Purported correlations between breeding productivities of arctic geese: a statistical artifact without implication for waders

L.G. Underhill

Underhill, L.G. 1993. Purported correlations between breeding productivities of arctic geese: a statistical artifact without implication for waders. *Wader Study Group Bull.* 70: 19-21.

L.G. Underhill, Avian Demography Unit, Department of Statistical Sciences, University of Cape Town, Rondebosch, 7700 South Africa, and Schweizerische Vogelwarte, Sempach, CH-6204 Switzerland.

INTRODUCTION

De Boer & Drent *(Wader Study Group Bulletin* 55: 11-17, 1989) tried to show that there were unexpectedly many correlations in breeding performance between arctic breeding geese. In this note, it is shown that their statistical argument was flawed. Therefore, the notion that that there is a worldwide synchrony in breeding productivity of arctic goose populations must be regarded as unproven.

De Boer & Drent (1989) attempted to demonstrate correlations between the breeding productivity of 21 populations of arctic-breeding geese. They alleged that there were significant ($P<0.05$) correlations between 58 of the 210 (=21 x 20/2) possible pairwise comparisons, instead of the 11 (=0.05 x 210) that they claimed could be expected by chance alone.

There are, however, several flaws in their arguments.

First (and worst), the authors have fallen into one of the basic statistical traps, the simultaneous inference situation (see, for example, Beal & Khamis 1991). Therefore, the logic behind the assertion that 11 significant correlations could be expected by chance alone is wrong. The situation here is analogous to the classic simultaneous inference situation, the test of whether the means of k populations are equal: analysis of variance (ANOVA) is used to control the overall level of significance at, say, 5%. Doing all possible pairs ($n=k(k-1)/2$ of them) of two sample t-tests at the 5% significance level is wrong. Likewise, looking at the correlations between all possible pairs of variables at the 5% level is wrong. It is wrong for the same reason: the overall significance level is no longer 0.05. What the "true" significance level is is impossible to determine. Whatever it is, it is greater than 0.05 and less than n x 0.05, where n is the number of tests performed. So de Boer & Drent can only claim is that their overall significance level lies somewhere above 0.05 and below 10.5 (=210 x 0.05) (but because probabilities have an upper limit of one, their significance level is between 0.05 and one). One way out of this problem is to use a conservative approach, the so-called Bonferroni method, and to divide the required signficance level by the number of tests being performed (Beal & Khamis 1991). In this case, the correlations would each need to be tested at the 0.05/210= 0.000238 significance level so that the overall significance level is definitely less than 0.05.

Secondly, econometricians know that simply cross-correlating pairs of time-series to find relationships can easily result in misleading and spurious correlations, because the observation in one year need not be independent of the observation in the next. The correct approach is to use the time-series modelling techniques of Box & Jenkins (1976), removing the time dependence between successive observations using autoregressive or moving average models, and then to cross-correlate the so-called white noise residuals.% (a term emanating from electrical engineering, not South African politics, this being written on Referendum Day, 17 March 1992!). For examples of ecological applications of the Box-Jenkins approach, see Shannon *et al.* (1988) and Peach *et al.* (1991). The folklore in econometrics is that the significance of correlations between raw time-series frequently evaporates when properly analysed.

Thirdly, the breeding productivities are, logically, not independent to begin with. Many of the species considered share ranges, and *a priori* can be anticipated to show positive correlations between their productivities. Thus, it is not surprising that Greylag and Pinkfooted Geese *Anser anser* and *A. brachyrhyncus*, both breeding in Iceland, showed positive correlations, and that

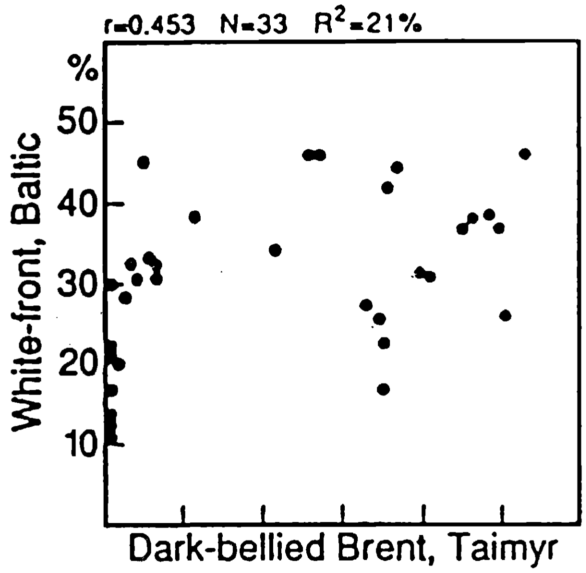

Figure 1

63% of the North American goose populations were positively intercorrelated.

Fourthly, one of the six plots in Figure 2 of de Boer & Drent, reproduced here as Figure 1, and purporting to exhibit a linear relationship, is nominally "significant" at $P<0.01$. However, the relationship appears parabolic, not linear. Even $P<0.01$ should be overturned by common-sense, if it produces biologically meaningless results. This is a manifestation of what is becoming known to statisticians as the "*P*-value culture", which is a consequence of the success with which statisticians have brainwashed editors of journals into rejecting papers unless $P<0.05$.

Fifthly, there is some confusion between one-sided and two-sided tests. Only one correlation was significantly negatively correlated, suggesting that a two-sided alternative was used, so that both large negative and large positive correlations would be seen as significant. If a two-sided alternative was used, then two of the six examples given in Figure 2 of de Boer & Drent are not, in fact, significant at the 1% level, as the caption to the figure claims.

On a related issue, there is one slightly misleading . statement in the paper. It is stated that "only one correlation had a negative coefficient". The reader is encouraged to infer that "only one correlation [of the 210] had a negative coefficient". In fact, what was intended is that "only one correlation [of the 58 that were significant at the 5% level] had a negative coefficient". In other words, there were many correlations, and not just one, that were negative.

So, with all these considerations taken into account, the statistical argument underpinning the paper is destroyed. In particular, the notion that there some kind of a worldwide synchrony in breeding productivity of arctic goose populations is unproven, Editors of journals should be alerted to look out for citations to this paper that seek to perpetuate the myth. It has not been shown.

De Boer & Drent (their Figure 4) plotted date of snow-melt against breeding success of Barnacle Geese *Branta leucopsis.* The correlation is significant at the $P<0.001$ level. The regression line is interpreted in the caption to the figure as showing that for every 10 days delay in snow-melt, breeding productivity drops 9.6%. There is no problem with this as a descriptive statement. However, it is a short step to conclude that date of snow-melt, i.e. the "weather", is the causal factor driving the process. The problem here is another basic statistical trap — the existence of correlation does not prove cause and effect. De Boer & Drent don't say this, but they come tantalizingly close to allowing the unwary reader to draw the conclusion that delays in snow melt actually cause breeding productivity to drop 9.6% per 10 days.

Owen & Norderhaug (1977), the paper from which the bulk of the data for the figure was taken, provide an alternative explanation, in terms of the predation hypothesis, for this correlation. When the thow is late, there is a persistence of ice-bridges that allow foxes to cross to the breeding islands, a situation that has been observed by Norderhaug (1970) to result in almost complete failure of breeding.

Three issues raised by Summers & Underhill (1987) which de Boer & Drent fail to address, and which appear to be critical to achieve their goal of demonstrating that "predation is an unlikely single cause of the fluctuations of breeding success" and showing that "body condition is probably the most important causal factor" are:

1. How do they explain, in terms of the body condition on arrival hypothesis and/or the timing of snow-melt hypothesis, the strong bimodality in breeding pro ductivity of Dark-bellied Brent Geese *Branta b. bernicla*? The geese tend to breed either very well, or poorly, with few years of intermediate production.

2., How do they explain, in terms of these hypotheses, the observed c. 3-year cycles in breeding success for this species?

3. How do they explain, in terms of the body condition

on arrival hypothesis, the bimodality in breeding success of waders breeding in the northern Taimyr Peninsula, and the general concordance between the breeding success of Brent Geese and these waders?

For the third question, the different strategies of geese and waders on arrival on the breeding grounds need to be noted. Brent Geese are determinate layers (Barry 1962). If the thaw does not come early, egg follicles become atretic, and productivity reduced. If the eggs are predated, replacement clutches are not possible. Geese are grazers, and no substantial amounts of new growth are available until several weeks after the thaw. Thus, the most of the breeding operation, almost until hatching, has to be undertaken on body reserves brought with them (Newton 1977).

In contrast, some invertebrate food for waders is available during the thaw (and possibly even beforehand) (Green *et al.* 1977; Uspenskii 1984 p. 48, Aitchison 1989; Koponen 1989; pers. obs). Mating takes place on the breeding grounds, clutch size is not reduced if the thaw is late, and replacement clutches may be laid if eggs are lost to predators in the early stages of incubation. Thus the breeding productivity of waders is not constrained by body condition on arrival in quite the same way it is for geese. It is not related, in the more or less linear way observed for geese, to timing of the thaw (except, of course, if the thaw is so late that the snow-free period is too short for incubation and fledging, breeding may be a total failure, a step relationship).

Thus the editorial comment which introduced de Boer & Drent's paper, that "their findings, although they mainly concerned geese, are of great relevance to wader-workers", was somewhat naïve. Waders and geese have totally differing breeding strategies, at least until incubation is underway. If correlations between their breeding productivity are observed, the logical place to start looking for causality is at the stage when they are doing things in common. This is the strength of the predation hypothesis put forward by Summers & Underhill (1987), it operates when both geese and waders are incubating their eggs and tending their young.

ACKNOWLEDGEMENTS

I am grateful to the Foundation for Research Development, the University of Cape Town and the Scheizerische Vogelwarte for support, and to the Russian Academy of Sciences for the opportunity to visit the Taimyr Peninsula.

REFERENCES

Aitchison, C.W. 1989. The ecology of winter-active collembolans and spiders. *Aquilo Series Zoologica* 24: 83-89.

Barry, T.W. 1962. Effects of late seasons on Atlantic Brant reproduction. *J. Wildl. Manage.* 26: 19-26.

Beal, K.G. & Khamis, H.J. 1991. A problem in statistical analysis: simultaneous inference. *Condor* 93: 1023-1025.

Box, G.E.P. & Jenkins, G.M. 1976. *Time-series Analysis: Forecasting and Control*, revised edition. Holden Day, San Francisco.

De Boer, W.F. & Drent, R.H. 1989. A matter of eating or being eaten? the breeding performance of arctic geese and its implications for waders. *Wader Study Group Bulletin* 55: 11-17.

Green, G.H., Greenwood, J.J.D. & Lloyd, C.S. 1977. The influence of snow conditions on the date of breeding of wading birds in north-east Greenland. *Journal of Zoology*, London 183: 311-328.

Koponen, S. 1989. Spiders (Araneae) on snow surface in subarctic Lapland. *Aquilo Series Zoologica* 24: 91-94.

Owen, M. & Norderhaug, M. 1977. Population dynamics of Barnacle Geese *Branta leucopsis* breeding in Svalbard, 1948-1976. *Ornis Scandinavica* 8: 161-174.

Newton, I. 1977. Timing and success of breeding in tundra-nesting geese. In: Stonehouse, B. & Perrins, C.M. (Eds). *Evolutionary Ecology*. London: MacMillan: 113-126.

Norderhaug, M. 1970. The present status of the Barnacle Goose (*Branta leucopsis*) in Svalbard. *Norsk Polarinstitutt Arbok* 1968: 24-35.

Peach, W., Baillie, S.R. & Underhill, L.G. 1991. Survival of British Sedge Warblers *Acrocephalus schoenobaenus* in relation to west African rainfall. *Ibis* 133: 300-305.

Shannon, L.V., Crawford, R.J.M., Brundrit, G.B. & Underhill, L.G. 1988. Responses of fish populations in the Benguela ecosystem to environmental change. *Journal du Conseil International pour l'Exploration de la Mer* 45: 5-12.

Summers, R.W. & Underhill, L.G. 1987. Factors related to breeding production of Brent Geese *Branta b. bernicla* and waders (Charadrii) on the Taimyr Peninsula. *Bird Study* 34: 161-171.

Uspenskii, S.M. 1984. *Life in High Latitudes: A Study of Bird Life.* Russian Translations Series 18. Rotterdam: A.A. Balkema.