Journal of Applied Ecology 2000, **37** (Suppl. 1), 13–27

Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent

YVONNE C. COLLINGHAM*, RICHARD A. WADSWORTH†, BRIAN HUNTLEY* and PHILIP E. HULME*

**Department of Biological Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, UK; and †Centre for Ecology and Hydrology, Monks Wood, Abbots Ripton, Huntingdon PE17 2LS, UK*

Summary

1. The existence of a hierarchical scheme of environmental controls on the spatial distribution of plant species was explored for three non-indigenous weeds, *Fallopia japonica*, *Heracleum mantegazzianum and Impatiens glandulifera*, in the British Isles.

2. Logistic regression analyses of the presence/absence of the three weed species examined the relative importance of 60 environmental variables, encompassing land cover, geology and climate. Analyses were undertaken using variables assessed at a hectad (10×10 km) or tetrad (2×2 km) resolution at national (England and Wales) and regional (County Durham, UK) spatial extents.

3. The ranges of all three species in the British Isles are currently increasing, and the non-equilibrium nature of their distribution limited the goodness-of-fit of logistic models. Interpretation of whether a species has expanded to occupy entirely its potential spatial range was scale-dependent, and species' distributions, when viewed at coarser spatial scales, may be more likely to be interpreted as having reached stasis.

4. Spatial autocorrelation was more evident at the finer tetrad spatial resolution for both *F. japonica* and *I. glandulifera*, but not evident at all for *H. mantegazzianum*. Only the distribution of *I. glandulifera* revealed significant spatial autocorrelation among hectads at the national scale. These patterns appear related to the different dispersal mechanisms of the three species.

5. The majority of the environmental variables identified as important at the tetrad resolution for County Durham were also important at the hectad resolution for England and Wales for both *F. japonica* and *I. glandulifera*, but not for *H. mantegazzianum*. However, for all three species the environmental variables identified as significant were consistent with qualitative descriptions of the species' habitat characteristics. There was no evidence of a hierarchy of environmental controls.

6. At the regional extent, scaling-up species' distributions from tetrads to hectads was relatively successful, but scaling-down was not. The coarser resolution models were too unrefined to model fine-scale distributions successfully. Similarly, at a coarse hectad resolution, regional models were poor predictors of national species' distributions. It therefore appears that scaling-up from fine to coarse resolution is appropriate when spatial extent is held constant, and focusing-down from large to small spatial extents is appropriate when data resolution is held constant.

Key-words: Fallopia japonica, *Heracleum mantegazzianum*, *Impatiens glandulifera*, landscape ecology, logistic regression, spatial autocorrelation.

Journal of Applied Ecology (2000) **37** (Suppl. 1), 13–27

Introduction

Plant invasions are a serious threat to natural and managed ecosystems world-wide (Hobbs & Humphries 1995; Manchester & Bullock 2000). There are

© 2000 British Ecological Society

Correspondence: Philip E. Hulme (fax 91 374 2417; e-mail P.E.Hulme@durham.ac.uk).

well-documented examples of invading species that degrade human health and wealth, alter the structure and functioning of otherwise undisturbed ecosystems, and/or threaten native biological diversity (Vitousek *et al.* 1997). In the British Isles over 40% of the vascular flora comprises established non-indigenous plant species (Clement & Foster 1994). To date, research on the non-indigenous flora of the British Isles has focused on assessments of rates of spread (Perrins, Fitter & Williamson 1993) or attempts to identify life-history traits that correlate with species invasiveness (Thompson, Hodgson & Rich 1995; Crawley, Harvey & Purvis 1996). However, it has recently been suggested that a focus on the characteristics of the invaded ecosystem, rather than the invader, is likely to be a more effective research strategy (Hobbs & Humphries 1995). Although a broad survey of the invasibility of British ecosystems has been undertaken (Crawley 1986), no attempts have been made to identify the characteristics of habitats prone to invasion by specific target species. This could be achieved by assessing current distributions with reference to suitable environmental variables, e.g. vegetation type. Such an assessment would facilitate the setting of management priorities for habitat protection and invasion control (Hobbs & Humphries 1995).

Plant species distribution data are available at a variety of spatial scales, including regional (Graham 1988), national (Perring & Walters 1976) and continental (Jalas & Suominen 1972–94). When combined with climate, soil and/or vegetation maps, these databases have presented numerous opportunities to study the factors underpinning plant species' distributions (reviewed by Franklin 1995). For terrestrial plants, a hierarchical scheme of environmental controls on species' distributions has been suggested in which climatic variables correspond to patterns at the largest spatial scales, followed by geology (and its effect on soil chemistry and nutrient availability) and topography, which moderates many of the macroclimatic variables (Franklin 1995). An understanding of the interplay among processes operating at different spatial, and necessarily temporal, scales will identify their relative importance and emphasize the most appropriate scale at which to predict future changes to the system (May 1989). However, to date few studies have specifically examined the correspondence between the factors important in species' distributions at one spatial scale and those operating at other spatial scales.

It is within this context that this study proposed to explore scale dependence in the environmental factors that influence the spatial distribution in the British Isles of three species of non-indigenous invasive weeds: *Fallopia japonica* Houtt. (Japanese knotweed), *Heracleum mantegazzianum* Sommier et Levier (giant hogweed) and *Impatiens glandulifera* Royle (Himalayan balsam). All three species were deliberately introduced to the British Isles as garden plants during the nineteenth century. However, following accidental introductions into natural habitats they have become widespread (Fig. 1), particularly along rivers and waterways (Ellis 1993). *Fallopia japonica* is a polycarpic perennial, introduced in 1825, that rarely sets seed in the British Isles and spreads locally through vegetative expansion and more widely through the dispersal of rhizome fragments in water or following anthropogenic transport of soil (Beerling, Bailey & Conolly 1994). *Heracleum mantegazzianum* is a monocarpic perennial, introduced in 1893, that in its second or third year sets copious amounts of seed that are dispersed by wind and/or water (Tiley, Dodd & Wade 1996). *Impatiens glandulifera* is an annual, first introduced in 1855, whose seeds are dispersed in the first place ballistically and then secondarily by water (Beerling & Perrins 1993). Their similar history in the British Isles, co-occurrence within riparian habitats and their markedly different life histories makes these species particularly suitable for comparative studies of their distribution. In addition, these species present problems of an economic (*F. japonica* damages riparian flood defence embankments), public health (physical contact with the sap of *H. mantegazzianum* has the potential to cause phytophotodermatitis) and ecological (all three species may outcompete and replace native herbaceous plants) nature (Roblin 1994). Attempts to eradicate these species locally pose a significant financial burden on statutory bodies and local wildlife groups (Roblin 1994; Sampson 1994). A better understanding of the factors determining their distribution will be a key step in the appropriate management of these invasive non-indigenous species. Furthermore, an understanding of the role played by climate and land use on the species' distributions will be pivotal in predicting their distributions in response to environmental change.

Progress towards the understanding, prediction and management of these invasive non-indigenous species must incorporate an awareness of the links among processes occurring at different spatial scales (Wadsworth *et al.* 2000). However, with regard to species distribution data, there is often an increase in data resolution with decreasing spatial extent. For example, the plant distribution data for Europe (10 382 000 km^2) are currently mapped on a 50×50-km resolution grid (Jalas & Suominen 1972–94), whilst for the British Isles (244 872 km^2) these data are available on a 10×10-km resolution grid (Perring & Walters 1976), and within the British Isles regional floras (e.g. County Durham, 263 km^2) may be mapped on a 2×2-km resolution grid (Graham 1988). Thus, comparisons of species–environment relationships at different spatial scales (regional,

© 2000 British Ecological Society

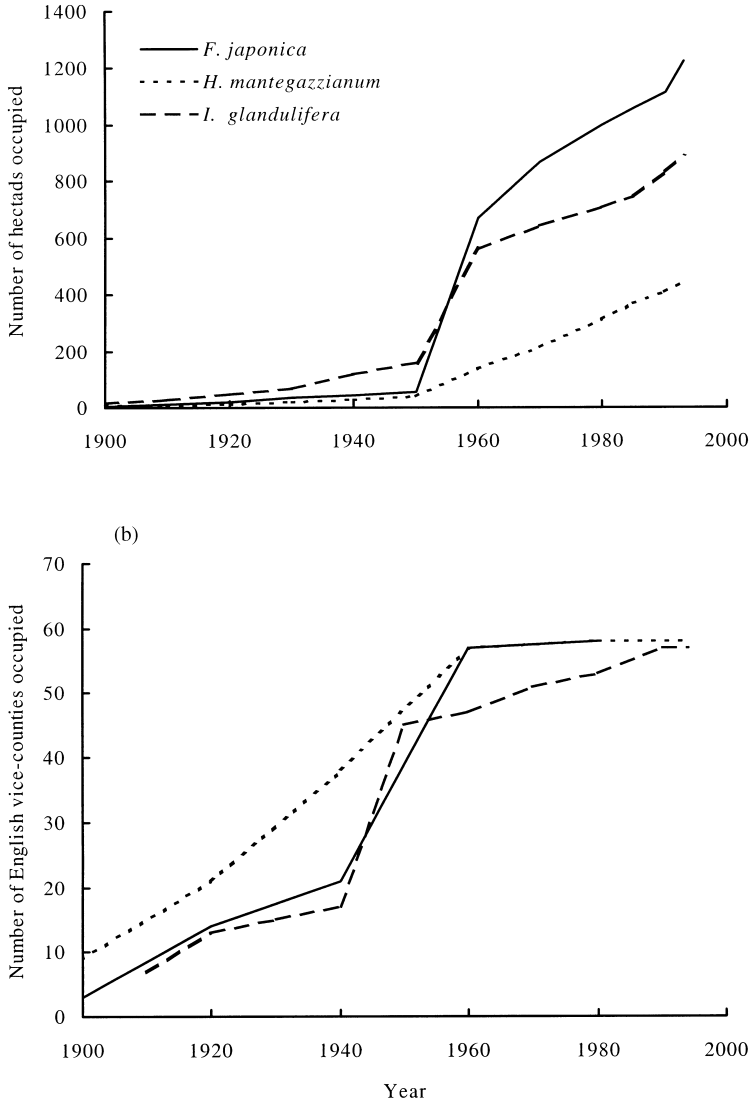


Fig. 1. The rate and extent of spread of *Fallopia japonica*, *Heracleum mantegazzianum* and *Impatiens glandulifera* as measured by (a) the cumulative number of hectads (10×10-km squares) in England and Wales; and (b) the cumulative number of English vice-counties, occupied since 1900. Data are from the Biological Records Centre.

national, continental) may confound the effects of resolution with that of spatial extent. For these two effects to be disaggregated, either extent or resolution must be held constant, e.g. rescaling the finer resolution to a coarser resolution to permit comparisons among data sets of different spatial extent. However, it still may be of applied interest to determine whether models developed from finer resolution regional data can be successfully applied to coarser resolution, national predictions or vice versa.

Unfortunately, to date only *F. japonica* is mapped at a European scale and thus detailed distribution data for all three non-indigenous species are currently only available for the British Isles. Using a common set of appropriately scaled predictor variables this study aimed:

1. For regional distributions, to develop separate statistical models of the presence/absence of the three non-indigenous species at both coarse (hectad) and fine (tetrad) spatial resolutions in order to identify key environmental variables that influence species' distributions at each spatial scale.

2. To evaluate the success of the statistical models developed at one spatial resolution to predict species' distributions at the other spatial resolution.

3. For national distributions, to develop separate statistical models of the presence/absence of the three non-indigenous species at a coarse spatial resolution in order to identify key environmental variables that influence species' distributions at the national scale.

4. To compare the national and regional models derived at a coarse spatial resolution to examine how spatial extent affects both the goodness-of-fit between species and environmental data and the relative importance of different environmental variables.

© 2000 British Ecological Society

5. To evaluate the success of regional models (both coarse and fine resolution) to predict species' distributions at the national scale and vice versa.

As part of the NERC/SERAD special topic on large-scale ecology, the work compliments contributions to this issue at other spatial scales (Baillie *et al*. 2000; Cowley *et al*. 2000; Gaston *et al*. 2000; Pettifor *et al*. 2000).

Methods and materials

SPECIES DISTRIBUTION AND ENVIRONMENTAL DATA SETS

National distribution data for the three species were obtained from the Biological Records Centre (Monks Wood, Abbots Ripton, Huntingdon), and included records for the years up to and including 1993 at a hectad resolution (10×10-km squares). Only national data for England, Wales and southern Scotland were used. Data from central and northern Scotland were omitted from the analysis because all three species are relatively scarce in these regions and probably only occupy a fraction of all suitable squares since species introductions occurred later than in England and Wales. Regional distribution data were taken from *The Flora and Vegetation of County Durham* (Graham 1988), which records species presence at a tetrad resolution (2×2-km squares) using records received up to October 1986. The County of Durham is situated in north-east England (Fig. 2) and encompasses a representative range of the topography (0–750 m a.s.l.), geology and land use of England and Wales (Graham 1988).

The proportion of different land cover types at both the hectad and tetrad resolution was obtained from the Institute of Terrestrial Ecology land cover data set (Fuller, Groom & Jones 1994). This data set was derived from satellite imagery using a classification process to derive a land cover class for each 25×25-m pixel of land. Twenty-six land cover classes were recognized, e.g. urban, woodland, etc. (see the Appendix). The 25×25-m pixel data set was then used to calculate the percentage cover for each land cover class found in each square at both the hectad and tetrad resolutions.

River networks were derived from national hydrology maps (Bartholomew's database; CHEST 1993) and converted into a 500-m grid using the Arc/Info Geographical Information System (GIS) (ESRI 1998). The 500-m cells that had a river flowing through them were classified as 'riparian' cells. The number of riparian cells in each square at the national (hectad) and regional (tetrad) resolutions was counted and then converted into a proportion by dividing by 400 and 16, respectively. This parameter gives an estimate of the degree to which a particular hectad or tetrad square is riparian. Roads (Bartholomew's database; CHEST 1993) were treated in a similar way, producing three variables reflecting the proportion of 500-m cells within a hectad or tetrad containing primary 'A' roads, secondary 'B' roads or minor roads.

Geological data were available only at a hectad resolution. However, these data were scaled to a tetrad resolution by scoring all tetrads falling within a particular hectad as having the geological characteristics of that hectad. The national land characteristic (NATLAC) data set (Ball, Radford & Williams 1983) was used, which was derived from mapping units of the 1 : 625 000 Institute of Geological Sciences Great Britain (North and South) maps of solid and drift geology. There were nine classes for bedrock stratigraphy, 13 for bedrock lithology and seven for surface geology (see the Appendix).

The minimum, mean and maximum elevations for each hectad were obtained from Ball, Radford & Williams (1983). For each tetrad, minimum, mean and maximum elevations were determined from the Institute of Hydrology's 50-m horizontal resolution digital terrain model using the Arc/Info GIS (ESRI 1998).

Laplacian thin-plate spline surfaces (Hutchinson 1989) fitted to meteorological station data for monthly mean temperature, monthly mean precipitation and monthly mean cloudiness, compiled initially by Leemans & Cramer (1991) and subsequently extended (W. Cramer, unpublished data), were used to interpolate values of these variables for locations at the centre of each hectad and tetrad. Following the methods described in Huntley *et al*. (1995) and Prentice *et al*. (1992), these interpolated values were then used to compute five bioclimatic variables: the number of degree days above 0 °C (GDD0), the number of degree days above 5 °C (GDD5), the mean temperature of the coldest month (MTCO), the mean temperature of the warmest month (MTWA) and the ratio of actual to potential evapotranspiration (APET) for the mean elevation in each square of the grids.

DATA ANALYSIS

Logistic regressions

Although a variety of statistical techniques has been used to model the spatial distribution of species, logistic regression analysis is a flexible and robust technique that has been applied frequently where species data are recorded as presence/absence (Buckland & Elston 1993; Franklin 1995). Prior to logistic regression analysis the Pearson correlations were calculated between all environmental variables at both hectad and tetrad resolutions in order to identify potential sources of multi-colinearity and singularity (Tabachnik & Fidell 1996). The majority of bioclimatic variables were highly correlated with elevation at both the hectad and tetrad resolutions (Pearson correlations $> 0{\cdot}70$). This was not surprising as elevation was one of the parameters that was

© 2000 British Ecological Society

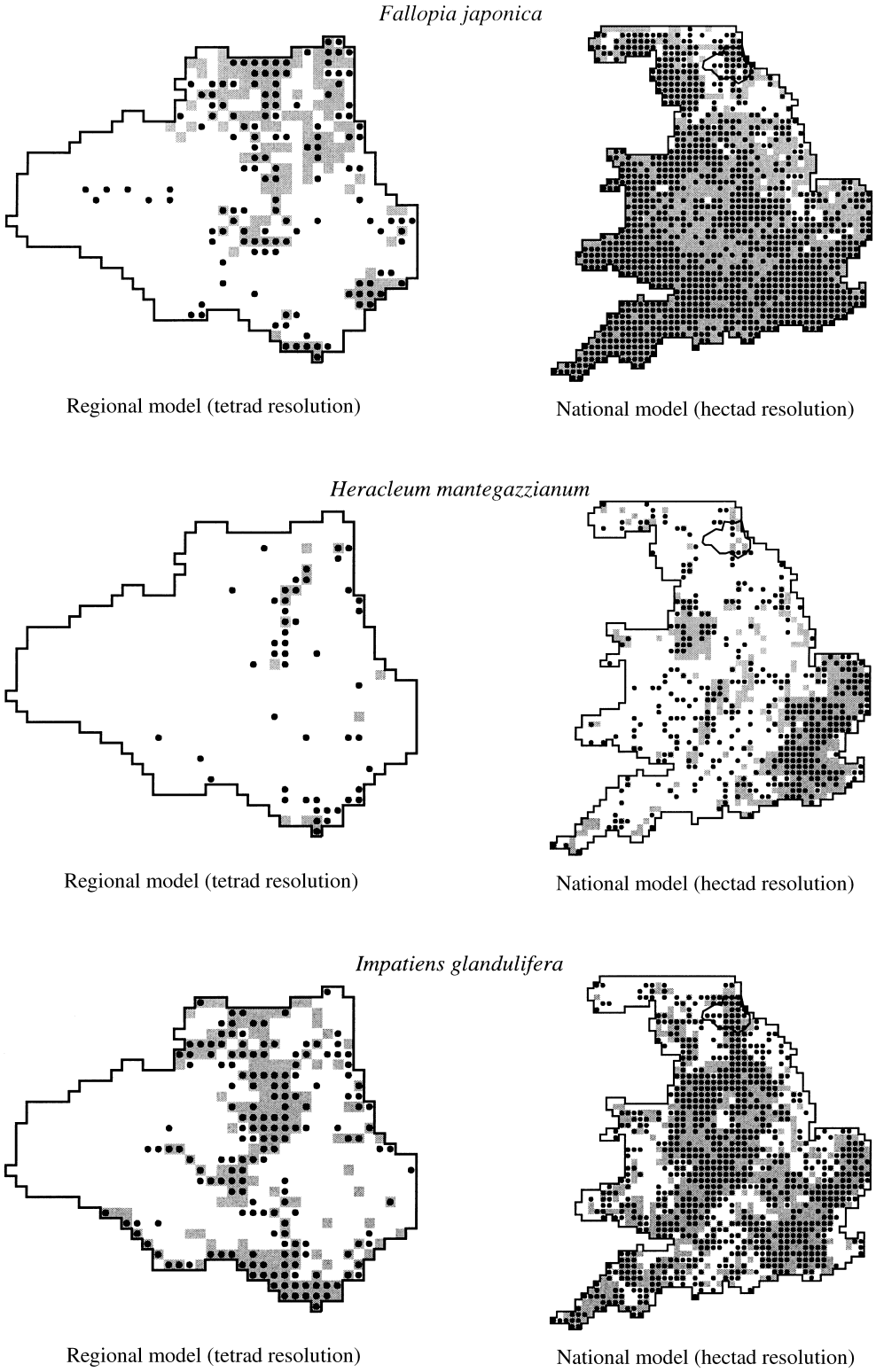


Fig. 2. Actual (points) and predicted (shading) presence of *Fallopia japonica* (top), *Heracleum mantegazzianum* (middle) and *Impatiens glandulifera* (bottom) at a tetrad resolution in County Durham (left) and at a hectad resolution for England and Wales (right). The County Durham boundary is outlined in the figures for England and Wales.

used in the spline surface interpolations, and it was therefore omitted from the variable list. Similarly large correlations were found between all bioclimatic variables apart from the ratio of actual to potential evapotranspiration. Thus, number of degree days above 0 °C, the number of degree days above 5 °C, the mean temperature of the coldest month and the mean temperature of the warmest month encompassed similar climatic variation. To limit the problems associated with this multi-colinearity, only

© 2000 British Ecological Society

one of these four variables, the number of degree days above 5 °C, was entered into the analyses. Thus the bioclimatic variables used in the analyses were the ratio of actual to potential evapotranspiration and the number of degree days above 5 °C.

At each spatial resolution and extent, separate stepwise logistic regression analyses were performed using SPSS (Norusis 1994), with the environmental variables as predictors of the presence/absence of each of the three non-indigenous species (decision rule: $\alpha = 0{\cdot}05$ for entry; $\alpha = 0{\cdot}1$ for removal). To allow testing of the regression models only half of each data set was used in the analyses (Fielding & Bell 1997). Following Smith (1994), every second square (tetrad or hectad) on the regular geographical grid was included in building the regression models, and these models were validated by their ability to classify correctly the remaining squares in the grid. In order to limit the problems associated with interpreting models resulting from stepwise regression procedures (James & McCulloch 1990), both forward and backward stepwise procedures were undertaken. The likelihood ratio test (Tabachnik & Fidell 1996) was used to assess the statistical significance of each environmental predictor variable in the logistic model. To ensure that no important predictor variables were excluded from the stepwise models, the parameter estimates from the backward and forward stepwise models were compared. Selection of the most parsimonious model involved comparison of: (i) the goodness-of-fit; (ii) the matching coefficient (percentage of cases correctly classified); and (iii) Cohen's κ-statistic.

Cohen's κ-statistic measures the proportion of specific agreement between observed and predicted presence/absence scores and assesses the improvement of model predictions over chance. It is the most appropriate method of assessing prediction error for presence/absence data because it makes full use of both presence and absence scores, measures agreement and not association, and is scaled between 0 and 1 (Fielding & Bell 1997). A subjective scale was used to classify the κ-statistic, with values $< 0{\cdot}4$ considered poor, 0·40–0·55 fair, 0·55–0·70 good, 0·70–0·85 very good and $> 0{\cdot}85$ excellent (Prentice *et al*. 1992).

For a binary-dependent variable consisting of species' presence/absence scores, logistic regression predicts continuous probabilities of presence (ranging from 0 to 1) that are then categorized in relation to a critical threshold probability value (usually 0·5) as either present (probability $\geq$ threshold) or absent (probability $<$ threshold). This threshold probability value (T) was selected on the basis of the optimum fit (as estimated by the κ-statistic) between observed and predicted presence/absence rather than a default threshold probability value of 0·5. Although stepwise procedures may identify a subset of predictor variables that provide useful overall predictions, caution should be applied when interpreting the magnitude of the parameter estimates as a measure of their individual influence (James & McCulloch 1990). However, variables with large and statistically significant parameter estimates that are consistently found in the forward and backward models are likely to be influential predictors. For each species at the regional scale, the best-fit logistic models at the tetrad resolution were run for the hectad data and vice versa. Similarly, the best-fit logistic models at the national hectad resolution were run for the regional hectad data and vice versa.

Spatial autocorrelation

The presence of a species in a tetrad or hectad will often tend to increase the likelihood that the species will be present in neighbouring tetrads or hectads (Smith 1994). Furthermore, neighbouring tetrads and hectads will tend to reflect similar environmental conditions (Augustin, Mugglestone & Buckland 1996). This spatial autocorrelation implies that data from individual tetrads or hectads are unlikely to be independent and thus breach one of the assumptions of regression analysis. Although the spatial autocorrelation in the environmental variables may account for some of the spatial autocorrelation in the species distribution data (Smith 1994), it is unlikely that it will account for it completely and thus the environmental variables may not model the species data satisfactorily (Augustin, Mugglestone & Buckland 1996).

The incorporation in logistic regression of a Gibbs sampler, a computer-intensive Markov Chain Monte Carlo routine, is recommended when autocorrelation occurs within presence/absence data (Augustin, Mugglestone & Buckland 1996, 1998; Huffer & Wu 1998). However, the Gibbs sampler algorithm is problematic when the logistic model contains many covariates (Huffer & Wu 1998), and thus could not be used in this study. As an alternative, an autologistic model was developed that, for each species, computed a separate neighbourhood statistic to account for the effects of autocorrelation (Smith 1994; Augustin, Mugglestone & Buckland 1996). As only every second square (tetrad or hectad) on the regular geographical grid was included in the model, the number of second-order neighbours (i.e. the number of presences in the diagonally adjacent squares) was computed for each tetrad or hectad using the Arc/Info GIS. This was then converted into a proportion based on the maximum possible number of nearest neighbours in order to avoid bias at the coast and, in the case of the tetrad model, at the county boundary. The use of second, rather than first (adjacent squares), order neighbours may weaken conclusions about autocorrelation but this does allow the efficacy of the model to be tested. It should be borne in mind that using this

© 2000 British Ecological Society

autologistic regression, rather than the Gibbs sampler approach, may lead to slightly higher misclassification rates (Augustin, Mugglestone & Buckland 1998).

Results

TETRAD RESOLUTION FOR COUNTY DURHAM

Although the best-fit logistic models for all three species produced matching coefficients greater than 80%, this was due to the relative ease of classifying absence in sparse distribution data, and the κ-statistic appeared to provide a more rigorous assessment of the goodness-of-fit (Table 1). The optimum thresholds for species presence were similar for all three species. The logistic fits to the observed distribution of *I. glandulifera* were good (as assessed by the κ-statistic), while the fits to the *F. japonica* and *H. mantegazzianum* data were fair (Table 1). However, only the *I. glandulifera* model was robust in making successful predictions for cases not used in the analysis. Mapping model predictions in relation to observed data for County Durham highlighted these findings (Fig. 2). For *I. glandulifera* the predictions broadly agreed with the observed spatial distribution of this species but were poor for the eastern coastal populations. Spatial correspondence between observed and predicted distributions was also found for *F. japonica*, although the regression model was particularly poor at predicting western upland populations. For *H. mantegazzianum* the model failed to identify southern populations. Inclusion of an autocorrelation term significantly increased the model goodness-of-fit for *F. japonica* ($\chi^2 = 9{\cdot}55$, d.f. 3, $P < 0{\cdot}05$), and particularly for *I. glandulifera* ($\chi^2 = 20{\cdot}45$, d.f. 1, $P < 0{\cdot}001$), but had no effect for *H. mantegazzianum* (Table 1).

Variables common to the models for all three species (Table 2) included deciduous woodland cover (d_wood) and annual temperature sum above 5 °C at the mean elevation (gdd5_mn); these shared variables may partly explain the significant spatial associations between *F. japonica* and both *H. mantegazzianum* ($\chi^2 = 17{\cdot}37$, d.f. 1, $P < 0{\cdot}001$) and *I. glandulifera* ($\chi^2 = 39{\cdot}81$, d.f. 1, $P < 0{\cdot}001$), as well as between *H. mantegazzianum* and *I. glandulifera* ($\chi^2 = 66{\cdot}46$, d.f. 1, $P < 0{\cdot}001$). It appears that *F. japonica* is associated with anthropogenic variables in the form of the proportion of A (primary) and B (secondary) roads (pr_a_rds, pr_b_rds), whereas *I. glandulifera* is similarly related to the proportion of B roads (pr_b_rds) but is the only species with a clear association with riparian habitats (pr_riv). These trends were robust to spatial autocorrelation, although taking it into account further highlighted the importance of anthropogenic land classes (urban, suburban and tilled) in the distribution of *F. japonica*.

HECTAD RESOLUTION FOR COUNTY DURHAM

For all three species, analyses at the hectad resolution for County Durham produced considerably better fits between the observed and predicted values, and the improvement in fit was particularly marked for *F. japonica* (Table 3). The best-fit models were robust and, while the optimum threshold for presence remained similar for *I. glandulifera*, it more or less doubled for the other two species. These trends

Table 1. Summary results of logistic regressions on the presence and absence of three species of non-indigenous weeds in County Durham, assessed at a tetrad (2×2-km squares) resolution. For each species, the results of the best-fit models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable are presented. The goodness-of-fit of the models is assessed by the χ^2-statistic and its associated degrees of freedom (d.f.). All models were statistically significant ($P < 0{\cdot}0001$). The optimum threshold for predicting presence (T), the percentage of correctly classified presences (P), absences (A) and total correct predictions, as well as the κ-statistic, are presented separately for cases used to build the model and for cases upon which the model was tested

Logistic model	χ^2	d.f.	Classification: Original data					Classification: Test data			
			P	A	Total	T	κ	P	A	Total	κ
Fallopia japonica											
Autocorrelation present	79·66	6	58·06	86·39	81·46	0·30	0·41	46·57	88·77	80·17	0·37
Autocorrelation removed	89·21	8	66·13	85·37	82·02	0·27	0·45	47·95	84·21	76·82	0·31
Heracleum mantegazzianum											
Autocorrelation present	40·60	3	30·00	99·40	95·51	0·34	0·41	10·71	98·79	91·90	0·15
Autocorrelation removed	40·60	3	30·00	99·40	95·51	0·34	0·41	10·71	98·79	91·90	0·15
Impatiens glandulifera											
Autocorrelation present	122·31	6	69·23	86·41	82·02	0·36	0·54	70·73	83·33	80·45	0·49
Autocorrelation removed	142·76	7	67·73	88·68	83·15	0·38	0·56	73·17	88·41	84·92	0·59

© 2000 British Ecological Society

Table 2. Climatic, land cover, geological and other environmental variables included in the best-fit logistic models of the presence and absence of three non-indigenous plant species in County Durham at a tetrad resolution. Results of models including spatial autocorrelation and those from which it has been removed by the fit of an additional neighbourhood variable are presented separately. Variables with positive parameter values are presented in upright type whereas variables with negative parameter values are shown italicized. Variables marked with an asterisk were significant terms in the hectad model for County Durham (addition of an autocorrelation term did not alter the hectad model). Only abbreviated variable names are presented, a full description of each variable can be found in the Appendix

Autocorrelation	*Fallopia japonica*		*Heracleum mantegazzianum*		*Impatiens glandulifera*	
Variables	Present	Removed	Present	Removed	Present	Removed
Climate	gdd5_mn*		gdd5_mn*	gdd5_mn	gdd5_mn*	gdd5_mn
Land cover	d_wood	d_wood	d_wood	d_wood	d_wood	d_wood
		suburb	l_bog	l_bog	*i_bg*	*i_bg*
		tilled				
		urban				
Geology	geol_17	geol_17			geol_19*	geol_19
Other	pr_a_rds	pr_a_rds			pr_b_rds	pr_b_rds
	pr_b_rds	*pr_m_rds*			pr_riv	pr_riv
	*pr_m_rds**	prop_fj				prop_ig

may reflect the fact that while the hectad data set comprised only 42 cases for each species, the proportion of hectads occupied by each species was of the order of 50% or more (Table 4). The differences in relative prevalence of the species at the hectad and tetrad resolutions reflected the fact that all three species are widely distributed in County Durham (Fig. 2). At the coarser hectad spatial resolution, the addition of an autocorrelation term had no effect on the goodness-of-fit of models for any of the three species. It appears that bioclimate (gdd5_mn) is of primary importance in determining the distribution of all three species at this spatial scale, with land cover and geology playing only a minor role (Table 3). It was therefore not surprising that at this resolution significant spatial associations were found to exist between *F. japonica* and both *H. mantegazzianum* ($\chi^2 = 8{\cdot}59$, d.f. 1, $P < 0{\cdot}001$) and *I. glandulifera* ($\chi^2 = 8{\cdot}98$, d.f. 1, $P < 0{\cdot}001$), as well as between *H. mantegazzianum* and *I. glandulifera* ($\chi^2 = 11{\cdot}19$, d.f. 1, $P < 0{\cdot}001$).

HECTAD RESOLUTION FOR ENGLAND AND WALES

The goodness-of-fit of the logistic models at the hectad resolution for England and Wales was worse than at either resolution for County Durham (Table 5). The fit was generally poor for all three species and, although incorporating a greater number of statistically significant environmental variables, the models were not particularly robust in their predictive ability. For all three species the matching coeffi-

Table 3. Summary results of logistic regressions on the presence and absence of three species of non-indigenous weeds in County Durham, assessed at a hectad (10×10 km squares) resolution. For each species, the results of the best-fit models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable are presented. The goodness-of-fit of the models is assessed by the χ^2-statistic and its associated degrees of freedom (d.f.). All models were statistically significant ($P < 0{\cdot}0001$). The optimum threshold for predicting presence (T), the percentage of correctly classified presences (P), absences (A) and total correct predictions, as well as the κ-statistic, are presented separately for cases used to build the model and for cases upon which the model was tested

Classification			Original data					Test data			
Logistic model	χ^2	d.f.	P	A	Total	T	κ	P	A	Total	κ
Fallopia japonica											
Autocorrelation present	20·10	2	93·33	100·00	95·24	0·62	0·89	100·00	50·00	85·71	0·59
Autocorrelation removed	20·10	2	93·33	100·00	95·24	0·62	0·89	100·00	50·00	85·71	0·59
Heracleum mantegazzianum											
Autocorrelation present	7·64	1	81·82	80·00	80·95	0·63	0·62	90·91	80·00	85·71	0·71
Autocorrelation removed	7·64	1	81·82	80·00	80·95	0·63	0·62	90·91	80·00	85·71	0·71
Impatiens glandulifera											
Autocorrelation present	18·74	2	100·00	85·71	95·24	0·34	0·89	94·12	75·00	95·00	0·69
Autocorrelation removed	18·74	2	100·00	85·71	95·24	0·34	0·89	94·12	75·00	95·00	0·69

© 2000 British Ecological Society

Table 4. Relative abundance of the three non-indigenous weed species in the three data sets

Proportion of hectads/tetrads occupied	Total number hectads/tetrads	*Fallopia japonica*	*Heracleum mantegazzianum*	*Impatiens glandulifera*
Tetrads in County Durham	714	0·19	0·07	0·24
Hectads in County Durham	42	0·71	0·48	0·74
Hectads in England and Wales	1658	0·70	0·25	0·52

cients were less than 80%, reflecting the fact that while prediction of species presence was improved at this spatial scale, the opposite was true for species absence. As both *F. japonica* and *I. glandulifera* occupied over 50% of all national hectads, these poor fits did not seem to reflect problems of data sparseness (Table 4). There does not appear to be any consistent trend in the way the optimal threshold for presence changes with an increase in spatial extent. Examination of the mapped predictions revealed similar broad correspondence between observed and predicted distributions for *I. glandulifera*, although presence in coastal areas was again poorly predicted (Fig. 2). The predictions for *F. japonica* were particularly poor in central England, whereas the *H. mantegazzianum* predictions were concentrated in the south-east. Inclusion of an autocorrelation term significantly increased the model goodness-of-fit for *I. glandulifera* ($\chi^2 = 13{\cdot}54$, d.f. 1, $P < 0{\cdot}001$) but had no effect for either of the other species (Table 5).

Climate (gdd5_mn) continued to be important for all three species in the national models (Table 6), and they were also all strongly influenced by anthropogenic variables (the proportion of A and B roads). These similarities between the three species may account for the spatial association at this coarser resolution between *F. japonica* and both *H. mantegazzianum* ($\chi^2 = 21{\cdot}10$, d.f. 1, $P < 0{\cdot}001$) and *I. glandulifera* ($\chi^2 = 47{\cdot}42$, d.f. 1, $P < 0{\cdot}001$), as well as for that between *H. mantegazzianum* and *I. glandulifera* ($\chi^2 = 58{\cdot}12$, d.f. 1, $P < 0{\cdot}001$). Again, only *I. glandulifera* was significantly associated with riparian systems (proportion of rivers and the cover of inland water). As in the regional models at the tetrad resolution, these trends were robust to spatial autocorrelation (Table 5).

SPATIAL SCALE-DEPENDENCE

Logistic models developed at one spatial resolution were not particularly successful at predicting the presence/absence of species at a different resolution (Table 7a). This was not surprising as the models were not especially robust when tested at even the same spatial resolution and extent (Tables 1, 3 and 5). For County Durham, there was evidence that using tetrad data to predict hectad distributions resulted in better predictions than scaling down (Table 7a). Scaling-down from hectad to tetrad data resulted in successfully predicting a high percentage of presences but only a low percentage of absences.

Table 5. Summary results of logistic regressions on the presence and absence of three species of non-indigenous weeds in England and Wales, assessed at a hectad (10 × 10-km squares) resolution. For each species the results of the best-fit models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable are also presented. The goodness-of-fit of the models is assessed by the χ^2-statistic and its associated degrees of freedom (d.f.). All models were statistically significant ($P < 0{\cdot}0001$). The optimum threshold for predicting presence (T), the percentage of correctly classified presences (P), absences (A) and total correct predictions, as well as the κ-statistic, are presented separately for cases used to build the model and for cases upon which the model was tested

Classification Logistic model	χ^2	d.f.	Original data P	A	Total	T	κ	Test data P	A	Total	κ
Fallopia japonica											
Autocorrelation present	188·76	11	77·93	63·81	73·55	0·64	0·40	76·53	63·64	72·77	0·38
Autocorrelation removed	187·34	7	77·41	61·87	72·58	0·64	0·38	79·08	63·22	74·46	0·41
Heracleum mantegazzianum											
Autocorrelation present	167·86	12	60·56	80·81	75·60	0·31	0·39	56·16	77·99	72·65	0·32
Autocorrelation removed	–	–	–	–	–	–	–	–	–	–	–
Impatiens glandulifera											
Autocorrelation present	216·51	19	80·50	59·18	70·41	0·47	0·40	76·85	55·28	66·51	0·32
Autocorrelation removed	230·05	18	81·19	60·46	71·38	0·46	0·42	73·84	87·54	66·02	0·32

© 2000 British Ecological Society

Table 6. Climatic, land cover, geological and other environmental variables included in the best-fit logistic models of the presence and absence of three non-indigenous plant species in England and Wales at a hectad resolution. Results of models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable are presented separately. Variables with positive parameter values are presented in upright type whereas variables with negative parameter values are shown italicized. Only abbreviated variable names are presented; a full description of each variable can be found in the Appendix

Autocorrelation	*Fallopia japonica*		*Heracleum mantegazzianum*		*Impatiens glandulifera*	
Variables	Present	Absent	Present	Absent	Present	Absent
Climate	gdd5_men	gdd5_mn	apet_mn gdd5_mn	apet_mn gdd5_mn	gdd5_mn	gdd5_mn
Land cover	d_wood *ds_heath* m_grass r_grass *tilled*	bracken *tilled*	*c_wood* g_heath *sea_est*	*c_wood* g_heath *sea_est*	d_wood ds_moor g_turf i_water m_grass mead_vr suburb tilled	d_wood ds_moor g_turf i_water m_grass mead_vr suburb tilled
Geology	geol_10 geol_13 geol_23 geol_29	geol_10 geol_23	*geol_04* *geol_06* geol_09 geol_14 geol_24	*geol_04* *geol_06* geol_09 geol_14 geol_24	*geol_03* geol_05 geol_10 geol_13 geol_17 geol_19 *geol_23* *geol_29*	geol_05 geol_13 geol_17 geol_19 *geol_23* *geol_29*
Other	pr_a_rds pr_b_rds	pr_a_rds pr_b_rds prop_fj	pr_a_rds pr_b_rds pr_m_rds	pr_a_rds pr_b_rds pr_m_rds	pr_a_rds pr_b_rds pr_riv	pr_a_rds pr_b_rds pr_riv prop_ig

Spatial autocorrelation is a scale-dependent measure and thus would not be expected to be of similar importance when the scale of predictions is changed from tetrad to hectad resolution data. As expected, extrapolating from a regional model to the national scale produced poor predictions, whereas national models were much better predictors of regional distributions (Table 7b). Once again, the trends in the matching coefficients reflected biases in the regional and national models (Tables 3 and 5). As no significant spatial autocorrelation was found for the regional hectad analyses (Table 5), national predictions for the regional hectad data were not expected to improve when spatial autocorrelation was taken into account. Although spatial autocorrelation was found to be significant for *I. glandulifera* in the national hectad analyses (Table 5), it is conceivable that the regional hectad data set was too small to test successfully for spatial autocorrelation at this resolution and extent.

Although comparisons confounded both the effects of resolution and extent, an examination of predictions of national distributions based on tetrad regional models, and regional distributions based on hectad national models, was instructive (Table 8). With the exception of *H. mantegazzianum*, scaling-down from national hectad models to regional tetrad distributions resulted in better predictions than scaling-up from tetrad models to national distributions. Taking account of spatial autocorrelation improved the fit of both the scaled-up and scaled-down predictions, but they remained poor.

Discussion

The ability to model the presence/absence of the three species of non-indigenous weeds varied considerably between species and also at the two different spatial resolutions and extents. Four aspects of the models were scale-dependent: (i) the goodness-of-fit of the predicted values to the observed distributions; (ii) the optimum threshold for species presence; (iii) the significance of spatial autocorrelation; and (iv) the environmental variables included in the best-fit model.

A greater likelihood of a good fit might be expected for tetrad rather than hectad data due to the greater spatial correspondence between species' distributions and environmental data. However, the reverse appears to be true for the County Durham analyses, with significantly better fits to data assessed at the hectad rather than tetrad resolution. Species' distributions in County Durham are certainly more sparse when viewed at the tetrad rather than hectad resolution (Table 4) and this undoubtedly limits the statistical robustness of the tetrad models (Table 1). However, the overriding importance of climate in the hectad models (Table 2) sug-

© 2000 British Ecological Society

Table 7. A summary of the extent to which logistic models produced at one spatial resolution/extent can predict the presence and absence of three non-indigenous species at another spatial resolution/extent. Comparisons are made between: (a) the two spatial resolutions (hectad and tetrad) for a single region (County Durham), and (b) two different spatial extents (national and regional) at a hectad resolution. For each species, predictions from models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable (where this was found to alter the logistic model) are presented. The goodness-of-fit of the predictions is assessed by: the percentage of correctly classified presences (P), absences (A) and total correct predictions; and the κ-statistic. The optimum threshold for predicting presence (T) is also shown

	Tetrad prediction for hectad data					Hectad prediction for tetrad data				
(a) County Durham	P	A	Total	T	κ	P	A	Total	T	κ
Fallopia japonica										
Autocorrelation present	90·00	91·67	90·48	0·60	0·78	80·74	52·50	57·84	0·96	0·20
Autocorrelation removed	86·67	66·67	80·95	0·15	0·53	–	–	–	–	–
Heracleum mantegazzianum										
Autocorrelation present	86·36	75·00	80·95	0·01	0·62	56·25	84·23	82·35	0·81	0·22
Autocorrelation removed	86·36	75·00	80·95	0·01	0·62	–	–	–	–	–
Impatiens glandulifera										
Autocorrelation present	96·77	72·73	90·48	0·03	0·72	88·44	44·73	55·32	0·91	0·21
Autocorrelation removed	100·0	72·73	92·86	0·06	0·80	–	–	–	–	–

	Regional prediction for national data					National prediction for regional data				
(b) Hectad data	P	A	Total	T	κ	P	A	Total	T	κ
Fallopia japonica										
Autocorrelation present	95·08	15·37	70·63	0·73	0·11	100·00	66·67	90·48	0·24	0·74
Autocorrelation removed	–	–	–	–	–	86·67	83·33	85·71	0·50	0·67
Heracleum mantegazzianum										
Autocorrelation present	46·88	76·81	69·30	0·05	0·23	77·27	80·00	78·57	0·05	0·57
Autocorrelation removed	–	–	–	–	–	–	–	–	–	–
Impatiens glandulifera										
Autocorrelation present	76·27	32·28	55·31	1·00	0·09	96·77	63·64	88·10	0·09	0·66
Autocorrelation removed	–	–	–	–	–	96·77	63·64	88·10	0·14	0·66

Table 8. A summary of the extent to which logistic models produced at one spatial scale and resolution can predict the presence and absence of three species of non-indigenous weeds at another spatial scale and resolution. Results of predictions are shown for each species in England and Wales at a hectad resolution using tetrad models developed for species' distributions in County Durham and predictions in County Durham at a tetrad resolution using hectad resolution models developed for species' distributions in England and Wales. For each species, predictions from models including spatial autocorrelation and those from which it has been excluded by the fit of an additional neighbourhood variable are presented. The goodness-of-fit of the predictions is assessed by the percentage of correctly classified presences (P), absences (A) and total correct predictions; and the κ-statistic. The optimum threshold for predicting presence (T) is also shown

	Tetrad prediction for hectad data					Hectad prediction for tetrad data				
Hectad data	P	A	Total	T	κ	P	A	Total	T	κ
Fallopia japonica										
Autocorrelation present	60·57	62·73	61·22	0·31	0·20	42·96	89·64	80·81	0·87	0·34
Autocorrelation removed	63·16	65·93	63·99	0·23	0·25	45·93	89·64	80·37	0·69	0·37
Heracleum mantegazzianum										
Autocorrelation present	58·65	64·41	62·97	0·68	0·19	54·17	70·27	69·19	0·35	0·09
Autocorrelation removed	–	–	–	–	–	–	–	–	–	–
Impatiens glandulifera										
Autocorrelation present	64·17	52·53	58·62	0·63	0·17	73·41	65·43	67·37	0·67	0·30
Autocorrelation removed	69·93	56·71	63·63	0·70	0·27	73·41	69·50	70·45	0·64	0·35

© 2000 British Ecological Society

gests that the coarser resolution analyses characterize, albeit successfully, the simple difference between upland and lowland hectads and the relatively greater invasibility of the latter. Thus while the hectad models present good fits, they provide little information on the relative susceptibility to invasion of different lowland habitats.

In contrast to the good fits for hectad distribution data in County Durham, national analyses at a hectad resolution were, at best, only fair. This is in contrast to Smith (1994), who modelled successfully plant distributions in the British Isles at a hectad resolution using a comparable statistical approach on a similar set of environmental variables. A fundamental assumption in correlative modelling of species' distributions is that species should be in equilibrium with their environment (Franklin 1995). In the British Isles, the distributions of all three non-indigenous species are still increasing (Rich & Woodruff 1996; Fig. 1a). This non-equilibrium distribution will limit the goodness-of-fit of regression models in three ways. First, logistic models are likely to perform poorly when either species' presences or absences are sparse. Hence, the sparse species' distributions at the tetrad resolution resulted in models that were relatively poor at predicting species presence (Table 1). The optimum threshold for species presence appears to be influenced by the sparseness of data, decreasing as the species distribution becomes more sparse and thus acting to increase the classification of species presence. Secondly, if a species does not occupy all suitable habitats then it is unlikely that the variables entered into regression models will be sufficiently comprehensive and thus predictions will be weak. Thirdly, even if the environmental variables capture the essence of the species distribution, models will consistently predict species presence in areas where the species has yet to invade. Thus the closer a species is to maximum potential range, the better the observed goodness-of-fit. This has considerable implications for any attempts at modelling the spread of invasive species. Problems of modelling the distribution of non-indigenous plant species have previously been found in New Zealand where, for some species, locations predicted as being climatically suitable had as yet not been colonized due to limited rates of dispersal (Wilson *et al*. 1992).

Interpretation of whether a species has expanded to occupy entirely its potential spatial range is a scale-dependent issue. In England, measures of the distribution of *F. japonica*, *H. mantegazzianum and I. glandulifera* at a vice-county scale (average size of a vice-county is 2200 km^2) reveal most vice-counties to have been colonized (Fig. 1b). However, at a hectad scale the species can still be seen to be increasing (Fig. 1a). It is therefore possible that species' distributions, when viewed at coarser spatial scales, may be more likely to be interpreted as having reached stasis (Wiens 1989). This may explain why Beerling, Huntley & Bailey (1995) were able to model the European distribution of *F. japonica* at a 50×50-km scale particularly well (κ-statistic = 0·69) using only bioclimatic variables. This effect may also help to explain why the regional hectad models for County Durham produced significantly better fits than the tetrad models. This implies that studies examining hierarchical constraints on species' distributions should control for both the extent and grain of their data (*sensu* Wiens 1989).

The distribution of an invasive species spreading across a landscape might be expected to exhibit spatial autocorrelation because the probability that an area will be colonized will be a function of its distance from neighbouring populations. In general, the scale over which spatial autocorrelation might be expected will reflect the dispersal efficiency of the invading species. Spatial autocorrelation is therefore expected to be more evident at finer spatial scales. This is the case for *F. japonica*, a species that spreads vegetatively and through dispersal of rhizomes (Beerling, Bailey & Conolly 1994) and for which weak autocorrelation was present in the tetrad data but not in either of the two hectad data sets. Although of greater significance at the tetrad resolution, spatial autocorrelation occurred for *I. glandulifera* at tetrad and national hectad scales. The lack of significant spatial autocorrelation at the regional hectad distribution may be due to the small data set used. The greater spatial scale of autocorrelation compared with *F. japonica* may reflect the efficient long-distance seed dispersal of this species in waterways (Beerling & Perrins 1993). No evidence of significant spatial autocorrelation was found for *H. mantegazzianum*, although this species is believed to disperse its fruits widely by both wind and water (Tiley, Dodd & Wade 1996). Although local aggregations of this species can be seen for both the regional and national distributions (Fig. 2), the majority of records are relatively isolated. This may reflect the importance of long-distance dispersal events (possibly anthropogenic) in the initial spread of this species.

Even on the occasions where the fit of the models to the observed distribution data was only fair, the environmental variables identified as significant were consistent with qualitative descriptions of the habitat characteristics of these species (Beerling & Perrins 1993; Beerling, Bailey & Conolly 1994; Tiley, Dodd & Wade 1996). Fewer variables were incorporated into the best-fit models at the regional hectad resolution that at either the regional tetrad or national hectad resolutions (Tables 2 and 6). This may simply reflect the smaller data set used in the regional hectad model and the subsequent decrease in degrees of freedom resulting in the selection of fewer significant variables. The greater prominence of geological variables at the national hectad resolu-

© 2000 British Ecological Society

tion is probably an artefact as geology was the only environmental variable measured uniquely at a hectad resolution and the importance of geology at the tetrad resolution may thus have been underestimated. Nevertheless, the majority of the environmental variables identified as important at the tetrad resolution were important at hectad resolutions for *F. japonica* (climate, deciduous woodland cover and the proportion of A and B roads) and *I. glandulifera* (climate, deciduous woodland cover, hard sandy sedimentary rocks and the proportion of A roads and rivers). However, for *H. mantegazzianum* only one environmental variable (gdd5_mn) was shared by the regional tetrad and national hectad models.

At the regional extent (County Durham), scaling-up from tetrad to hectad species' distributions was relatively successful. This reflected the fact that the tetrad models encompassed all the environmental variables found to be significant at the coarser hectad resolution. Scaling-down from hectad resolution to a tetrad resolution produced poor predictions for exactly the same reason. The coarser resolution model was too unrefined to model fine-scale distributions successfully. The same logic explains why, at a coarse hectad resolution, regional models were poor predictors of national species' distributions. The degree of environmental variation within a small region, e.g. County Durham, is likely to be insufficient to encompass the national variation across England and Wales. This is not the case for the national models, which are capable of successfully modelling species' distributions within particular national subregions. It therefore appears that scaling-up from fine to coarse resolution is appropriate when spatial extent is held constant, and focusing-down from large to small spatial extents is appropriate when data resolution is held constant. It is therefore not surprising that attempts to examine scale-dependence simultaneously across different spatial resolutions and extents are unlikely to be successful (Table 8).

The key issue in the study of scale-dependence is the extent to which processes acting at large or small spatial resolutions determine species' distributions, and hence what the most appropriate scale may be to predict future changes in species distribution. This study has revealed that, although correspondence exists in the key environmental variables at different spatial resolutions, scaling down from a coarse to a finer resolution was not particularly successful. Yet, if spatial resolution requirements lead to sacrifices in the spatial extent of data, the generality of models is compromised. This highlights the importance of modelling species' distributions at more than one spatial scale (Ormerod & Watkinson 2000). This is a logical approach from the viewpoint of both statistical modelling and weed management. Control and conservation efforts will generally be applied at relatively fine spatial scales (Roblin 1994; Sampson 1994), and predictions at fine spatial resolutions will therefore be most instructive to these efforts, yet for long-term management a knowledge of processes acting at larger spatial scales will be essential (Wadsworth *et al.* 2000). For example, the frequent importance of climatic factors in the distribution models of all three species suggests their present ranges may be strongly influenced by future environmental change.

© 2000 British Ecological Society

Acknowledgements

This research was funded by the Natural Environment Research Council and the Scottish Executive Rural Affairs Department under the Large-Scale Processes in Ecology and Hydrology Thematic Programme. The authors would like to thank Chris Preston for providing the BRC species distribution data for Great Britain and the Reverend Gordon Graham for making available the species distribution data for County Durham. Wolfgang Cramer provided the spline surfaces used to interpolate the climatic data. The authors are very grateful to four anonymous referees for the detailed comments on a previous version of the manuscript.

References

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 677–692.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1998) The role of simulation in modelling spatially correlated data. *Environmetrics*, **9**, 175–196.

Baillie, S.R., Sutherland, W.J., Freeman, S.N., Gregory, R.D. & Paradis, E. (2000) Consequences of large-scale processes for the conservation of bird populations. *Journal of Applied Ecology*, **37** (Suppl. 1), 88–102.

Ball, D.F., Radford, G.L. & Williams, W.M. (1983) Land characteristic data banks developed from map-derived material. *Ecological Mapping from Ground, Air and Space* (ed. R.M. Fuller), pp. 28–38. Institute of Terrestrial Ecology, Cambridge, UK.

Beerling, D.J. & Perrins, J.M. (1993) Biological flora of the British Isles: *Impatiens glandulifera* (Royle). *Journal of Ecology*, **81**, 367–382.

Beerling, D.J., Bailey, J.P. & Conolly, A.P. (1994) Biological flora of the British Isles: *Fallopia japonica* (Houtt.) Ronse Decraena. *Journal of Ecology*, **82**, 959–979.

Beerling, D.J., Huntley, B. & Bailey, J.P. (1995) Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science*, **6**, 269–282.

Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, **30**, 478–495.

CHEST (1993) *Bartholomew Digital Map Data 1993 GB 1: 250, 000*. University of Bath, Bath, UK.

Clement, E.J. & Foster, M.C. (1994) *Alien Plants of the British Isles*. BSBI, London, UK.

Cowley, M.J.R., Wilson, R.J., Leon-Cortes, J.L., Gutierrez, D., Bulman, C.R. & Thomas, C.D. (2000) Habitat-based statistical models for predicting the spatial distribution of butterflies and day-flying moths in a

fragmented landscape. *Journal of Applied Ecology*, **37** (Suppl. 1), 60–72.

Crawley, M.J. (1986) What makes a community invasible? *Colonization, Succession and Stability* (eds A.J. Gray, M.J. Crawley & P.J. Edwards), pp. 429–453. Blackwell Scientific Publications, Oxford, UK.

Crawley, M.J., Harvey, P.H. & Purvis, A. (1996) Comparative ecology of the native and alien floras of the British Isles. *Philosophical Transactions of the Royal Society of London, Series B*, **351**, 1251–1259.

Ellis, R.G. (1993) *Aliens in the British Flora: An Account of Some of Our Plant Invaders*. National Museum of Wales, Cardiff, UK.

ESRI (1998) *Arc/Info*. Environmental Systems Research Institute Inc., Redlands, CA.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.

Fuller, R.M., Groom, G.B. & Jones, A.R. (1994) An automated classification of the Landsat thematic mapper data. *Photogrammetric Engineering and Remote Sensing*, **60**, 553–562.

Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R. & Lawton, J.H. (2000) Abundance–occupancy relationships. *Journal of Applied Ecology*, **37** (Suppl. 1), 39–59.

Graham, G.G. (1988) *The Flora and Vegetation of County Durham*. The Durham Flora Committee and the Durham County Conservation Trust, Durham, UK.

Hobbs, R.J. & Humphries, S.E. (1995) An integrated approach to the ecology and management of plant invasions. *Conservation Biology*, **9**, 761–770.

Huffer, F. & Wu, H. (1998) Markov chain Monte Carlo for auto-logistic regression models with application to the distribution of plant species. *Biometrics*, **54**, 509–524.

Huntley, B., Berry, P.M., Cramer, W. & McDonald, A.P. (1995) Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography*, **22**, 967–1001.

Hutchinson, M.F. (1989) *A New Objective Method for Spatial Interpolation of Meteorological Variables from Irregular Networks Applied to the Estimation of Monthly Mean Solar Radiation, Temperature, Precipitation and Windrun*. CSIRO Technical Report, CSIRO, Canberra, Australia.

Jalas, J. & Suominen, J. (1972–94) *Atlas Florae Europaeae: Distribution of Vascular Plants in Europe*. Societas Vanamo, Helsinki, Finland.

James, F.C. & McCulloch, C.E. (1990) Multivariate analysis in ecology and systematics: panacea or Pandora's box. *Annual Review of Ecology and Systematics*, **21**, 129–166.

Leemans, R. & Cramer, W. (1991) *The IIASA Database for Mean Monthly Values of Temperature, Precipitation and Cloudiness of a Global Terrestrial Grid*. Research Report RR-91–18. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Manchester, S.J. & Bullock, J.M. (2000) The impact on non-native species on UK biodiversity and the effectiveness of control. *Journal of Applied Ecology*, in press.

May, R.M. (1989) Levels of organization in ecology. *Ecological Concepts* (ed. J.M. Cherrett), pp. 339–363. Blackwell Scientific Publications, Oxford. UK.

Norusis, M.J. (1994) *SPSS Advanced Statistics 6.1*. SPSS Inc., Chicago, IL.

Ormerod, S.J. & Watkinson, A.R. (2000) Editorial. Large-scale ecology and hydrology: an introductory perspective from the editors of the *Journal of Applied Ecology*. *Journal of Applied Ecology*, **37** (Suppl. 1), 1–5.

Perring, F.H. & Walters, S.M. (1976) *Atlas of the British Flora*, 2nd edn. EP Publishing for the Botanical Society of the British Isles.

Perrins, J., Fitter, A. & Williamson, M. (1993) Population biology and rates of invasion of three introduced *Impatiens* species in the British Isles. *Journal of Biogeography*, **20**, 33–44.

Pettifor, R.A., Caldow, R.W.G., Rowcliffe, J.M., Goss-Custard, J.D., Black, J.M., Hodder, K.H., Houston, A.I., Lang, A. & Webb, J.N. (2000) Spatially explicit, individual-based, behavioural models of the annual cycle of two migratory goose populations. *Journal of Applied Ecology*, **37** (Suppl. 1), 103–135.

Prentice, C.I., Cramer, W., Harrison, S.P., Leemans, R., Monserud, R.A. & Solomon, A.M. (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, **19**, 117–134.

Rich, T.C.G. & Woodruff, E.R. (1996) Changes in the vascular floras of England and Scotland between 1930 and 1960 and 1987–88: the BSBI monitoring scheme. *Biological Conservation*, **64**, 370–392.

Roblin, E. (1994) Alien invasive weeds – an example of National Rivers Authority sponsored research. *Ecology and Management of Invasive Riverside Plants* (eds L.C. de Waal, L.E. Child, P.M. Wade & J.H. Brock), pp. 189–194. John Wiley & Sons, Chichester, UK.

Sampson, C. (1994) Cost and impact of current control methods used against *Heracleum mantegazzianum* (giant hogweed) and the case for investigating a biological control programme. *Ecology and Management of Invasive Riverside Plants* (eds L.C. de Waal, L.E. Child, P.M. Wade & J.H. Brock), pp. 55–65. John Wiley & Sons, Chichester, UK.

Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, **4**, 47–61.

Tabachnik, B.G. & Fidell, L.S. (1996) *Using Multivariate Statistics*, 3rd edn. Harper Collins, New York, NY.

Thompson, K., Hodgson, J.G. & Rich, T.C.G. (1995) Native and alien invasive plants: more of the same? *Ecography*, **18**, 390–402.

Tiley, G.E.D., Dodd, S. & Wade, P.M. (1996) Biological flora of the British Isles: *Heracleum mantegazzianum* (Sommier & Levier). *Journal of Ecology*, **84**, 297–319.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L., Rejmanek, M. & Westbrooks, R. (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology*, **21**, 1–16.

Wadsworth, R.A., Collingham, Y.C., Willis, S.G., Huntley, B. & Hulme, P.E. (2000) Simulating the spread and management of alien riparian weeds: are they out of control? *Journal of Applied Ecology*, **37** (Suppl. 1), 28–38.

Wiens, J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.

Wilson, J.B., Rapson, G.L., Sykes, M.T., Watkins, A.J. & Williams, P.A. (1992) Distributions and climatic correlations of some exotic species along roadsides in South Island, New Zealand. *Journal of Biogeography*, **19**, 183–193.

Received 19 January 1999; revision received 24 May 2000

© 2000 British Ecological Society

Appendix

A list of the environmental variables used in the fitting of logistic regression models to the presence/absence of *Fallopia japonica*, *Heracleum mantegazzianum* and *Impatiens glandulifera* at both tetrad and hectad resolutions. Sources of these data are described in the text. Both the abbreviated codes used in the text and the spatial resolution of the data are indicated.

Code	Variable description	Code	Variable description
ITE LAND COVER (percentage cover based on data resolved for 25 × 25 m squares)			
beach	beach/mudflat/cliffs	mead_vr	meadow/verge/semi-natural
bracken	bracken	os_heath	open shrub heath
c_wood	coniferous woodland	os_moor	open shrub moor
d_wood	deciduous woodland	r_grass	rough/marsh grass
ds_heath	dense shrub heath	r_weed	ruderal weed
ds_moor	dense shrub moor	s_marsh	saltmarsh
f_forest	felled forest	scr_orch	scrub/orchard
g_heath	grazed heath	sea_est	sea/estuary
g_turf	mown/grazed turf	suburb	suburban/rural development
I_bg	inland bare ground	tilled	tilled land
I_water	inland water	u_bog	upland bog
l_bog	lowland bog	unclass	unclassified
m_grass	moorland grass	urban	continuous urban
BIOCLIMATIC (interpolations computed at tetrad and hectad resolutions)			
apet_mn	ratio of actual to potential evapotranspiration at mean elevation		
gdd5_mn	growing degree days above 5(C at mean elevation		
ROADS and RIVERS			
Number of 500 × 500-m squares containing one of the following variables expressed as a proportion of the maximum number of squares in a tetrad or hectad			
pr_a_rds	A road		
pr_b_rds	B road		
pr_m_rds	M road		
pr_riv	River		
SPATIAL AUTOCORRELATION (determined using either hectad or tetrad data)			
Number of second order neighbours occupied by each plant species (expressed as a proportion of the maximum number which may be occupied)			
prop_fj	*Fallopia japonica*		
prop_hm	*Heracleum mantegazzianum*		
prop_ig	*Impatiens glandulifera*		
GEOLOGY (hectad resolution and scaled to tetrad resolution)			
geol_01	Pre-Cambrian		
geol_02	Cambrian, Ordovician and Silurian		
geol_03	Devonian		
geol_04	Carboniferous		
geol_05	Permian and Triassic		
geol_06	Jurassic		
geol_07	Cretaceous		
geol_08	Tertiary		
geol_09	Igneous rocks of unspecified stratigraphic age		
geol_10	Intrusive acid igneous rocks and lithologically similar metamorphic rocks		
geol_11	Extrusive acid igneous rocks and lithologically similar metamorphic rocks		
geol_12	Intrusive basic igneous rocks and lithologically similar metamorphic rocks		
geol_13	Extrusive basic igneous rocks and lithologically similar metamorphic rocks		
geol_14	Igneous or metamorphic rocks of high ferro-magnesian composition		
geol_15	Metamorphic quartzose rocks (quartzite)		
geol_16	Metamorphic calcareous rocks (metamorphosed limestone)		
geol_17	Hard sandy sedimentary rocks (hard sandstones, grits)		
geol_18	Soft sandy sedimentary rocks (soft sandstones and sands)		
geol_19	Hard clayey sedimentary rocks (slates and shales)		
geol_20	Soft clayey sedimentary rocks (clays)		
geol_21	Hard calcareous sedimentary rocks (limestone)		
geol_22	Soft calcareous sedimentary rocks (chalk)		
geol_23	Bedrock		
geol_24	Boulder clay or other drift cover with a medium-to fine-textured matrix		
geol_25	Sand and gravel drift cover		
geol_26	'Clay-with-flints' cover		
geol_27	Alluvial cover		
geol_28	Wind-blown cover		
geol_29	Peaty cover		

© 2000 British Ecological Society