The use of GIS in estimating spatial variation in habitat quality: a case study of lay-date in the Great Tit *Parus major*

TEDDY A. WILKIN,* CHRISTOPHER M. PERRINS & BEN C. SHELDON

Edward Grey Institute of Field Ornithology, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

Finding the most biologically meaningful scale at which to describe environmental variation is a persistent problem in ecology. Most studies of forest passerines are conducted at the scale of the habitat or woodland and do not account for environmental variation between individual breeding sites. Here we employ two GIS models, and four spatial scales, to describe environmental variation among 4683 Great Tit *Parus major* breeding sites occupied over a 32-year period in Wytham Wood, Oxford, UK, and use these data to help explain variation in an environmentally sensitive trait, first egg date. Model 1 used Thiessen polygons to generate individual spatial scales for each breeding pair, while Model 2 used a range of predetermined radial scales around each breeding site. Environmental variables included local altitude and aspect estimated from a Digital Terrain Model, and the number of Oak *Quercus robur* trees around each nest-site, used here as a surrogate for local food (caterpillar) availability. In a Linear Mixed Model, Model 2 explained the most variation in lay-date at a scale of 25 m from each nestbox. However, the model that returned the lowest Akaike Information Criterion (AIC) contained environmental variables from Model 2, but measured at different fixed spatial scales of between 25 and 75 m. Results from this final model showed that birds breeding in low-altitude, north-facing and Oak-rich areas bred earlier than those in high-altitude, south-facing and Oak-poor areas, at radial scales of 25, 75 and 75 m from each nestbox, respectively. In addition, birds in interior sites bred earlier than those nearer the woodland edge, although this edge effect was only apparent on north-facing slopes. Thus, the current study demonstrates that a range of GIS models can be effectively used to scale and describe environmental variation between individual breeding sites, and that fine-scale topographic variation, food availability and edge proximity can affect the breeding date of Great Tits.

Keywords: aspect, digital terrain model, tessellation, Thiessen, Voronoi.

Finding the most biologically meaningful scale at which to describe environmental variation is a persistent problem in ecology. Most studies of forest passerines are conducted at the scale of the habitat or woodland and do not account for environmental variation between individual breeding sites. Here we employ two GIS models, and four spatial scales, to describe environmental variation among 4683 Great Tit *Parus major* breeding sites occupied over a 32-year period in Wytham Wood, Oxford, UK, and use these data to help explain variation in an environmentally sensitive trait, first egg date. Model 1 used Thiessen polygons to generate individual spatial scales for each breeding pair, while Model 2 used a range of predetermined radial scales around each breeding site. Environmental variables included local altitude and aspect estimated from a Digital Terrain Model, and the number of Oak *Quercus robur* trees around each nest-site, used here as a surrogate for local food (caterpillar) availability. In a Linear Mixed Model, Model 2 explained the most variation in lay-date at a scale of 25 m from each nestbox. However, the model that returned the lowest Akaike Information Criterion (AIC) contained environmental variables from Model 2, but measured at different fixed spatial scales of between 25 and 75 m. Results from this final model showed that birds breeding in low-altitude, north-facing and Oak-rich areas bred earlier than those in high-altitude, south-facing and Oak-poor areas, at radial scales of 25, 75 and 75 m from each nestbox, respectively. In addition, birds in interior sites bred earlier than those nearer the woodland edge, although this edge effect was only apparent on north-facing slopes. Thus, the current study demonstrates that a range of GIS models can be effectively used to scale and describe environmental variation between individual breeding sites, and that fine-scale topographic variation, food availability and edge proximity can affect the breeding date of Great Tits.

Keywords: aspect, digital terrain model, tessellation, Thiessen, Voronoi.

Finding the most biologically meaningful scale at which to describe environmental variation is a persistent problem in ecology. Most studies of forest passerines are conducted at the scale of the habitat or woodland fragment. For example, several studies have contrasted the mean breeding date of birds breeding in deciduous, compared with coniferous, woodlands (Lambrechts *et al*. 2004, Tremblay 2005). However, there is likely to be considerable environmental variation within these broad habitat classifications and the characteristics which define a habitat type are unlikely to be randomly distributed within any such scale. Therefore, within-scale environmental variation is unaccounted for in such analyses.

Environmental variation can take many forms. Traditional sources of variation relate to factors...
associated with habitat quality and are usually, but not exclusively, formed from interactions between the availability of food and the distribution of competitors. At the level of the population, these factors are linked to broad habitat classifications and population size, while at the level of the individual they are determined by territory size and quality, and by the distribution of other competing species. For Great Tits *Parus major*, food availability, and hence an important aspect of territory quality, is strongly associated with the distribution of Oak *Quercus robur* trees, upon which the main food supply for their nestlings, caterpillars, are found at highest densities (Summerville et al. 2003). Other sources of variation relate to the optimum conditions required by the focal species and its food supply, and include factors such as temperature and exposure. At the level of the population these are linked to local meteorology, while at the level of the individual territory they are determined by fine-scale variation in topography.

We argue that the most appropriate scale to adopt when investigating the reproduction of forest passerines is the level of the individual breeding site, as this is the scale at which competition and selection are expected to operate. Yet assessing territory size and habitat use in the field is prohibitively time-consuming and it is impossible to apply such methods to the analysis of long-term, historical data sets. However, recent advancements in Geographic Information Systems (GIS) have allowed biologists to make quantitative predictions of territory size and shape based solely on the breeding location of focal individuals and their neighbouring conspecifics (e.g. Wilkin et al. 2006). We propose that the Thiessen polygons formed by such GIS models may also offer individual and biologically meaningful scales at which to describe environmental variation between breeding sites. However, these models are at present unable to account for interstices between territories and consequently they are likely to overestimate territory sizes in areas of low breeding density. An alternative method is to examine environmental variation within search buffers formed with fixed and predetermined radii around each breeding location. A clear advantage of this approach is the ability to examine environmental variation over a range of spatial scales by altering the radius of the search buffers. However, this technique is biologically less appealing as it results in the non-mutual exclusivity of resources in areas where the distance between breeding sites is less than double the search radius’ although it is by no means certain that Great Tits maintain entirely exclusive territories on this small scale. Furthermore, a given radius may be appropriate in some areas but not others, and for some resources but not others, depending on the quality of the environment, the number of individuals present, and the way that resources are exploited.

The aims of the current paper were two-fold. First, we compared the utility of two GIS models by forming both Thiessen polygons and fixed-radius search buffers at three spatial scales, around occupied breeding sites in each year. The resulting polygons were then used to describe variation in topography, food availability, edge proximity and interspecific competition, between 4683 Great Tit breeding sites occupied over a 32-year period. Secondly, we compared the utility of the GIS models in explaining variation in first egg dates between individual females, and use the results to generate hypotheses relating to causation.

**METHODS**

**Field data**

Data were obtained during the Edward Grey Institute’s long-term study of the Great Tit at Wytham Woods, near Oxford, UK. The data used in the present study were collected between 1965 and 1996, in accordance with methods described previously (e.g. Perrins 1965, 1979, Gosler 1993). Great Tits breed almost exclusively in nestboxes in Wytham, the number and locations of which have remained more or less constant throughout the current study period, unless a short-distance move was necessitated by tree fall. The locations of each nestbox (*n* = 1020) and each Oak tree (*n* = 3926) were digitally mapped in the field during the winter of 2004–05 by using a differential GPS system (submetre accuracy) with a laser rangefinder to map distant objects (Wilkin et al. 2007). It was not necessary to impose a minimum size for an Oak tree to be mapped as no saplings were present at the study site.

Nestboxes were visited at least weekly to determine clutch initiation date (lay-date, where 1st April = 1). Parents were trapped at nestboxes after the chicks reached 7 days of age (to reduce desertion risk), aged and sexed (Svensson 1992), and ringed or their identities established from rings already fitted.

We analysed first clutches (second clutches are very rare in Wytham) defined as those laid within 30 days of the first clutch in each year (as in van Noordwijk et al. 1995), and only those in which four or more eggs were laid. Habitats were broadly
classified into the following five categories as in Gibson (1988) and Gosler (1990): (1) ancient semi-natural woodland, (2) 18th and 19th century plantations, (3) 20th century plantations, (4) secondary regenerated wood pasture and (5) grassland (containing 545, 232, 146, 97 and 0 nestboxes, respectively). The spring temperature each year was recorded as the sum of the maximum temperatures for each day between 1 March and 25 April (the warmth sum), as in McCleery and Perrins (1998).

Map Info Professional v7.8 and Vertical Mapper v3 were used to produce annual maps detailing the location of nestboxes in Wytham Woods that were occupied by Great Tits during the breeding seasons of 1965–1996.

Model 1: Thiessen polygons

The GIS software was used to generate quantitative predictions about habitat use based on the relative location of occupied nestboxes within the perimeter of the wood (as in Wilkin et al. 2006). By using a Dirichlet tessellation technique, we formed Thiessen polygons (Rhynsburger 1973, Tanemura & Hasegawa 1980, Stoyan et al. 1987) around occupied nestboxes (Fig. 1a). A tessellated polygon is a geometric construct that contains all points closer to the generating point, in this case a nestbox, than to any other. This is achieved by placing boundaries exactly mid-way between all adjacent neighbours up to and including the woodland perimeter. Thiessen polygons are contiguous and cover all of the available area, so at low nest densities, some polygons were excessively oversized, some exceeding 10 ha in size. As a result the scales at which Thiessen polygons describe environmental variation are more biologically meaningful for high-density, rather than low-density, individuals and the ability of the model to explain environmental variation for low-density individuals will depend on the spatial heterogeneity of the environmental variables. The areas of Thiessen polygons, capped at a maximum of 1 ha, were used in the current analyses as a measure of breeding density, as this capping imitates interstices at a tested level in the current population (as in Wilkin et al. 2006).

Model 2: Radial buffers

Using the GIS software, three sets of search buffers were formed around each breeding location. The buffers measured 25, 50 (Fig. 1b) and 75 m in radius from each nestbox corresponding to areas of 0.20, 0.79 and 1.77 ha, respectively, and were used to examine environmental variation at these spatial scales.

Environmental variables

A three-dimensional model of Wytham was formed from an Inverse Distance Weighting (IDW) interpolation of a 10-m resolution Land Form PROFILE Digital-Terrain-Model (DTM) data set provided by Ordnance Survey. By running the GIS models and the DTM simultaneously we were able, by means of region inspection queries, to extract the mean within-polygon altitude, slope aspect (relative to due north)
and gradient strength (the last is calculated for each point by averaging the difference in altitude between the surrounding eight points on a grid square). Aspect angles are circular in form and so were transformed into a linear measurement of northness, equal to the cosine of the aspect angle. The resulting measure of northness relates to the degree to which the slope faces north on a scale of 1, for due north, to –1, for due south, while points facing east or west have a zero value for northness. The GIS software was also used to measure the shortest distance between the midpoint of each polygon (the nestbox in the case of the radius model) and the woodland perimeter as a measure of edge proximity, the most common (median) habitat category, the number of Oak trees and the number of nestboxes occupied by a close relative, the Blue Tit Cyanistes caeruleus, within each type of polygon.

**Statistical analyses**

Four Linear Mixed Models (LMMs) with normal errors were constructed with Genstat v8 (VSN Intl 2005) to assess the effects of environmental variation on lay-date. The nestbox, year of reproduction and female identity were included as random effects. By including nestbox as a random effect we control for non-independence between breeding attempts in the same nestboxes in different years, by including year for non-independence between breeding attempts in the same nestboxes in different years, by including year

Asymptotically as \( \chi^2 \) (VSN Intl 2005). All first-order interactions were considered except those between variables correlated with each other due to non-random variation within the environment. For example, as Wytham is a hill-top woodland, edge distance is necessarily correlated with both altitude and slope strength, i.e. low-altitude territories are usually both near the edge and on level ground. Early lay-dates have been consistently associated with older females, resident females and warmer springs (Wilkin et al. 2006 and references therein). Therefore, the age of the female (exact age for local recruits and minimum age for immigrants) and the status of the female (resident or immigrant, defined as whether hatched in a Wytham nestbox or not, respectively) were included in the model as continuous and categorical variables, respectively. Spring temperature (warmth sum) and other environmental variables were treated as continuous variables, with the exception of habitat type.

First, four separate LMMs were run, each containing all the environmental variables for a spatial scale (Thiessen polygons, 25-, 50- and 70-m buffers), together with spring temperature and female age/status. The amount of variation in lay-date explained by the fixed effects was calculated by the percentage of the difference between the residual variance when only random effects were included and the residual variance of each model. The final model was constructed by entering all the environmental variables at each of the four spatial scales into a single all-subsets regression. Spring temperature and a female's age and status were forced into the model. The best model was selected as the combination of variables that returned the lowest Akaike Information Criterion (AIC) value. For ease of presentation, data are displayed graphically as residuals estimated by removing the predictor variable from the model and re-running the analysis.

**RESULTS**

The GIS models were applied to 7856 occupied nestboxes over the 32 years of the study. Of these, 4683 reproductive attempts (mean = 165 per year) met the conditions detailed in the previous section and were thus suitable for analysis.

**Model comparisons**

Four LMMs were run, each containing every environmental variable measured at the given spatial scale. In each of these four models all of the environmental variables were significantly associated with Great Tit lay-date (all \( P < 0.009 \)), with the exception of gradient slope and the number of Blue Tits in the vicinity of each nestbox. The fixed effects in the Thiessen polygon model explained 5.87% of lay-date variation, while the fixed effects in the 25-, 50- and 75-m-radius
models explained 7.04, 6.43 and 5.92% of the variation in lay-date, respectively (Fig. 2). This suggests that environmental variation is most important at a fixed scale of 25 m from each nestbox. However, it is not unreasonable to assume that different environmental factors will be important at different scales. Therefore, the scale at which each environmental factor was most important in determining Great Tit lay-date was determined by the combination of factors and spatial scales that returned the lowest Akaike value in an all-subsets regression.

**Final model**

Results from the final model are shown in Table 1. As has been shown previously for the current population (e.g. Perrins 1979, McCleery & Perrins 1998, Wilkin et al. 2006), lay-date was negatively associated with female age ($\chi^2 = 151.6, P < 0.001$), spring temperature measured as warmth sum ($\chi^2 = 57.33, P < 0.001$) and whether birds were immigrants or not ($\chi^2 = 16.74, P < 0.001$). Birds fledged from Wytham nestboxes bred, on average, 1.6 days earlier than immigrants, an effect that persisted throughout their lifetimes, as shown by the absence of an interaction between female age and status ($\chi^2 = 0.00, P = 0.974$). In addition, there was a weak negative relationship between lay-date and the size of the Thiessen polygon capped at 1 ha ($\chi^2 = 6.37, P = 0.012$; Fig. 3a), suggesting that lay-date was only weakly related to local population density (as in Wilkin et al. 2006). Each of these fixed effects was forced into the final model (Table 1).

The final analysis showed strong relationships between lay-date and both mean altitude within 25 m radius ($\chi^2 = 46.34, P < 0.001$; Fig. 3b) and Oak tree abundance within 75 m radius ($\chi^2 = 29.13, P < 0.001$; Fig. 3c), showing that earlier lay-dates were achieved at low altitude and in territories with high Oak abundance. We also found a strong effect of mean northness within 75 m radius ($\chi^2 = 43.93$,
P < 0.001; Fig. 3d) and edge distance ($\chi^2_1 = 25.66, P < 0.001$; Fig. 3e) on lay-date, such that birds bred earlier on north-facing slopes and in interior areas. An interaction between these two variables was also retained ($\chi^2_1 = 3.9, P = 0.048$; Fig. 3f), suggesting that the edge effect was only apparent in north-facing ($\chi^2_1 = 25.05, P < 0.001$), as opposed to south-facing ($\chi^2_1 = 1.75, P = 0.187$), territories (Fig. 3f). Lastly, there was a significant effect of median habitat type within 25 m radius on lay-date ($\chi^2_1 = 61.23, P < 0.001$). The general lay-date trend with respect to median habitat types was that birds breeding in areas dominated by later successional habitats such as ancient semi-natural woodland and secondary regeneration bred earlier than birds breeding in territories dominated by less established habitats such as 20th century plantations and grasslands. No significant interactions were found between environmental variation and the other, forced variables in the model, suggesting that the above environmental effects were equally important for birds differing in age and immigration status, and also between springs of differing temperatures. Note that all significant effects were present simultaneously and independently of each other, as each of the environmental factors discussed were included in the same analysis.

DISCUSSION

Breeding date was most strongly associated with environmental variation within 25 and 75 m of occupied focal nestboxes, depending on the environmental factor being measured (Table 1). This demonstrates that a range of fixed-radius buffers formed around breeding locations can effectively be used to select the scale that is most biologically meaningful for a given environmental factor. Our analyses have shown that, in the current population, Great Tit breeding date responded significantly to topographical variation such as altitude and aspect, and also varied with
habitats type, Oak tree abundance and woodland edge proximity. These effects occurred independently of one another and, as female identity was included as a random effect, the effects were apparent within repeat breeding individual females (Nussey et al. 2005 and references therein). This means that for a given individual, any change in an environmental factor between successive breeding attempts was mirrored by a change in lay-date. This strongly suggests that these effects were direct environmental effects upon individuals and did not occur because young, poor quality or newly immigrated individuals had been displaced to poor quality environments, where they bred later in the season.

Thiessen polygons explained less lay-date variation than the fixed-radius models (Fig. 2). This is somewhat surprising given that the former generates mutually exclusive and individual scales at which to describe environmental variation. One possible reason for the reduced explanatory power is that as Thiessen polygons are entirely contiguous, either with each other or with the perimeter of the woodland, extremely large and biologically meaningless polygons are formed in areas where the breeding density is low enough for areas of unutilized habitat, or interstices, to be present between territories. For this reason, at very low densities, which are found in some areas of this study population, the efficacy of the territory model in measuring environmental variation at a local scale is reduced. The Thiessen polygons do explain more variation in lay-dates when the low-density areas are excluded from the analysis but this data restriction results in a considerable reduction in sample size and biased sample. We therefore propose that the Thiessen polygons method may still represent the most suitable model for assigning variation to individuals in the environment, but only in populations where all breeding locations are known and where individuals breed at sufficiently high density that all the available habitat is utilized.

Laying date was earlier with increasing female age and spring temperature, and birds born in Wytham nestboxes bred earlier than immigrants (Table 1). These results have been reported before in the current population (e.g. Perrins 1979; McCleery & Perrins 1998, Wilkin et al. 2006). The advancement of laying date with age may reflect, partially, improved competence, or learning, and the matching of the birds’ breeding time to annual variation in temperature is an adaptation to unpredictable timing of development of the birds’ main nestling food source in different years. In contrast, the difference in lay-date between immigrants to the woods and residents is less easy to explain. This effect could be due to some maternal effect of being born outside of the main woodland which is perhaps poorer habitat, or in a natural cavity rather than a nestbox. What is particularly intriguing is that this effect is persistent throughout the lifetime of an individual, such that an immigrant, who has bred in Wytham for five consecutive years, is still expected to breed later than a similarly aged resident. This effect warrants further investigation.

We found only a weak relationship between lay-date and individual Thiessen polygon size (capped at 1 ha), suggesting that lay-date is only weakly density-dependent at the level of the individual (Fig. 3a). This is somewhat supported at the population level by the lack of a clear relationship in the current data set between mean lay-date and the number of pairs breeding in Wytham each year (General Linear Model, $t_{28} = -1.41, P = 0.169$), when controlling for differences in spring temperature. However, recent and current work on this population have shown that individuals breeding at very high density, i.e. those with the smallest tessellated territories, display a significant negative relationship between territory size and breeding date (Wilkin et al. 2006; Fig. 3a). This effect is masked at the level of the population because, in a given year, up to 60% of the population are at low enough density, i.e. territories larger than 1 ha, to be unconstrained with respect to their breeding dates.

**Environmental effects**

Lay-date was delayed with increasing altitude at a scale of 25 m from each nestbox (Fig. 3b). This altitude affect has been seen before in the current population (Wilkin et al. 2006) and can be explained by a phenological phenomenon predicted by Hopkins law of bioclimatics, i.e. spring is often delayed by 1 day for every 30 m rise in altitude, probably due to differences in temperature, which affect developmental budgets (Hopkins 1938). Wytham has around 100 m range in altitude and so we might expect a 3-day difference in breeding date between birds at the top of the hill and those at the bottom. In the current dataset, a linear regression of altitude on lay-date does indeed show almost a 3-day delay between high-altitude birds (fitted lay-date = 29.9) and low-altitude birds (fitted lay-date = 27.1) ($t_{489} = 5.67, P < 0.001$). However, the effect size in the final model only shows about 1 day’s delay when controlling for potentially confounding factors such as spring
temperature and aspect (Fig. 3b). Therefore, of the 
3-day difference observed in the raw data between 
the high- and low-altitude birds, 1 day’s variation is 
directly attributed to altitude in our model, probably 
as a surrogate for temperature, while the remaining 
2-day variation is probably attributable to other 
topographic factors also related to temperature 
differences within the landscape.

Birds breeding in Oak-rich areas bred earlier than 
those in Oak-poor areas, at a fixed scale of 75 m from 
each nestbox (Fig. 3c). Our a priori reason for mapping 
and counting the number of Oak trees within the 
locality to each nestbox was that we expected the 
number of Oak trees to act as a surrogate measure 
for caterpillar availability during the provisioning 
period. However, as the majority of caterpillars have 
not emerged at the time when the birds are preparing 
to lay eggs, the caterpillars cannot be responsible for 
earlier laying in Oak-rich areas. However, Oak trees 
are also known to harbour very high densities of 
other invertebrates, which may enable birds in 
Oak-rich territories to reach breeding condition 
early. For example, about one-third of the Great Tit 
Oak-rich territories to reach breeding condition 
early. For example, about one-third of the Great Tit 
lay-dates on southern slopes. This intriguing possibility 
is currently under investigation.

We found a strong advancement of lay-date with 
increasing northness around each nestbox at a 
fixed scale of 75 m (Fig. 3d). This was somewhat 
unexpected as south-facing environments, which are 
presumably warmer, would normally be associated 
with less costly reproduction and faster development. 
However, as north-facing slopes receive less solar 
radiation than southern slopes they may be damper 
and more humid, possibly increasing invertebrate 
abundance, which would enable birds to reach 
breeding condition earlier than birds on south-facing 
slopes where invertebrates would be more likely to 
face desiccation pressure. An alternative hypothesis 
is differential likelihood of replacement clutches 
between north- and south-facing slopes. For example, 
if females are more likely to lay replacement clutches 
following earlier failed attempts on southern than on 
northern slopes, then this would result in later mean 
lay-dates on southern slopes. This intriguing possibility 
is currently under investigation.

Birds breeding nearer the woodland edge laid later 
than birds in more interior territories (Fig. 3e). Edge 
effects on the timing of breeding have been shown 
before in the current population (Wilkin et al. 2007), 
for Great Tits breeding in small as opposed to 
large woodlands (Hinsley et al. 1999) and in linear 
habitats such as hedgerows (Riddington & Gosler 
1995). A possible reason for this edge effect is that 
seed fall and/or agricultural waste from outside the 
woodland has degraded the woodland around the 
edge, making it more difficult for ‘edge’ birds to get 
into breeding condition as early as ‘interior’ birds. 
However, as Great Tits feed on a wide range of food 
items that are found primarily on mature trees it is 
unlikely that a change in understorey vegetation 
would render the habitat suboptimal. Another 
possible explanation is that increased wind exposure 
at the woodland edge may reduce the density of 
invertebrates via desiccation pressure or temperature 
changes. This hypothesis is supported by our novel 
finding that edge effects were only apparent on north-

The results of this study suggest that a range of 
fixed-radius GIS models offer the most effective 
method for describing environmental variation on 
biologically meaningful scales. In this population, we 
have shown that topographical variation, such as 
alitude and aspect, the abundance of Oak trees, 
habitat type and edge proximity can affect the lay-
date of breeding Great Tits.
We are grateful to the many people who collected data during the long-term study of the Great Tit at Wytham and to two anonymous referees for their helpful comments on the manuscript. T.A.W. was funded by a PhD studentship from Biotechnology and Biological Sciences Research Council (BBSRC).

REFERENCES


