

How complete are the species lists of breeding birds in censuses of large areas?

OLLI JÄRVINEN & RISTO A. VÄISÄNEN

JÄRVINEN, O. & R. A. VÄISÄNEN 1977: *How complete are the species lists of breeding birds in censuses of large areas?* — *Ornis Fennica* 54: 160—165.

The number of land bird species breeding in 54 squares (100×100 km) of the Finnish uniform grid was estimated from distribution maps and compared with the number observed in line transect censuses. The ratio of the two numbers (*SC/S*) was used as the dependent variable in a stepwise regression analysis with ten independent variables. The most important variable was the logarithmic number of pairs observed in a square. The next variable to enter into regression was the number of species having their range boundary in the square. These two independent variables accounted for 77 % of the variation in *SC/S*. The model implies that more working hours are needed in the north than in the south in order to observe the same proportion of species, because the density of birds decreases sharply northwards in Finland. Applications of the model in faunistic work, Atlas studies, conservation and land use planning and theoretical studies of ecosystem dynamics are discussed.

Olli Järvinen, Dept. Genetics, University of Helsinki, P. Rautatiekatu 13, SF-00100 Helsinki 10, Finland

Risto A. Väisänen, Zoological Museum, University of Helsinki, P. Rautatiekatu 13, SF-00100 Helsinki 10, Finland

Introduction

Species lists of breeding birds are compiled for a variety of purposes. In this paper we report methodological observations made when the reliability and completeness of such lists was assessed with the aid of a stepwise regression analysis of Finnish line transect data.

Material and methods

The data used were obtained in line transect censuses of breeding birds in Finland. The material was divided into 54 squares, each 100 × 100 km, and the

study area thus covers the whole of Finland, most of northern Norway and small parts of northern Sweden and north-western Russia (Leningrad area). The census method has been described in detail by JÄRVINEN & VÄISÄNEN (1976b). Here, it is sufficient to note that the different habitats are represented in their true proportions in the line transect censuses.

The material comprises nearly 90 000 observations, or, on average, more than 1600 pairs of land birds per square. In addition, maps for all the land bird species breeding in the study area were constructed on the basis of those published by HAFTORN (1971). As HAF-

TORN's detailed maps are not available for all the species breeding in the study area, other recent standard handbooks were also used. In our maps the occurrence of each species in the 100 km × 100 km squares of the Finnish uniform grid system (HEIKINHEIMO & RAATIKAINEN 1971) was recorded as follows:

- 0 = absent
 1 = boundary of range running through the square
 2 = breeding throughout the land area of the square

The number of species breeding in a square (either code 1 or 2) is denoted by *S*. Those breeding in only part of a square are called "boundary species" (code 1), and their number is denoted by *SB*. Finally, the number of species observed in a square in the line transect censuses is denoted by *SC*.

The data were analysed with the aid of a stepwise regression analysis. The independent variables used were (cf. also Table 1):

1—2) Coordinates of the square in the Finnish uniform grid system (south-north, east-west).

3) Density of birds (pairs/km², from line transect data).

4—5) Diversity of birds (measured with the Shannon index *H'*, but corrected using the formula given by HUTCHESON 1970) and $\exp(H')$ or "the number of equally common species".

6—7) Length of transect censused in a square and its logarithm.

8—9) Number of pairs censused and its logarithm.

10) *SB*.

All the logarithms used in this paper are natural logarithms,

The dependent variable chosen for closer investigation in a stepwise regression analysis was defined as 100 *SC/S*, that is, the percentage ratio of land bird species censused to the true number derived from distribution maps. The analysis method (see e.g. DRAPER & SMITH 1966:171—172) may be described briefly as

TABLE 1. Averages, standard deviations and ranges of the most important variables used in the stepwise regression model described in the text. The data stem from line transect censuses made in 54 squares (100 km × 100 km).

Variable	$\bar{x} \pm S.D.$	Range
<i>S</i> (see text)	111.1 ± 17.2	50—151
<i>SC</i> (see text)	65.9 ± 17.1	29—97
100 <i>SC/S</i>	58.7 ± 9.8	30.2—75.2
<i>SB</i> (see text)	26.1 ± 14.1	11—68
Density (pairs/km ²)	144.5 ± 72.0	31—292
Diversity (<i>H'</i>)	3.14 ± 0.29	2.37—3.65
$\exp(H')$	24.1 ± 6.3	10.7—38.4
Transects (km)	52.9 ± 23.7	22—154
Number of pairs (<i>N</i>)	1658 ± 962	195—4220
$\ln N$	7.23 ± 0.66	5.27—8.35

follows. The computer program starts with a number of independent variables (in this case, variables 1—10). First a correlation matrix is calculated using all the variables. The independent variable most highly correlated with the dependent variable is then entered into regression. Using the partial correlation coefficients, the program selects, as the next variable to enter regression, that independent variable whose partial correlation with the dependent variable is highest. Further variables are then entered into regression in the order determined by the partial correlation coefficients. Since a variable which may have been the best single variable to enter at an early stage may, at a later stage, be superfluous because of the relationships between it and other variables now in regression, the contribution made by each variable is re-examined at every stage of the regression. Any variable which makes a non-significant contribution can thus be removed from the model. In short, the formulae obtained are ordinary regressions of the dependent variable upon several independent variables, but the model is constructed stepwise.

Since both *S* and *SB* were derived from the distribution maps, it may appear that circular reasoning is involved when *SB* is used to explain variation in a function of *S* (the dependent variable 100*SC/S*). However, these variables correlate only weakly (they share 8.8 % of their variance).

Results

The first variable to enter into regres-

sion is the logarithmic number of the observations ($\ln N$):

$$100SC/S = -28.1 + 12.0 \ln N \quad (1)$$

The model accounts for as much as 66.1 % of the variance ($r = 0.81$) in $100SC/S$, and it is highly significant ($F = 101.4$, d.f. = 1; 52). The relationship (1) implies that the most important single variable contributing to the completeness of the species lists is $\ln N$. The number of observations as such, N , gives a considerably poorer result ($r = 0.75$; 56.0 % of the variance explained). As shown in the following tabulation, the completeness of the species list increases relatively slowly with increasing census effort:

N	$100SC/S$
100	27.2
1000	54.8
10000	82.4

In other words, when census effort ($= N$) increases 10-fold, the completeness of the species list increases by 27.6 percentage units. The formula should not be extrapolated beyond the range of N shown in the tabulation.

The second variable to enter into regression was SB :

$$100SC/S = -4.34 + 9.65 \ln N - 0.25 SB \quad (2)$$

This model, with two independent variables, explains 77.0 % of the variance in $100SC/S$, and it is highly significant ($F = 85.5$, d.f. = 2; 51). Of the two variables, $\ln N$ is the more important one. If $\ln N$ is deleted, 34.4 % of the variance explained is lost, while deletion of SB yields the model (1). The two independent variables are correlated ($r = -0.44$), and the model therefore poses certain problems of interpretation: when the value of one of the independent variables is changed by one unit, the value of the dependent

variable does not change by exactly the amount indicated by the regression coefficient, since, owing to the correlation, the value of the other independent variable is also changed. However, as the two independent variables do not share more than 19.1 % of their variance, we believe that the model is not seriously affected by the correlation between them. This correlation arises from the fact that fewest pairs were observed in those squares where SB is relatively high. Indeed, if the most extreme observation is deleted, only 12.5 % of the variance is shared by $\ln N$ and SB .

The decrease of $100SC/S$ with increased SB ($r = -0.65$) implies that the completeness of the census tends to decrease with an increase in the number of species occurring in only part of a square. This finding is of methodic importance — for example, equal degrees of completeness can be achieved in two squares only if more birds are censused in the square with the greater proportion of species occurring in only part of the square.

Model (2) is illustrated by two maps. Fig. 1 shows the completeness of the species lists ($100SC/S$) obtained from 25-km line transect censuses (the data for this map have been calculated from the number of pairs observed per km in each square). Fig. 2 gives the length of transect that has to be censused in order to observe 50 % of the species breeding in the square. The maps show clearly that equal census effort by no means guarantees equal completeness — as the density of birds decreases considerably from south to north, a much shorter transect gives the same percentage of species in the south as in the north. The number of "boundary species" increases northwards, which increases the difference in $100SC/S$ between the south and the north. However, the maximum ef-

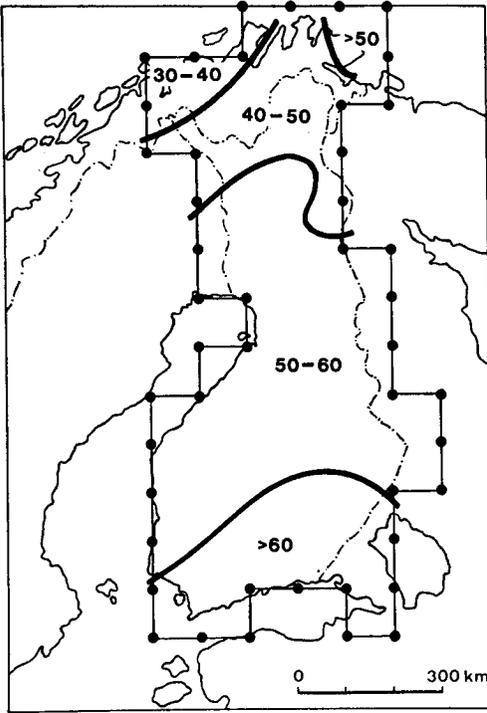


FIG. 1. The percentage ratio of land bird species censused on 25-km line transects in 100-km squares to the true square numbers derived from distribution maps. The percentage has been calculated from the regression model (2). The boundaries of the study area are shown by dots.

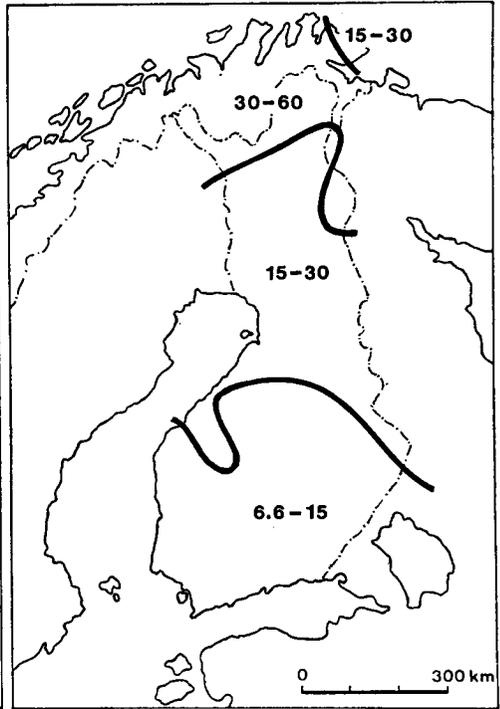


FIG. 2. The length of line transects (in km) which must be censused in order to observe 50% of the breeding land bird species in a 100-km square, calculated from the regression model (2).

fect of the latter trend is $0.25 \times 57 = 14.2$ percentage units, since the minimum *SB* is 11 and the maximum 68. Thus a greater proportion of the north-south difference is due to the number of birds observed per km.

Discussion

Both the independent variables entered into regression are plausible: the number of species observed should increase with the logarithmic number of observations (e.g. JÄRVINEN & VÄISÄNEN

1976a) and "boundary species" are presumably less likely to be encountered in the censuses than those occurring in the whole square.

The further steps in the analysis were not described above, because they increased the total variance explained relatively little. We also preferred a simple model to a more complicated one. After all, anything can be explained if a sufficient number of independent variables are incorporated into the model, even if the variables are random numbers!

The application of our result in other

contexts depends on the similarity of the methods of compiling species lists. In line transect censuses, the census-maker walks along a predetermined transect and writes down the observations made during the census. For one km of line transect censuses, about 45—60 minutes are required. Thus the census-maker does not use his time in searching for the nests of the birds or in visiting attractive bird localities in the neighbourhood of the transect. (Visiting such localities presumably increases the number of observed species more rapidly than the rate found in line transect censuses.)

Our results may have applications in at least the following cases:

Faunistic lists. As seen in Fig. 1, 25 km of line transects yields about 60 % of the species in southern Finland and about 40 % in Lapland; 25 km can be equated with 20 hours of field-work in the morning. Such knowledge may be valuable when faunistic lists published by earlier authors are evaluated.

Atlas studies. Atlases of breeding birds (DYBBRO 1976, SHARROCK 1976, YEATMAN 1976) contain information on the birds breeding in 10-km squares (or larger or smaller areas). In Atlas work, the field-worker tries to establish whether the birds observed really breed in the square studied or not. Thus the time used per bird is much greater than in our censuses. Atlas work also differs from our censuses in the time of day at which the work is performed (line transect censuses are made early in the morning, when the detectability of birds is best; cf. JÄRVINEN et al. 1977) and in the possibility of highly selective excursions by the Atlas worker. However, our model is probably roughly applicable to evaluating the completeness of Atlas work in an area.

Conservation studies. The number of

species, especially that of the endangered species, has been used by BEZZEL & RANFTL (1974) in land use planning. The results presented here are applicable as long as the areas chosen for study are not very small. (As a guess, we suggest that 10-km squares are the minimum for the present model.)

Ecosystem dynamics. Dynamic aspects of island ecosystems are often studied by comparing two species lists compiled at different times with each other. Counting the number of extinctions and immigrations is a way of estimating the turnover in the island communities. As revealed by the present model, the effort needed to obtain a full list of species may be considerable on a large island, since new species are added in proportion to $\ln N$.

As the present analysis is evidently restricted to relatively large areas, we recommend that the completeness of species lists should also be studied in small areas (e.g. 10 ha or 1 km²). A preliminary estimate can be made from the simulations of JÄRVINEN & SAMMALISTO (1973). These authors showed that about 60 % of the pairs in a community chosen at random included more than 80 % of the species. The communities studied were small (6 and 20 species), and, accordingly, the number of species found in smaller areas may approach completeness relatively rapidly (see also JÄRVINEN & LOKKI 1978).

Acknowledgements. We are grateful to Prof. S. Haftorn for allowing us to use the original maps of Norges fugler.

Selostus: Kuinka täydellisiä ovat pesimälinnuston laskennoissa kootut lajiluettelot?

1. Pesivien maalintulajien lukumäärä arvioitiin käsikirjojen levinneisyyskarttojen perusteella yhtenäiskoordinaatiston 54 ruudusta (100

×100 km). Saatua lajilukua (*S*) verrattiin samoissa ruuduissa tehtyjen linjalaskentojen lajilukuun (*SC*).

2. Askeltavassa (*stepwise*) regressioanalyysissä käytettiin kymmentä riippumatonta muuttujaa (taul. 1) selittämään riippuvassa muuttujassa 100SC/*S* havaittua vaihtelua.

3. Analyysissä päädyttiin malliin $100SC/S = -4.34 + 9.65 \ln N - 0.25 SB$, missä $\ln N$ = ruudussa havaitun parimäärän logaritmi ja *SB* = niiden lajien lukumäärä, joiden levinneisyysraja kulkee ruudun kautta. Malli on erittäin merkitsevä ($P < 0.001$) ja selittää 77 % riippuvan muuttujan vaihtelusta. Muuttujista $\ln N$ on selvästi tärkeämpi kuin *SB* (näiden poistaminen mallista vähentää selittävyyttä 34 ja 11 %-yksikköä). Malli siis ennustaa, että lajiluettelo tulee varsin hitaasti täydellisemmäksi (suhteessa havaitun parimäärän logaritmiin) ja että luettelon täydentämistä haittaa jonkin verran, mikäli kovin monen lajin levinneisyysraja sattuu alueelle.

4. 25 linjalaskentakilometriä tuottaa n. 60 % 100 km:n ruudun lajeista Etelä-Suomessa, mutta n. 40 % Pohjois-Suomessa ja -Norjassa (kuva 1). Jotta havaittaisiin noin puolet ruudun pesivistä maalintulajeista, Etelä-Suomessa täytyy tehdä n. 10 km laskentoja, mutta karkeasti viisinkertainen määrä on tarpeen tutkitun alueen pohjoisosissa (kuva 2). Jyrkkään eroon vaikuttaa pääasiassa pesimälinnuston tiheyksien selvä väheneminen pohjoiseen.

5. Mallin sovellutuksia faunistisissa yhteyksissä, atlasyössä, maankäytön suunnittelussa ja ekosysteemien dynamiikkaa koskevista teoreettisista töistä käsitellään lyhyesti. Tutkimuksen tulosten laajentaminen koskemaan pieniä alueita olisi tärkeää.

References

- BEZZEL, E. & H. RANFTL 1974: Vogelwelt und Landschaftsplanung. Eine Studie aus dem Werdenfeler Land (Bayern). — Tier und Umwelt N.F. 11/12:1—92.
- DRAPER, N. R. & H. SMITH 1966: Applied regression analysis. — New York.
- DYBBERO, T. 1976: De danske ynglefugles udbredelse. — Copenhagen.
- HAFORN, S. 1971: Norges fugler. — Oslo.
- HEIKINHEIMO, O. & M. RAATIKAINEN 1971: The recording of localities of biological finds in Finland. — Ann. Entomol. Fennici 37(1a):1—30.
- JÄRVINEN, O. & J. LOKKI 1978: Indices of community structure in bird censuses based on a single visit: Effect of variation in species efficiency. — Ornis Scandinavica 9 (in press).
- JÄRVINEN, O. & L. SAMMALISTO 1973: Indices of community structure in incomplete bird censuses when all species are equally detectable. — Ornis Scandinavica 4:127—143.
- JÄRVINEN, O. & R. A. VÄISÄNEN 1976a: Species diversity of Finnish birds, II: Biotopes at the transition of taiga and tundra. — Acta Zool. Fennica 145:1—35.
- JÄRVINEN, O. & R. A. VÄISÄNEN 1976b: Finnish line transect censuses. — Ornis Fennica 53:113—117.
- JÄRVINEN, O., R. A. VÄISÄNEN & Y. HAILA 1977: Bird census results in different years, stages of the breeding season and times of the day. — Ornis Fennica 54:108—118.
- SHARROCK, J. T. R. 1976: The Atlas of breeding birds in Britain and Ireland. — Aylesbury.
- YEATMAN, L. 1976: Atlas des oiseaux nicheurs de France de 1970 à 1975. — Paris.

Received March 1977