Review of

Thames Basin Heath Special Protection Area (SPA) Study Final Report

(Environmental Dimension Partnership)

Dr Jennifer Gill

School of Biological Sciences

University of East Anglia

Norwich NR4 7TJ

j.gill@uea.ac.uk

The Thames Basin Heaths (TBH) Special Protection Area is the subject of an English Nature (now Natural England, NE) Draft Delivery Plan, relating to guidelines for planning permission for future residential developments. The ***Environmental Dimension Partnership*** (EDP) consultancy was commissioned to undertake a review of these guidelines. The NE Draft Delivery Plan uses information from a scientific study of the impact of urban development in Dorset heaths on nightjars, ***Caprimulgus europaeus*** (Liley & Clarke 2003), one of the three Annex 1 species occurring on the Thames Basin Heaths SPA. The EDP review includes an ecological study of the influence of urban development on the distribution of nightjars, woodlarks, ***Lullula arborea***, and Dartford warblers, ***Sylvia undata***, largely carried out through an attempt to replicate the methods adopted by Liley & Clarke (2003). In this report, I review the section of the EDP report dealing with the ecological study and the replication of the Liley & Clarke (2003) approach. This report therefore relates solely to the scientific framework within which the EDP study was carried out, and to the conclusions and interpretations of the research.

The EDP report explores the evidence for urban development impacting on numbers of nightjars within the Thames Basin Heaths, firstly by applying the methods of Liley & Clarke (2003) to sites within the TBH, and secondly by relating the distribution of the three bird species to measures of habitat type and numbers and types of paths within squares within each SPA. However, there are four major problems with the study which mean that, in the present form, the EDP research would not be acceptable for publication in a peer-reviewed scientific journal. These problems are summarised here and detailed in the point-by-point description below.

1. The first section of EDP study is based on only 13 sites, which severely limits the potential for statistical and biological analysis of these data, and the interpretation of the resulting relationships.

2. The areas within which bird densities are calculated are very likely to include unsuitable habitats, and hence the estimates of bird densities used in the analyses are likely to be misleading.

3. The extension of the nightjar analysis to the two other species, woodlark and Dartford warbler, does not take account of the other factors influencing the distribution and dynamics of these species.

4. The second section of the EDP study uses inappropriate univariate statistics to address a problem which requires multivariate statistics.

Details of specific issues

4.50. ***Small sample size issues in multiple regression.***

Multiple regression analysis is, in theory, the correct way to analyse these data, because the relative effect of each predictor variable on nightjar density can be assessed. However, multivariate analyses generally require ***at least*** 10 datapoints per predictor variable, in order to be able to assess how much of the variation in the data can be assigned to each variable (e.g. Tabachnick & Fidell 1996). In this study there are only 13 sites, which is too few to assess the effect of more than one variable. A multiple regression analysis incorporating three predictor variables would require a minimum of 30 sites. The Liley & Clarke (2003) study used 36 sites, and so a model with three predictors (patch area, urban cover and woodland cover) was appropriate. In the EDP study, 13 sites was not a sufficient number for this analysis.

4.46. ***Problems with covariance among predictor variables.***

In Figures 2 & 3, the relationships between nightjar density and the two predictor variables, urban development cover within 500m and woodland cover within 500m, are assessed using two separate correlation analyses. These univariate analyses suffer from the problem that the covariance between predictor variables is not assessed. Covariance is important because it can result in variables being falsely identified as significant or non-significant predictors of species presence, because they happen to co-occur with another variable. For example, Liley & Clarke (2003) showed that nightjar densities in Dorset declined with increasing levels of urban development, and increased with increasing woodland cover. These relationships were not apparent in the 13 sites in the EDP study because, for example, several sites had low urban cover (<20%) but low also nightjar density (<0.07 per ha) (Figure 2). However, Figure 3 shows that the site with the lowest nightjar density has both low urban cover and low woodland cover. Thus the reason for the low nightjar density in such a site could be the low woodland cover, irrespective of the urban cover. Similarly, low densities in some sites may be for reasons such as the habitat being unsuitable, the site being too isolated from other nightjar populations etc. A multivariate analysis incorporating all the factors likely to influence bird numbers would be necessary to accurately assess these relationships. As described above, this method would require far more than 13 sites.

4.16. ***The assumption of a normal distribution in the analysis of nightjar count data.***

The Poisson distribution is the appropriate distribution to use when data are non-negative whole numbers (in this case, counts of nightjars). Using the normal distribution to model such data could result in the model predicting negative numbers of nightjars. The normal distribution can generally only provide a good approximation to Poisson-distributed data when the average of the data (ie the nightjar counts) is greater than 10, as the shapes of the Poisson distribution and the normal distribution are then sufficiently similar for the normal approximation to be used. For averages less than 10, the Poisson distribution is truncated at zero, whereas the normal distribution will include negative count data. Whether or not the use of the normal distribution was appropriate in this case therefore depends on whether the average number of nightjars on the sites exceeded 10.

4.5 (iv). ***Estimating bird densities***.

A common problem in ecological studies is accurately measuring the area of habitat available for a species to use. The Liley & Clarke study used the CEH Dorset Heathland Survey to identify patches of suitable heathland which were then specifically surveyed for nightjars. The patches were therefore selected on the basis of the heathland habitat that they contain. The estimates of bird density used in that study were therefore very likely to reflect true nesting densities of nightjars on Dorset heathland. In the EDP studies, the sites are the SPA's, and nesting densities are estimated from the number of territories of each species within each SPA component, divided by the estimated area of heathland (identified from aerial photographs) within each SPA component. There are two problems with this approach: 1. sites may largely comprise habitats unsuitable to the birds, thus their inclusion may seriously distort the results of analyses, 2. if some of the heathland within a site is not suitable for nightjars, woodlarks or Dartford warblers, then the estimates of densities will be incorrect, making all subsequent analyses involving these data impossible to interpret.

(Note: the data in Appendix 2 show heathland as only being present in 3 of the 13 sites (and no woodland within 500m of any of the sites), which is presumably a mistake).

4.51. ***Predicting nightjar numbers from the Liley & Clarke model***

The lack of fit between the actual numbers of nightjars on these sites and the numbers predicted by the Liley & Clarke (2003) model can result either from the model not being applicable to the TBH ***or*** from the estimates of numbers of nightjars, woodland cover and urban cover in the TBH being inappropriate to the model. Given the limitations of the EDP data described above, it is very unlikely that the true relationships between nightjar numbers and urban and woodland cover in the TBH have been identified with this small number of sites. Thus, a larger number of sites in which breeding densities and the influence of woodland and urban cover had been measured more accurately, may well have resulted in a better fit between model predictions and actual values.

4.52. ***Inappropriate use of correlation analysis.***

The analysis of the differences between observed and predicted densities in relation to SPA area uses a parametric correlation. However, this type of correlation analysis assumes that the data are evenly distributed along both axes (homoscedasticity), which is clearly not the case here. When the area of the SPA component is less than 300 ha, the difference between observed and predicted values ranges between +1 and -1. However, at higher SPA areas, the differences only vary between approximately +0.2 and +0.7. The correct analysis of these data is a rank correlation (e.g. Spearmans rank correlation, in which the data are ranked on both axes and the ranks are compared), which would show no significant association.

4.53. ***Analyses of woodlark distribution.***

The analysis of the effect of urban and woodland cover on woodlark distribution suffers from all of the statistical problems outlined above for nightjars. In particular, the estimates of density based solely on estimated area of heathland within the SPA are very unlikely to be correct, as woodlarks nest in a wide variety of habitats (including fields, paddocks and clearfell areas), and much heathland is unsuitable for woodlarks because of their very specific habitat requirements (Mallord ***et al.*** 2006).

4.57. ***Analyses of Dartford warbler distribution.***

The statistical problems outlined above also apply to the analysis of Dartford warbler distribution. In addition, Dartford warblers in the UK are currently expanding in number and distribution, probably as a result of increased winter survival following a series of mild winters. Analyses of numbers on any particular site must therefore also take account of the overall increase in the UK population size and the probability of a site being occupied, which is influenced both by habitat quality and distance from other occupied sites.

4.61, 4.62 & 4.63. ***Inappropriate use of univariate statistics.***

Applying a series of univariate tests (e.g. the t-test or its non-parametric equivalent, the Mann-Whitney test) to identify environmental predictors of species presence or absence is a highly simplistic approach that takes no account of any covariance between the environmental variables. For example, if squares with a high proportion of heath tend also to have a low proportion of woodland, univariate tests may find both variables to be significant predictors of nightjar presence or absence when in fact only one may actually influence nightjars. The habitat variables recorded here are almost bound to covary since, if, for example, woodland and heath are the two dominant habitat types, when a square comprises a high percentage of heath, it can only comprise a small percentage of woodland, and ***vice versa***. Thus the apparent avoidance of mature woodland and preference for immature woodland by nightjars may simply be because mature woodland is rare when immature woodland is common. This type of analysis can therefore not be used to accurately assess species' habitat preferences.

When datasets contain multiple predictors (e.g. habitat variables), multivariate tests are the appropriate methods of analysis because they allow the statistical significance of each variable to be calculated while controlling for the effects of the other variables. In this case, a binomial logistic regression analysis is the appropriate analysis, as it allows the relative effect of multiple predictor variables (e.g. extent of urban development, area of woodland etc) on the probability of a species occurring in a site to be calculated. (Note: this dataset may not actually be sufficient for a binomial logistic regression, since it would also be necessary to control for similarities amongst squares within each site, by including site within the analyses.)

4.64. The apparently greater distance from the SPA boundary of occupied squares than randomly selected squares could arise because of differences in habitat structure across the SPA (ie preferred habitats may be further from the boundary). Distance from the SPA boundary should therefore have been incorporated into the logistic regression analysis that should have been used for 4.61 - 4.63.

4.65. The analysis of differences in length of path within squares occupied or not by each species is similarly confounded by not including the habitat and distance effects explored above (4.61 - 4.64). Thus, the probability of a square being occupied or not may be influenced by habitat type, distance or number of paths, and the effects of each of these variables can only be explored in analyses which include all of them. Using inappropriate univariate analyses such as these means that it is impossible to know whether, for example, paths lengths are greater in unsuitable habitat and therefore path length appears not to affect bird distribution. The problems with this highly simplistic approach to assessing the effects of human disturbance on wildlife have been described previously (Gill ***et al.*** 1996, 2001).

In summary, the statistical procedures used to address the key ecological questions in this study are flawed to such an extent that it is highly unlikely that the true relationships between bird distribution and the key environmental factors have been correctly identified. The questions of whether the Liley & Clarke approach is appropriate to the TBH, and whether the distribution of nightjars, woodlarks or Dartford warblers in the TBH is influenced by urban development therefore remain to be answered.

References

Gill, J.A., Sutherland, W.J. & Watkinson, A.R. (1996) A method to quantify the effects of human disturbance for animal populations. ***Journal of Applied Ecology***, **33**, 786-792.

Gill, J.A., Norris, K. & Sutherland, W.J. (2001) Why behavioural responses may not reflect the population consequences of human disturbance. ***Biological Conservation***, **97**, 265-268.

Liley, D. & Clarke, R.T. (2003) The impact of urban development and human disturbance on the numbers of nightjar, ***Caprimulgus europaeus***, on heathlands in Dorset, England. ***Biological Conservation***, **114**, 219-230.

Mallord, J.W., Dolman, P.M., Brown, A.F. & Sutherland, W.J. (2006) Linking recreational disturbance to population size in a ground-nesting passerine. ***Journal of Applied Ecology***, doi: 10.1111/j.1365-2664.2006.01242.x

Tabachnick, B.G. & Fidell, L.S. (1996) ***Using multivariate statistics***. Harper Collins Inc.