Integration of Scanned Infrared Air Photos and Coordinate Files in Arc/Info - a Biohabitat Example from Southern Sweden

Andreas Oxenstierna

Remote Sensing and GIS Laboratory
Department of Physical Geography - Solvegatan 13
S-223 62 Lund Sweden - Fax Int +4646143859

Görgen Göransson

Department of Ecology - Ecology Building
S-223 62 Lund Sweden

ABSTRACT

The article describes a methodology developed to determine pheasant habitat selection and survival fitness in Scania, southern Sweden.
A vegetation class database was created by combining remote sensing techniques with various kinds of field data. The databse is used as input for calculating several bioecological parameters.

An infrared (sensitivity 0.5-0.9 µm) air photo from 4 600 meters was scanned, geometrically corrected against 1:10.000-maps and resampled to 2x2 metres resolution. The resulting image size is 2000 x 2000 pixels. Fifty-three training areas, taken from field data, were used to calculate autocorrelation distances between 22 proposed vegetation classes (and one shadow class). After evaluation and further field checks, 16 classes remained for the maximum likelihood classification.

The classified image was cleaned-up twice, using a mode filter, with a window size of 5x5 pixels. The image was then exported and vectorized into ARC/INFO geographical information system. The vectorized lines were generalized by eliminating unnecessary vertices. All polygons not belonging to tree or bush classes with areas less than 10 m^2 were eliminated. Small tree and bush patches were preserved because they are important for the pheasant habitat selection.

Activity aeras were determined by daily monitoring through radio transmitter on selected pheasant individuals and computed by harmonic mean calculations. The coordinate files, describing the perimeter of each activity area, were imported into ARC/INFO and used in order to build polygons. The polygons were used as clipping polygons against the vegetation class coverage. From the resulting "activity polygons", vegetation class statistics were calculated for each activity area.
The vegetation statistics were correlated against the number of hatched chickens and individual fitness (i.e. survival rate).

The database will be used to print several kind of maps (vegetation, topographic, bioecological etc.) and to evaluate other animal habitat selection strategies.

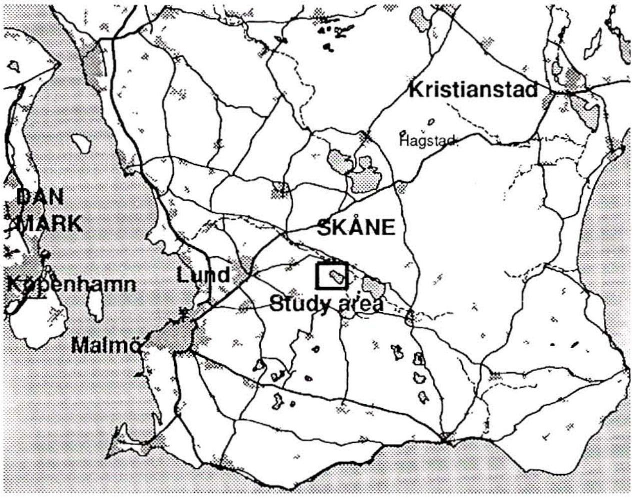
1. BACKGROUND

Animals are rarely distributed at random but to the occurence of essential resources. The availability of different necessary resources predict the survival, growth and reproductive output of animals. The resource availability is in turn determining the quality of a territory or a homerange. One method to evaluate territory or homerange quality is to measure individual fitness in different territories and use the fitness as a quality index.

In this study we have monitored males and females in a ringneck pheasant (*Phasianus colchicus*) population inhabiting 500 hectares area during 7 breeding seasons (1984-1990) and measured fitness as the number of chicks produced by each individual for its life- span and which chicks have reached at least 6 months of age.

To objectively measure habitat features of territories and homeranges remote sensing techniques were used. This article describes the remote sensing techniques used to determine and quantify the vegetation distribution in the pheasant homeranges and how the created vegetation database is used together with several biological parameters to get a better understanding on the biological and ecological behaviour of the pheasant.

The studied area covers 500 hectares in southern Sweden (see map). The area is situated around a minor lake and constitutes of habitats ranging from small lakes and marshes to dry sandy ground. It is primarily used as a foraging area for cattle.
To determine the vegetation quality and to obtain an accurate vegetation classification, remote sensing techniques were decided to be necessary.



The study area in southern Sweden. Scale 1:1.000.000.

2. IMAGE CORRECTION AND RECTIFICATION

An official air photo, taken the 30th of June 1986, time 10.17, with three different channels ranging from 0.5 to 0.9 μm, from 4.600 metre height by the Swed Survey in scale 1:30.000 was available.

The airphoto was scanned by FOA (the Swedish Military Site, Linkoping) with an Optronics Scanner to 0.75 metres resolution.

The scanned image was converted to the image analysing system Dipix ARIES III residing on a MicroVaxII at the Remote Sensing and GIS Laboratory in Lund. The ARIES III system is a raster based remote sensing analysing system, consisting of several subsystems that are used for different tasks. For further description of the system, see the manuals (Dipix, 1985).

2.1 Radiometric Corrections

Radiometric correction were not done. This decision was made for two reasons:

- Necessary data concerning atmospheric and ground (soil) parameters were missing or not easily available
- Supervised classification was to be used which minimize the need of radiometric correction

2.2 Geometric Corrections

Twenty-six control points were identified both in the image and on the official economic map in scale 1:10.000. After evaluating the control points root mean square error, 7 control points were removed and the 19 remaining control points, with a mean residual error of around 1.8 metres, were used for the geometric rectification and the resampling to 2 metres resolution with the cubic convolution method.
The image then consisted of 2000*2000 pixels, i.e. 4.000.000 pixels.

3. IMAGE CLASSIFICATION

After preliminary field checks, around 25 different classes were defined. It was possible to identify 53 homogeneous training areas in 22 classes. The training areas were identified on the screen and class signatures, using 2-5 training areas for each class, were calculated for the three different wavelength bands (see table I). After evaluation, new classes for water, fields and shadows were introduced to make it possible to classify the whole area.

A test maximum-likelihood classification was done, using the classifier in the ARIES III system. As a first result, the green (0.5-0.6 μm) wavelength band was found to be closely correlated to the red (0.6-0.7 μm) wavelength band. The green band was henceforth not used in the processing.
Next, it was found that the quota between the infrared (0.7-0.9 μm) and red wavelength band yielded the most distinct and homogeneous class signatures amongst several tested arithmetic combinations. This is the commonly accepted result when classifying images for vegetation purposes and the quota was used in the further processing without more investigation about the possibilities of using any other signature arithmetic characteristic.

TABLE I TRAINING AREA STATISTICS

Class	Nr. of pixels	Infrared		Red		Green		%
		Mean	Std	Mean	Std	Mean	Std	
Urtica	1 553	193.7	11.3	153.0	10.1	170.5	12.5	3.4
Ungrazed meadow	10 250	159.7	11.8	163.5	12.6	181.6	12.8	16.4
Juncus & Deschampsia	3 881	168.9	11.3	148.3	18.2	164.4	16.4	6.4
Anthriscus & Urtica	2 474	174.1	12.3	163.3	10.9	182.5	13.0	3.6
Grazed meadow	28 324	155.1	7.1	167.5	9.9	184.7	11.0	(16.4)
Humid grazed *Dactylis*	5 696	168.6	6.3	152.5	8.4	171.3	9.8	9.0
Alnus	723	126.4	46.4	108.9	30.2	128.5	26.3	4.2
Carex	677	184.8	15.1	151.7	8.8	168.6	11.0	(3.4)
Tall herb meadow	3 483	191.4	9.8	161.3	8.5	175.4	10.4	3.1
Tall herb meadow	473	193.6	11.3	160.9	10.2	177.0	11.8	(3.1)
Salix	5 363	160.2	40.0	139.3	30.1	153.5	30.8	3.9
Fagus	17 569	134.9	51.1	118.2	34.4	130.5	31.2	2.5
Quercus	276	150.6	40.0	119.2	27.1	128.5	25.0	(2.5)
Betula	5 992	120.2	46.2	98.9	30.0	122.5	26.3	(4.2)
Cladium	714	124.2	10.1	158.7	12.3	173.7	11.4	7.0
Fields	10 141	177.2	7.6	196.1	18.9	203.7	20.5	5.7
Picea	15 777	87.2	35.2	94.0	24.6	114.9	21.6	4.8
Pinus	6 905	70.0	34.1	92.9	24.2	110.4	20.4	5.2
Open water	2 014	31.0	8.0	119.8	6.9			7.2
Marsh	1 827	50.3	43.2	90.0	31.0			4.0

Values in brackets in the percentage column denotes merged classes, i. e. (4.2) for *Betula* means that *Betula* is merged with the class *Alnus* and that they have a total pixel percentage of 4.2.

To determine which classes that is possible to distinguish with some certainty, the closeness in the n-dimensional space has to be measured. The ARIES III system has the possibility of calculate the autocorrelation signature similarity measure that determines the correlation between any two classes.

The autocorrelation signature similarity measure between the different classes for the IR/Red quota were calculated as (Dipix, 1985):

$$AC = -10^{(\sqrt{\int (P1(x)^{*} P2(x))})}$$

or

$$AC = ((U-V)^{**} T (C^{-1})^{*} (U-V) + \log(\det(C)) - 1/2\log(\det(A)) - 1/2\log(\det(B)))$$

where A and B are the covariance matrices, C is (A+B)/2 and U and V are the mean vectors, for the signatures respectively probability distributions P1 and P2. **T indicates transpose of the matrix.

The probability distribution is defined as (n is the number of dimensions):

$$P(x) = \exp(-0.5((x-U)^{**}T)^{*}(A^{-1})^{*}(x-U)) / (((2\pi)^{n/2})^{*} \det(A)^{0.5})$$

The autocorrelation distances can approximatively be explained as follows:

$AC_{distance}$

2.0 ≈ 10% correlation
1.0 ≈ 30% correlation
0.5 ≈ 50% correlation
0 100% correlation, i.e. identical signature files

After evaluation of the autocorrelation distances (see table II) between the classes, checking for non-Gaussian distributions and comparing against the ground truth evaluation, several classes were merged. Sixteen classes remained for the final classification, amongst them one shadow class.

	Urt ica	Ung mea	Jun cus	Ant hris cus	Graz mead	Dac tylis	Aln us	Ca rex	Tall herb	mea dow	Sal ix	Fa gus	Que rcus	Bet ula	Clad ium	Fie lds	Pic ea	Pin us
Urtica	-																	
Ungrazed meadow	3.6	-																
Juncus	1.7	1.0	-															
Anthriscus	1.9	0.4	0.5	-														
Grazed meadow	6.6	0.2	2.3	1.3	-													
Dactylis	2.2	0.8	0.3	0.6	1.6	-												
Alnus	1.8	3.5	1.5	3.2	5.2	2.6	-											
Carex	0.2	1.9	0.6	1.1	3.4	0.8	1.7	-										
Tall herb -	0.4	3.0	1.4	1.3	6.1	2.2	2.5	0.4	-									
meadow	0.2	3.0	1.4	1.3	5.9	2.1	2.3	0.3	0.0	-								
Salix	0.9	1.3	0.5	0.9	2.2	1.0	0.4	0.6	0.9	0.8	-							
Fagus	1.4	2.5	1.0	2.1	3.7	1.8	0.2	1.1	1.6	1.5	0.2	-						
Quercus	1.5	4.0	1.5	3.6	5.9	2.8	0.3	1.6	2.4	2.2	0.4	0.2	-					
Betula	2.5	4.4	2.0	4.2	6.2	3.5	0.1	2.5	3.6	3.3	0.7	0.3	0.3	-				
Cladium	16.	2.9	9.2	6.0	3.3	8.6	6.6	8.3	17.	16.	3.6	5.0	8.3	7.0	-			
Field	4.0	1.9	1.9	1.7	3.5	3.4	4.8	3.1	3.0	3.2	2.5	3.6	4.6	5.7	9.3	-		
Picea	4.7	4.2	3.0	4.4	5.3	4.1	0.4	3.9	5.0	4.9	1.0	0.4	1.1	0.4	4.6	6.4	-	
Pinus	7.1	5.1	4.7	5.5	6.1	5.6	1.0	5.4	6.8	6.9	1.7	1.0	2.2	1.1	5.0	7.5	0.2	-

Table II describes the autocorrelation distance matrix between 18 different classes in a two-dimensional signature space. Shaded boxes marks classes to be merged.

Spruce (Picea) and fur (Pinus) could not be distinguished, as well as the different species of deciduous wood, alder (Alnus), oak (Quercus), beech (Fagus) and birch (Betula) which were to closely correlated to each other to be distinguished as separate classes. The only deciduous class that was possibly to keep as separate class was the Salix class, although it is, not surprisingly as they both belong to humid surroundings, quite closely correlated to the Alnus class.

Important to the result of this study was that seven different classes of meadows could be distinguish with sufficient significance.

The following 16 classes were possible to identify with the infrared/red quota in the digital air photo:

Open water, marsh, open sand, coniferous forest, deciduous forest, Cladium, Salix, Urtica-Carex, Juncus - Deschampsia, humid meadow, humid grazed Dactylis, tall herb meadow, fields, ungrazed - grazed meadow, Anthriscus, and unclassified pixels (mainly roads).

A mode filter with a widow size of 5x5 pixels was applied twice to clean up the classified image. This reduced the "noise" in the image considerably.

4. ARC/INFO - COVERAGE GENERATION

The classified image were exported from ARIES III via SVF-format to ARC/INFO, using first conversion programme, written at the Remote Sensing and GIS Laboratory and then ARC/INFO's vectorization tool gridpoly.

The vectorization tool generates a polygon topology map with boundaries between each class.
For visual understanding, see figure 1. Further description of the ARC/INFO system is available in the manuals (ESRI, 1989).

1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1
3	1	1	2	2	1	1	1
3	3	1	2	2	1	1	5
1	3	1	1	1	1	5	5
1	3	4	4	4	5	5	5
1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	2

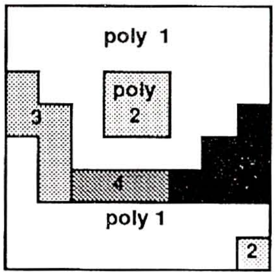


Fig. 1 - Classified raster image   Poligon ARC/INFO coverage.

Real map coordinates were regenerated to the ARC/INFO-coverage, to match the coordinates against the coordinates generated from the radio-tracking of the pheasants.

All shadow polygons were merged with its nearest tree/bush-class neighbour. The merging were done interactively on the graphics screen.
Generalization was necessary to do in order to reduce the amount of the data. Figure 1 points out the general problem with ARC/INFO:s vectorization tool. The polygon boundaries is defined strictly along the pixel borders and are therefore very "data expensive". The data size for the raster image was around 8 MByte and for the vectorized ARC/INFO-coverage around 22 MByte.
Next problem was that the smallest features, i.e. bushes and clumps of trees, were the most important ones. So the automatic generalization method provided in the system, weeding arc vertices with a specified weeding distances, would not work.
Instead, polygons belonging to a tree or bush class with an area smaller than 10 m^2 were selected and put to a temporary coverage. All other polygons with an area smaller than 10 m^2 were eliminated.
The border lines between the remaining polygons were generalized with a weedtolerance of 4 metres. The small bush and tree polygons in the temporary coverage were imported again.
This generalization process gave some minor incongruities, when the weeding changes the borderline in a large polygon into a small tree or bush polygon. Although, it was possible to automatically identify these problems and correcting them.
After the generalization process, the size of the ARC/INFO- coverage was diminished from 25.000 to 25.500 polygons and from 22 to 9 MByte in datasize.

Preliminary maps were plotted on a Versatec electrostatic plotter, both in colour and black-and-white and in various scales, and used for field checking.

After the field checking some minor corrections, mainly of the unclassified areas, were performed. The resulting vegetation map database was thereafter used for determine pheasant habitat selection and calculate statistics, as deswcribed in the next section.

5. METHODS FOR PHEASANT HABITAT DETERMINATION

5.1 Radio-tracking

The information in this chapter is taken from Göransson et al. (1990).

Two different types of radio-transmitters have been used. In 1984 and 1985 a light (3-5 g) tail-mounted transmitter but it had a short life expectancy and the birds often shed the feather where the transmitter was mounted. From 1986 a heavier (20 g) back- mounted transmitter have been used. It has been very reliable and since 1986 radio contact have been lost only with a few individuals during the breeding season.
With an accuracy of 10 m, 12.300 radio-fixes, have been recorded at least three times a week in 1984 and since 1985 every weekday.
All pheasants in the study area are annually trapped in December and their physical status (weight, spur lenght, age) are measured. They are ring-marked and blood-sampled to use DNA- fingerprinting techniques for parental identification.

Immigrants during the winter were also trapped. In all, 81 males and 101 females have been trapped and ratio-tracked. On the first weekday of April, a radio-transmitter was mounted on each individual and the birds were chased out of the enclosure.

5.2 Territorial Establishment

To determine the date when a bird was established in a territory or homerange, the ellipse model of Jennrich & Turner (1969) was used. The coordinates of the consecutively collected radio-fixes were introduced stepwise into the calculation formula. After one calculation, the set of coordinates was extended by the next consecutive radio-fix. The area of the ellipse was calculated to include 90% of the computed radio-fixes, i.e. the territory/homerange is defined operationally as the area in which a given bird was recorded 90% a time.

When plotting the calculated ellipse area for an individual pheasant we obtained a curve that increased at first, but as the bird became more stationary the curve declined. An individual was considered as established on his territory at the first day after the elliptic curve began to decline if it did not increase again for 2 consecutive days during the following 5 days. By this definition, any territorial period must last for at least 5 days of radio-tracking to be recognized. If the area of the ellipse began to increase monotonically at a later date, for 3 or more consecutive days, the pheasant was considered as having abandoned its territory on the first day of increase.

To describe the territorial/homerange borders we used the harmonic mean method of Dixon & Chapman (1980) as modified by Spencer & Barrett (1984). The method uses a grid, in our case of 50m, to find an area that contains 90% of the locations closest to each other. The area is enclosed by an isodistance line (isocline) with a metric value depending on the degree of the dispersion of the locations.

5.3 Statistical methods

The 1400 files containing the coordinate pairs describing the territorial borders for the pheasants, as computed with the methods described above, were imported from IBM PC to VAX/VMS.
Coverages were generated automatically, using especially written macros, for each of coordinate files, which have been converted to match the ARC/INFO generate format. These "territorial coverages" were used as clipping coverages against the map database. The output territorial coverage consists of the classified vegetation information for its territory.

Vegetation statistics concerning areas and parameters for each vegetation class were generated for each territory and transferred to IBM Excel for further analysis.

Class	Nr of polygons	total area	tot. perimeter	mean area	mean perim.	% of total area
Open water	4	3156,5	570,2	789,1	142,5	2,73
Marsh	27	6614,6	2057,4	245,0	76,2	5,72
Open sand	3	321,9	34,5	107,3	44,8	0,28
Coniferous forest	1	63,4	36,0	63,4	36,0	0,05
Deciduous forest	41	36745,9	7102,8	896,2	173,2	31,75
Cladium	0					
Salix	27	2541,3	1170,1	94,1	43,3	2,20
Urtica - Carex	46	4524,5	2166,6	98,4	47,1	3,91
Juncus - Deschampsia	49	3833,6	1861,4	78,2	38,0	3,31
Humid meadow	6	4065,8	7797,9	677,6	1299,6	3,51
Dactylis	43	5952,1	2256,9	138,4	52,5	5,14
Tall herb meadow	48	15807,7	4361,6	329,3	90,9	13,66
Field	0					
Grazed meadow	37	16563,4	3024,4	447,7	81,7	14,31
Anthriscus	27	12146,5	2805,8	449,9	103,9	10,50
Unclassified	13	3383,7	950,6	260,3	73,1	2,92
Total	372	115720,9	2913,3	311,1	7,8	100

Table III - Example of statistical output for one territoriali coverage (values in m^2).

6. BIOLOGICAL ANALYSIS

Homeranges were analyzed from data collected from 1984 to 1990. The distribution and amount of different habitat classes was correlated to the number of hatched pheasant chicks and individual fitness. Individual fitness was measured as the number of chicks produced by each individual for its life-span and which chicks survived at least 6 months. Insect density was sampled in different habitats for the post hatching period.

Analysing is still in a preliminary stage. The results presented here are more thorougly described in Göransson et al. (in press).

Successfully breeding females, i.e. females that hatched chicks, were analyzed for territorial behaviour during the egg-laying period (three weeks before incubation start) and during the brooding period (three weeks after hatching). The territory (homerange) was significantly smaller during the egg-laying ($x = 172\ m^2$, $sd = 238\ m^2$) than during the brooding period ($x = 555\ m^2$, $sd = 245\ m^2$). The area increase after hatching mainly consisted of the tall herb meadow and the deciduous forest habitats.

A comparison between successful (did hatch) and unsuccessful (did not hatch) with respect to habitat composition during egg-laying period, showed a significant difference. The homeranges of successful females were significantly richer in the marsh and deciduous forest habitat than were those of unsuccessful females ($p < 0.05$, $N = 36$; Spearman rank corr.).

During the first three week of brooding the chicks predominantly utilized the tall herb meadow habitat. Random samples on insect biomasse from the broods' forage areas were significantly larger ($x = 6.4\ gm^{-2}$) than were random samples from the tall herb meadow habitat as a whole ($x = 2.6\ gm^{-2}$).

CONCLUSIONS

This study demonstrates the possibilities of using an image analysing raster system and a geographical information vector system to integrate remote sensing data with radio transmitted data and several other data sources.

The maximum likelihood-classifiction method described made it possible to classify surprisingly many different (seven) meadow habitats. The habitats are, although, quite well separated in humidity and height of the vegetation. The reason why the different coniferous and deciduous tree species could not be distinguished is probably primarily due to the large amount of internal shadows in the tree crowns. The shadows is a more important feature for determining the spectral signature than the difference in spectral signatures between the individual species.

ARC/INFO:s powerful macro language facilitates the automatic import of 1400 files and of creating statistics automatically. Several kinds of maps have been and will be produced during the project, both as overlays to existing official maps and as separate maps.
Problems encountered were mainly to handle the size of the dataset and the lack of an operational supervised generalization process.
This article describes the initial work in the project, to create the vegetation database, to determine the territorial borders and to define other parameters measured during the investigation of the pheasants survival strategies.
The advanced and powerful geographical analysing techniques possible in ARC/INFO, as overlaying of different habitat and homerange types, connectivity and closeness analysis, perimeter effects etc., will then be used.

Next step will be to perform a principal components analyse on the different biotop statistics to define an objective, size- independent value for the homerange or territorial quality (in terms of fitness for the individual concerned). Other goals are to describe ecological connections between different species, i.e. quantify the territorial quality depending on the different species involved.

REFERENCES

DIPIX. 1985. "Aries System User's Manual".

Dixon, K.R. & Chapman, J.A. 1980. "Harmonic mean measure of animal activities areas". Ecology, 61, 1040-1044.

ESRI. 1989. "ARC/INFO vers. 5.0 manuals".

Goransson G., von Schantz, T., Froberag, I., Helgèe, A. & Wittzell, H. 1990. "Male characteristics, viability and harem size in the pheasant, Phasianus colchicus". Anim. Behav., 1990, 40, 89-104.

Goransson, G., Grahn, M., Oxenstierna A., & von Schantz, T., in press. "Territory quality in pheasants measured from air photos". Trans. XXth Int. Congr. Game Biol., Goddolo Hungary 1991.

Jennrich, R.I. & Turner, F.B. 1969. "Measurement of non-circular home range". J. theor. Biol., 22, 227-237.

Spencer, W.D. & Barrett, R.H. 1984. "An evaluation of the harmonic mean measure for defining carnivore activity areas". Acta zool. fenn., 171, 255-259.