands prior to being perpendicular. In low density herds, inexperienced observers may feel that they aren't seeing "enough" in the "A" band, giving the benefit of a doubt to that distance band. A histogram like the actual example shown in Figure 5.6 B raises suspicion that some observations were incorrectly assigned because of fewer clusters than expected being recorded in the "B" band. Training observers and experience can help minimize such problems.

Probability Function Estimation - Chi-Square Goodness of Fit Test*: The Chi-Square Goodness of Fit test indicates how well the detection curve "fits" the observed histogram. In some cases, the specified estimator may generally fit the overall histogram, yet fail to fit the "A" band properly (e.g., the first model in Appendix VII does not fit the A band well). In those cases, the model may seriously underestimate the true density. This problem is more common where sample sizes are relatively small, the "A" band is substantially higher than the others, and where the limited number of distance bands imposed by the Wyoming Technique does not allow sufficient degrees of freedom to test a higher-order model (i.e., one with more than two parameters so that the detection curve is more "flexible"). Compare the fit of the negative exponential and hazard models in Appendix VII to the first model (uniform model with one cosine adjustment in that example). The power of the Chi-Square test if fairly low, but nonsignificant values indicate an acceptable fit for well-behaved data.

Expected Cluster Size Estimation*: This analysis is output when "OBJECT=CLUSTER" is specified. It provides diagnostic information to evaluate the magnitude of cluster size bias. The (uncorrected) mean cluster size can be compared to the expected (corrected) cluster size. Results of a t-test are included which can be used to assess the magnitude of size bias evident in the data. If size bias is not significant, using the mean cluster size will improve precision of the estimate. The analysis in Appendix VII shows an example where cluster size bias was not significant.

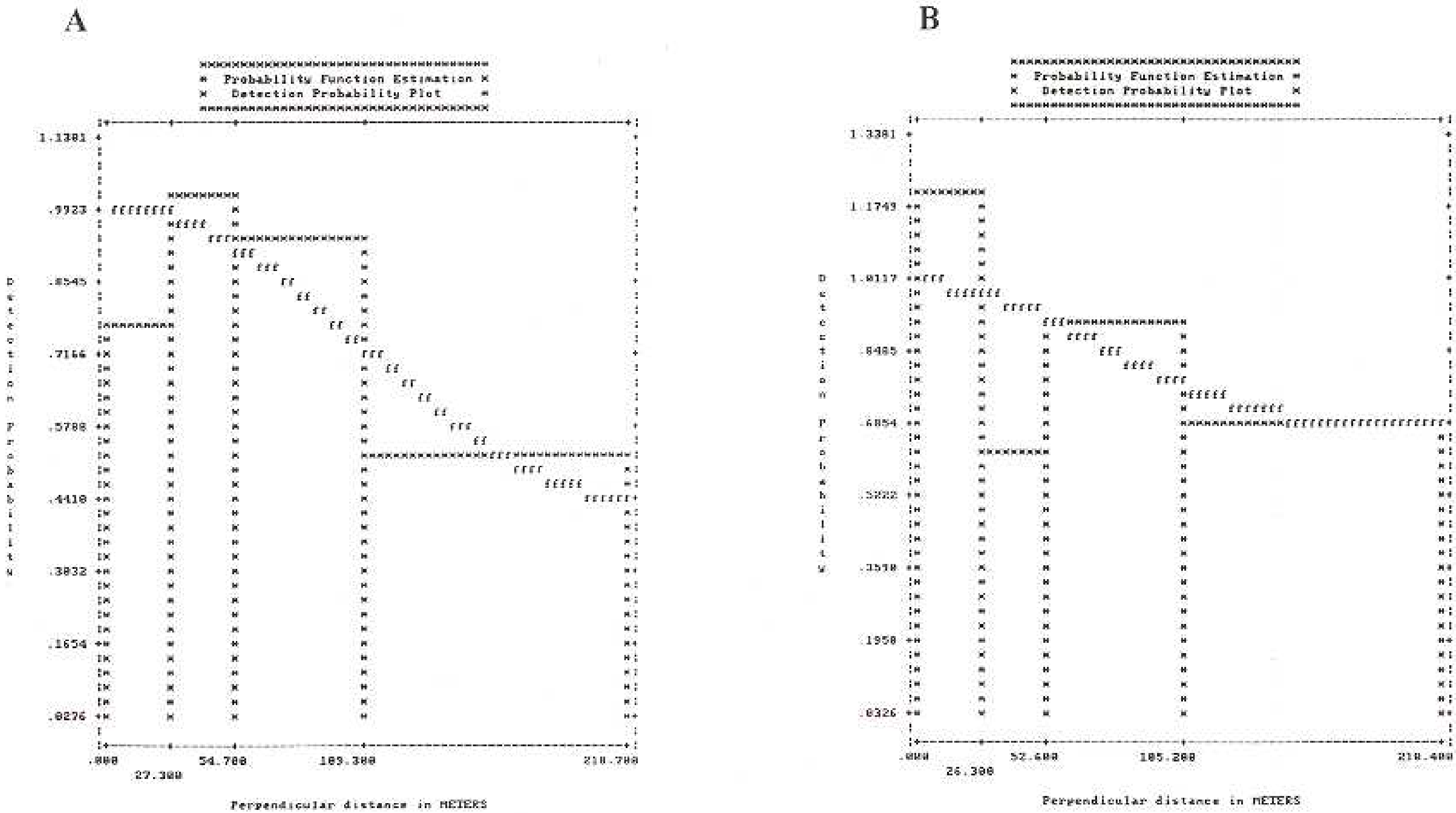


Figure 5.6. Example histograms and detection curves for aerial line transect surveys of pronghorn using the Wyoming Technique (actual survey data) suggesting improper survey procedures: (A) evidence that animals on and near the line were missed or moved before being detected, and (B) evidence that animals were misassigned to the inner distance band.

Cluster Size Regression Plot: This plot is generated whenever "OBJECT=CLUSTER" and the size-bias regression are specified. If DISTANCE selects the "best" model, then only the plot for that model is output.

Density Estimation Results*: This table summarizes some of the most important results of the analysis for the best model. It lists the specific model and gives point estimates, standard errors, percent coefficients of variation (CVs) and 95% confidence intervals for several important parameters including:

f(0)	-	the value of the probability density function evaluated at the line; i.e., the correction factor for density
ESW	-	the effective strip width
n/L	-	the encounter rate; number of objects detected per unit length of transect
DS	-	density of clusters
E(S)	-	expected cluster size
D	-	density of individuals
N	-	population total (if an area is specified)

The percent CV is helpful in comparing the relative precision of estimates. Ideally, surveys should have CVs for density less than 15%. In some herds, CVs $\leq$20% are acceptable. Estimates with larger CVs should receive greater scrutiny.

Additionally, this section of the DISTANCE output includes a breakdown of the component percentages of the variance of density into probability of detection, encounter rate, and cluster size. Such information is helpful in refining survey designs to improve precision.

The table is repeated for all requested estimators when "PICK=NONE" is assigned.

Estimation Summary - Encounter Rates: This table breaks down the estimate of encounter rate into its components and duplicates some of the information presented in the "Density Estimation Results" table.

Estimation Summary - Detection Probability: This table provides a summary of parameters for the estimation of detection probability such as the key + adjustment model, the number of parameters, f(0), AIC, and the Chi-square probability. It also duplicates some of the information in the table of "Density Estimation Results." This table is most useful for comparing several estimators when "PICK=NONE" is specified.

Estimation Summary - Expected Cluster Size: Summary results of the size-bias regressions to correct expected cluster sizes are shown in this table for specified models. The correlation (usually the natural log of cluster size on detection), probability and expected cluster size are listed. Again, this table is most informative where several estimators and "PICK=NONE" are specified.

Estimation Summary - Density/Abundance*: Like the other summary tables, this one is most useful for comparing the results of several estimators when "PICK=NONE" is defined. The table provides the estimate, coefficient of variation and 95% confidence interval for the density of clusters, the density of individuals and the total for each requested estimator. This is often one of the most useful of the Estimation Summary tables.

Bootstrap Summary of Density/Abundance: This table gives the bootstrapped coefficient of variation and confidence interval when requested. This is especially useful as a way to build in model selection uncertainty into the analysis.

Final Model Selection and Use

The estimation process should be refined until an acceptable (in a statistical sense) estimate is obtained or no further improvement is possible with the data:

- Compare the estimated population size from DISTANCE and the Quick & Dirty method to see whether or not these agree or if there is a reason why they diverge.
- Evaluate the overall fit of the model with the lowest AIC.
- How well does the model fit the histogram of observed clusters (i.e., GOF chi-square)?
- How well does the detection curve model the A band?

At that point, users should review the final analysis, evaluate the reasonableness of estimates, and make interpretations. Suggestions for improving the survey design should be formulated and reported with results. In some cases, estimates will be poor and users should acknowledge the their limitations.

Less stringent confidence limits (e.g., 80%, 85%) may be acceptable for many management purposes. DISTANCE Versions 2.2 (Laake et al. 1996) and earlier only provide 95% confidence intervals. Dr. Jeffrey L. Laake has generously provided an Excel spreadsheet, LIMITS, which allows users to take DISTANCE output and recalculate log-based confidence intervals for user-specified levels. LIMITS is discussed in Appendix VIII.

The interpretation of aerial line transect estimates and their use in management is reviewed in the next chapter.

The implementation of line transect sampling to pronghorn management in Wyoming should be adaptive (see Guenzel 1994). Surveys should continue to be improved based on results from prior surveys. This manual does not discuss all possibilities for improving survey designs and analysis.