

Predicting weed distribution at the landscape scale: using naturalized *Brassica* as a model system

D. A. Peltzer*, S. Ferriss and R. G. FitzJohn

Landcare Research Manaaki Whenua, PO Box 40, Lincoln 7640, New Zealand

Summary

1. Quantifying and predicting the distribution of naturalized plant species is a major concern for weed risk assessment and understanding the potential impacts of gene-flow between crop species and feral or naturalized populations.

2. We developed a rapid field assessment to quantify the distribution of naturalized *Brassica* populations at the landscape scale, using the Canterbury Plains region of New Zealand as a model system. Internationally, brassicas are one of the most widely cultivated crops, and have well-documented problems of crop escapes, gene-flow and hybridization with wild relatives. Brassicas are cultivated intensively in the study system for both seed production and forage, are widely naturalized and have raised concerns for gene-flow from feral populations to cultivated varieties.

3. Generalized linear models (GLM) were constructed in order to predict the presence–absence of brassicas using survey data of naturalized populations located in 50 3 × 3 km plots. This model was validated subsequently both by ground-truthing predicted presence–absence in a second field survey, and by comparing cumulative probability distributions from a presence-only botanical survey.

4. The best model produced from the field survey predicted naturalized *Brassica* presence–absence robustly ($r^2 = 0.52$) and was well validated against results from the presence-only survey carried out in the same year. Results from the second field survey during a different growing season did not match predictions from the original model, suggesting that strong between-year differences influence large-scale *Brassica* distribution. The differences between years were driven by an overall increase in the prevalence of brassicas over the course of the study that was not associated with any predictive variable or climate.

5. *Synthesis and applications.* These results have important implications both for developing models of species distribution and for predicting gene-flow risk at the landscape scale. Despite constructing a model that was subsequently validated from presence-only data in one year, it did not predict *Brassica* distribution in a subsequent year; this suggests that between-year or stochastic factors operate strongly and are a major consideration for risk assessment. Models used to predict species distributions, in general, need to account for these between-year effects either by incorporating mechanistic processes or long-term data.

Key-words: *Brassica*, crop escapes, distribution and abundance, field survey method, generalized linear models, geographical information systems, monitoring, weed risk assessment

Introduction

The increasing need to understand and manage plant invasions has generated international interest in predicting the distribution and abundance of weeds in both anthropogenic and natural systems. In addition to the widely documented environmental impacts of weeds on community and ecosystem

processes (e.g. Mack *et al.* 2000; Levine *et al.* 2003), there is an increasing realization that risks can also arise from gene flow between crops and wild relatives or through the naturalization of cultivated species (e.g. Raybould & Gray 1993; Scheffler & Dale 1994; Beckie *et al.* 2001; Gray 2004; Pilon & Prendeville 2004; Snow *et al.* 2005). These latter concerns are driven largely by those seeking to understand and quantify the risks associated with the release of genetically modified organisms, but are also more widely applicable to environmental

*Correspondence author. E-mail: PeltzerD@LandcareResearch.co.nz

weeds that escape management (Wolfenbarger & Phifer 2000; Armstrong *et al.* 2005; Andow & Zwahlen 2006). An unresolved problem is how to quantify these risks at appropriate spatial scales to describe the distribution and abundance of weeds that have the potential for gene flow or hybridization between managed species, feral plants and fully naturalized populations.

Despite international concern surrounding weed distributions and the likelihood of gene flow from naturalized plant populations, no international consensus on how to quantify or assess these risks has been developed (Ellstrand, Prentice & Hancock 1999; Pilson & Prendeville 2004). For example, some countries in the European Union have used regional surveys to detect feral crop species or hybrids in naturalized populations, and then generate spatial information about population distributions to quantify gene-flow risk (e.g. de Vries, van der Meijden & Brandenburg 1992; Flannery, Meade & Mullins 2005). In contrast, the United States and Canada generally assume that gene-flow will take place and should be managed accordingly (e.g. Wilkinson *et al.* 2003b; Traxler 2006). Australia has assumed that gene-flow will occur but also attempts to understand these risks and their impacts (e.g. Rieger, Preston & Powles 1999; Rieger *et al.* 2002; Godfree *et al.* 2004). Others have used remote-sensing technology to generate national-scale estimates of gene-flow, for example, of brassicas in the United Kingdom (Elliott *et al.* 2004). The fundamental constraint on these assessments is the trade-off between detailed studies of gene-flow or risk at relatively few sites vs. large-scale surveys of weed distributions.

Here, we develop an approach to generate spatial models quickly for predicting the distribution of naturalized *Brassica* populations at the landscape scale. Brassicas are one of the most commonly grown genetically modified crop species, are known to escape from cultivation, have well-documented gene-flow risks via pollen, seeds and crop escapes and can hybridize with wild relatives (e.g. Rieger, Preston & Powles 1999; Wolfenbarger & Phifer 2000; Beckie *et al.* 2001; Pessel *et al.* 2001; Wilkinson *et al.* 2003a). We are not examining gene-flow in this study *per se*, but the approach we generate here can be used to quantify the risk of gene-flow via crop escapes and naturalization, and is more widely applicable to predictions of species presence-absence at the landscape scale. This approach uses a combination of field surveys, geographic information systems (GIS) and spatial models to quantify the distribution and abundance of naturalized brassicas in the most important arable region in New Zealand for this genus. Based on the studies above, we predict that propagule pressure (i.e. proximity to seed storage facilities or transportation routes through seed spillage) should drive naturalized *Brassica* distribution, as should climatic favourableness. However, our study was not designed to generate data for formal hypothesis testing or to generate mechanistic models, but rather the overall aim of our project was to develop predictive models of naturalized *Brassica* distribution that could be used subsequently for understanding the risk associated with gene-flow or hybridization at a landscape scale.

Methods

STUDY SYSTEM

New Zealand is not currently growing genetically modified crops at a commercial scale; however, the moratorium banning their production has recently been lifted, even though the risks involved with this activity are largely unknown (Royal Commission on Genetic Modification 2002). The vast majority of New Zealand's *Brassica* seed and forage crops are grown in the arable portion of the Canterbury Plains, South Island, New Zealand. There is a relatively long history of cultivating brassicas for both seed and forage production in this region (Heenan, FitzJohn & Dawson 2004 and references therein). Both feral and naturalized *Brassica* populations in this system are comprised largely of *B. rapa*, with smaller numbers of *B. napus*, *B. juncea* and *B. oleracea*. These populations occur typically in disturbed habitats such as road verges or coastal areas (Heenan, FitzJohn & Dawson 2004), are ruderal species characterized by rapid growth and high fecundity (*sensu* Grime 2001) and can hybridize readily (Heenan *et al.* 2007; FitzJohn *et al.* in press); these characteristics make brassicas potentially risky with respect to gene-flow. We used the Canterbury Plains as a model system for quantifying the distribution of naturalized brassicas at the landscape scale. An advantage of using this system for generating spatial models is that the Plains have a roughly uniform climate (1971–2000 annual precipitation = 648 mm, mean annual temperature = 12.1 °C, National Institute for Water and Atmosphere, New Zealand) topography and soils over the region studied (Fig. 1 inset; see also Leathwick *et al.* 2003), but have diverse farming systems that are typically small holdings (< 100 ha) of mixed arable and pasture (dairy, sheep) production.

GIS AND FIELD SURVEY DESIGN

We used randomly located plots placed throughout the study system in order to generate unbiased observations of both presence and absence of *Brassica* populations. Such systematic surveys have advantages over convenience samples, or presence-only botanical surveys or herbarium data that lack observations or collections of absences, and hence introduce a source of unknown observational bias. We defined a broad study area using the ArcView GIS system as land within the Canterbury Plains regions having a slope of less than 10° (a total area of 9543 km²); this included the arable plains but excluded mountain lands, high country grazing systems and Banks Peninsula (Fig. 1). Within this region, we generated 100 randomly located 3 × 3 km plots, and then excluded those plots that overlapped, were located over rivers, lakes or towns and contained < 3 km of roads (see Heenan, FitzJohn & Dawson 2004 for details). The decision to use 50 large 3 × 3 km sample plots rather than more numerous, smaller plots was made due largely to logistical constraints. We estimated 50 plots to be the largest sample size that could be surveyed feasibly over the peak flowering period, and provided good coverage of the study system (i.e. we surveyed 4.7% of the total area). In addition, a 3 × 3 km area is approximately the isolation distance between farms that *Brassica* seed growers use in New Zealand, and so is relevant to managing gene-flow risk.

FIELD SURVEYS

Detailed field survey methods are presented elsewhere (Heenan, FitzJohn & Dawson 2004). Briefly, we surveyed, by car, every road

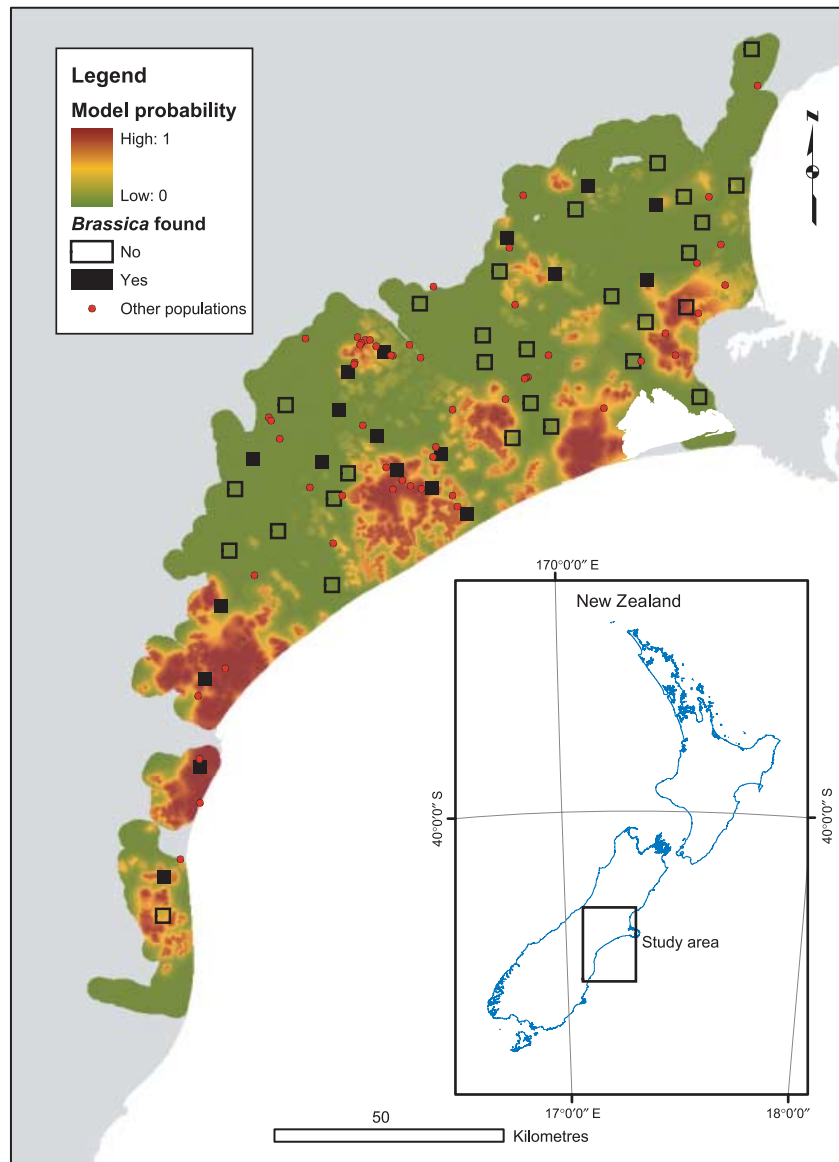


Fig. 1. Model output for the probability of naturalized *Brassica* presence (red) or absence (green) on the Canterbury Plains, South Island, New Zealand generated from field survey plots in 2003. Squares represent sampling plots in which brassicas were either present (filled) or absent (open). Circles represent other *Brassica* populations observed en route between sampling plots (presence-only survey). Inset panel: the study system encompasses the main arable region within which brassicas are grown in New Zealand.

within each random plot during late September and October 2003 to coincide with the peaking flowering period of brassicas (i.e. the southern hemisphere spring). The geographical position of any *Brassica* population observed on roadside verges was recorded both within the 50 sample plots and en route between plots (i.e. to generate presence-only survey data). Roadside verges are managed in such a way (i.e. disturbed through spraying, mowing or soil perturbation) that they are the primary habitat for *Brassica* in this system. In addition, c. 50-km stretches of two large braided river systems in the region were surveyed by helicopter; only a single *Brassica* plant was observed in this way (Heenan, FitzJohn & Dawson 2004). Our roadside survey included riparian habitats and minor waterways that are largely comprised of managed drainage ditches. Such roadside or streamside surveys have been used internationally for gene-flow or hybridization risk assessment and quantifying the distribution of weeds at the landscape scale (e.g. Wilkinson *et al.* 2003a; Rentch *et al.* 2005). Plants outside cultivation are by definition not planted, and represent both crop escapes (feral populations) and fully naturalized populations; these cannot be distinguished in the field, and we term all brassicas outside cultivation as 'naturalized' here.

For each naturalized population, we recorded species identity (*B. oleracea*, *B. rapa*, *B. napus* or *B. juncea*) and population size. Populations were defined by flowering individuals only, and were not distinguished by other methods (e.g. flow cytometry or genetic markers). Presence-absence was recorded at the plot level and was not defined as minimum distance to another *Brassica* population of the same or different species. Brassicas are easy to detect in the study system because of their conspicuous flowering and similar flowering phenology during our surveys; in addition, no native species in this system have similar floral displays.

In consultation with local *Brassica* and weed experts (P. Heenan, botanist, Landcare Research; A. Stewart, crop breeder, PGG Seeds; E. Johnson, agricultural consultant, Agriquality New Zealand), we identified factors that varied over a regional scale that might correlate with *Brassica* abundance and calculated spatial information about each sampling plot including: distance to seed producing farms, seed storage facilities, distance from a plot to the nearest coast, state and provincial highways, lake, railway, river or town and length of roads within a plot. These factors are related to either propagule pressure (i.e. seed production or dispersal) or climatic favourableness (e.g.

distance to coast). Seed production farms and seed storage facilities were identified using a crop producers database which provides details of *Brassica* crop production on farms (AgriBase, Agriquality, New Zealand). Distance to the nearest coast, state and provincial highways, lake, railway, river or town were generated in a GIS from Land Information New Zealand (LINZ) spatial data layers. To calculate the distance between plots and these factors, we generated the distance from the edge of the plots in the GIS system; where a feature occurred within a plot, this distance was zero. We also calculated spatial information about each sampling plot for climatic factors and soil types (from Leathwick *et al.* 2003), but were not included as these factors did not vary appreciably across the Canterbury Plains (data not shown). We also included road length contained within each plot as a factor in the models as a measure of sampling effort. These factors are unlikely to control *Brassica* distribution directly, but are probably correlated with various factors or mechanisms that are. For example, distance from the coast represents a complex gradient of urbanization, elevation and other factors, whereas distance to roads capture disturbance regime, largely through verge management (i.e. spraying, mowing or soil perturbation) or livestock transportation (e.g. sheep herding). The hypotheses of propagule pressure and climatic favourableness are implicit in our model development. However, our study was not designed to generate data for formal hypothesis testing or to generate mechanistic models. Our motivation for choosing these variables was to assess rapidly the distribution of naturalized brassicas rather than reveal the biological mechanisms controlling their distribution; this would require either fine-scale measurements, experimental manipulations or long-term data which are not available.

STATISTICAL ANALYSES

Because we found relatively few populations within the random plots (usually a single population), we calculated presence-absence at the plot level for any species of *Brassica* (19 of 50 plots contained *Brassica* populations), rather than considering the number of populations within a plot or population size. We constructed generalized linear models (GLMs) with a binomial error distribution (i.e. logistic regressions) to predict the presence-absence of *Brassica* populations using the landscape-level factors described above (R Development Core Team). Because the vast majority (> 90%) of brassicas encountered were *B. rapa*, there were insufficient data to run GLMs for the other species recorded. Preliminary models did not differ using either *B. rapa* data alone or including data for all *Brassica* species; all subsequent analyses included data for all *Brassica* species. All predictor variables were included in the initial model, and the model was refined by removing variables by backwards stepwise deletion; model fits were compared using Aikake's information criterion (AIC). Predictor variables were square root-transformed prior to analyses to correct for data skewing and zero values. Although the sample size used here is modest ($n = 50$ plots), this is more than sufficient for the two- or three-factor models that were fitted to the data (see below).

Using the best GLM model predicting *Brassica* presence-absence, we generated a predictive GIS layer across the entire sample region. Although others have used alternative approaches such as generalized additive models (GAM) or classification and regression trees (CART) to produce distributions of species using presence-absence data, these approaches do not generate predictive models that can be easily validated subsequently (GAM) or require larger data sets than we have available to produce robust predictions (CART) (Austin 2002; Bourg, McShea & Gill 2005; see also Thuiller, Araújo

& Lavorel 2003; Drake, Randin & Guisan 2006). To address the issue of whether low visibility of *Brassica* populations could result in false negatives for population presence (i.e. small populations were not detected), we ran additional GLM models by assigning additional presence values exhaustively to all combinations of one, two or three of the 31 plots in which no *Brassica* populations were found. These analyses revealed that the models we used were robust for predicting *Brassica* distribution even with a 5–10% rate of false absences; thus increasing the population threshold of detection (i.e. including additional *Brassica* populations) did not change our models of *Brassica* distribution appreciably. This approach is more appropriate than splitting the data set into model training and verification portions because of modest sample sizes.

MODEL VALIDATION AND RESURVEY

We validated the model in two ways: (1) using presence-only data from the original 2003 survey and (2) using an independent field survey in a different growing season (2005). We first validated the model using *Brassica* populations observed en route between plots during the original 2003 field survey; these data thus represent a quasi-independent data set of populations. Although these data are from an opportunistic presence-only survey, they yielded many of the *Brassica* populations observed (60 of 107 total populations). To determine if we found such populations at random, or in keeping with model predictions, we compared the distribution of predicted probabilities where *Brassica* populations were found with the distribution of probabilities across the sample space (within 20 m of roads). To test if the model predicted *Brassica* presence better than by chance, we compared the distributions with a one-tailed Kolmogorov-Smirnov (K-S) test. We also carried out a second field survey in 2005 in the same study system using newly generated plots rather than resampling the (2003) plots. The 2005 field plots were located randomly within areas predicted by our best GLM model (derived from the 2003 survey plots) to either contain brassicas (> 50% predicted probability of *Brassica* presence; $n = 22$ plots) or not (< 5% probability; $n = 22$ plots). As for the 2003 survey, we recorded the number of populations and plants, and generated presence-absence information for brassicas at a plot level.

We compared the results of the 2005 validation survey with the (2003) data in three ways: (1) contingency table analysis of predicted presence-absence from both years; (2) by fitting a second GLM using the same factors as the (2003) model and comparing model fit with the 2005 data; and (3) by generating the best model using all predictive factors used initially in the (2003) model and removing these by backwards stepwise deletion, then comparing the predicted probability surface of this model with that of the original model graphically.

Results

FIELD SURVEYS

Naturalized *Brassica* populations were distributed widely across the landscape, occurring in 19 of 50 sample plots in 2003 (see map in Heenan, FitzJohn & Dawson 2004). A total of 47 populations were recorded within the 21 plots. Most *Brassica* populations were small; 57% of the populations were comprised of less than five individuals.

The best predictive model from the 2003 survey data retained distance to the nearest coast, seed certification farm

Table 1. Summary of generalized linear model results for 2003 and 2005 *Brassica* surveys in the Canterbury Plains, South Island, New Zealand. *Brassica* presence and absence was recorded in 50 3 × 3 km field survey plots in 2003. An additional 44 plots were surveyed in 2005 for the purpose of model validation. Factors represent distances (km) from survey plots generated from spatial data layers. Further details of these factors and model construction are summarized in the Methods section

Factor	Estimate	AIC	Deviance (explained)	F	P
2003 model + 2003 data					
Coast	-0.027	44.7	31.79 (6.92)	10.0	0.003
Seed certification	-0.072	60.1	38.71 (22.28)	32.2	< 0.001
Town	-0.058	52.5	54.08 (14.68)	21.2	< 0.001
Total	9.648	39.8	31.79		
2003 model + 2005 data					
Coast	0.004	52.5	46.17 (0.34)	0.3	0.59
Seed certification	-0.012	53.8	46.51 (1.67)	1.4	0.24
Town	-0.016	55.8	47.81 (3.65)	3.2	0.83
Total	2.008	54.2	46.17		
2005 model + 2005 data					
Seed store	-0.021	53.6	46.51 (5.05)	4.6	0.038
Total	2.700	50.5	51.56		

and town as predictor variables and explained 52.1% of the variance in *Brassica* presence-absence (Fig. 1, Table 1). Few plots (10%) were misclassified in 2003. The probability of *Brassica* presence was correlated negatively with distance to each of these factors (i.e. the probability of *Brassica* presence decreased as distance increased). The length of road within plots was not a significant factor in predicting *Brassica* population presence. In summary, the distribution of brassicas in this system could be predicted reasonably accurately using a GLM model from the 2003 field survey data.

The 2003 model predicted that the presence of *Brassica* populations between random plots was significantly better than would be expected by chance alone. Comparison of the distribution of probabilities where *Brassica* populations were found with the distribution of probabilities over the entire sampled area showed that we found *Brassica* populations in higher-probability areas (see inset Fig. 2, one-tailed K-S test: $D = 0.26$, $P = 0.001$).

FIELD RESURVEY

Naturalized *Brassica* populations were more widespread during the 2005 field validation survey than in the 2003 survey. Thirty-two of 44 plots (72%) contained naturalized *Brassica* populations. Of 32 plots with naturalized *Brassica* populations, 29 contained *B. rapa* (three contained other *Brassica* species but not *B. rapa*). Similarly, of 142 populations observed in 2005, 125 were *B. rapa*. Of 22 'expected' plots (i.e. those plots surveyed in 2005 that were predicted to contain brassicas from the 2003 model), 18 (82%) had *Brassica* populations, whereas of the 22 'not-expected' plots (i.e. those plots surveyed in 2005 that were not predicted to contain brassicas), 14 (64%) contained brassicas (overall contingency table of 2003 vs. 2005 predictions of *Brassica* presence: $\chi^2_{d.f.} = 1.83$, $P = 0.18$). These results suggest that *Brassica* presence was reliable and could be predicted across years based on 2003 model. However, presence in other areas could not be predicted well, being absent in some years (2003) and present in others (2005).

Using the same model generated from the 2003 data, no predictive factors were significant when applied to 2005

survey data (Table 1). The best model for the 2005 data (Fig. 3a) retained only distance to seed storage facility as a significant factor, and explained only about 10% of the variance in *Brassica* presence-absence (Table 1). The predicted presence of *Brassica* populations at the landscape scale differed between the models generated in different years. A K-S test comparing the 2003 and 2005 distributions of predicted probabilities of *Brassica* presence showed that these were different ($D = 0.63$, $P < 0.0001$). This was driven by several factors: (1) the overall mean probability of *Brassica* presence was much higher in the 2005 model than the 2003 model (Fig. 3); (2) the 2005 model also predicted higher probability of *Brassica* presence in inland areas; and (3) all high probability areas (generally coastal) in the 2003 model were predicted by the 2005 model. The major area of disagreement was where the 2003 model predicted no *Brassica*, but the 2005 model predicted *Brassica* presence (37% of the modelled area). Cases where *Brassica* presence or absence was predicted by both models accounted for 19% and 0% of the model area, respectively. In summary, *Brassica* distribution at the landscape scale could be predicted well using results from a rapid field survey, but not consistently between years.

Discussion

Our results demonstrate that both presence and absence data can be used to construct robust models predicting the landscape scale distribution of naturalized brassicas. Despite this model being well validated by a presence-only botanical survey and making accurate predictions of *Brassica* presence that were reliable across years, presence in other areas could not be predicted well, being absent in some years (2003) and present in others (2005) (Fig. 3). Across years, *Brassica* presence is related most strongly to seed storage facilities and transportation routes (Table 1), suggesting that human-mediated dispersal may be a major pathway for creating naturalized populations. Below we discuss the implications of these findings for distribution models in general, and for understanding weed distribution and the potential of gene-flow risk.

Model validation is an important step in producing robust, predictive models. Although there is a large amount of

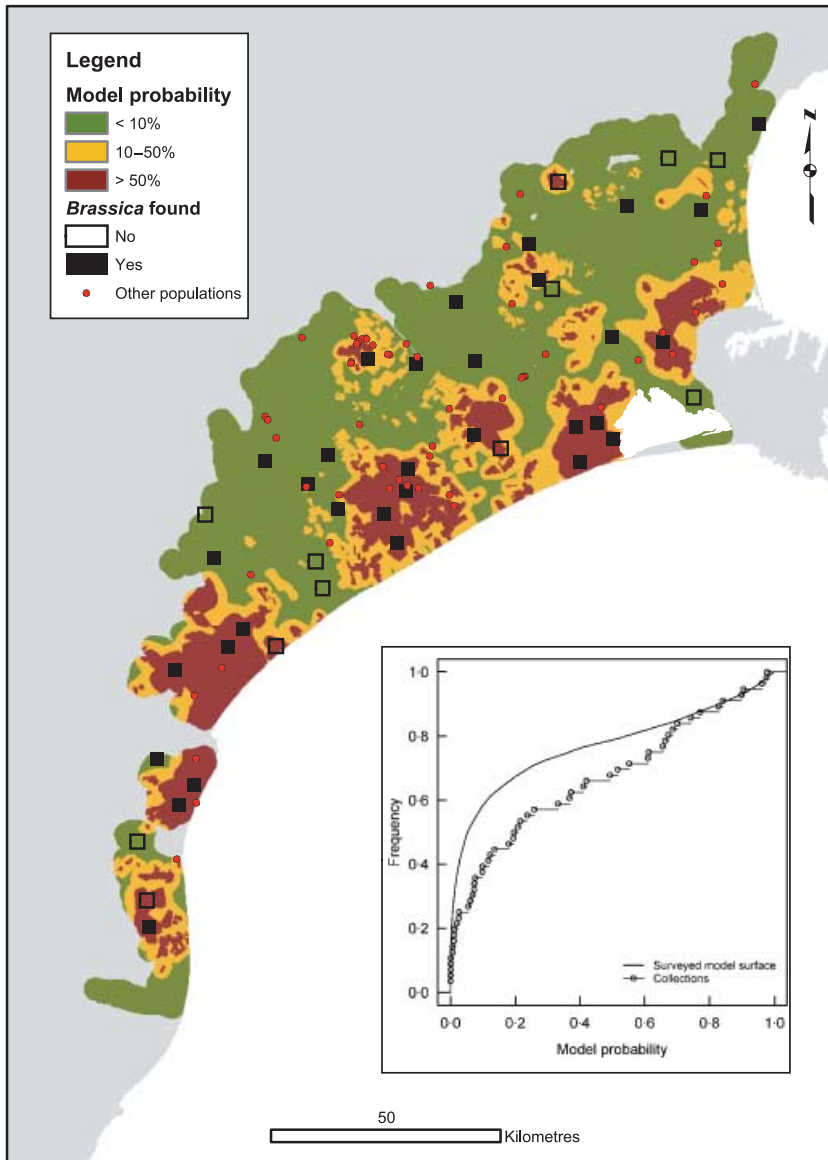


Fig. 2. The expected likelihood of naturalized *Brassica* presence from the original (2003) survey plot model summarized into three probability bins: *Brassica* > 50% probably present (probably present; red fill), 10–50% (perhaps present; yellow fill) or < 10% (probably absent; green fill) on the Canterbury Plains, South Island, New Zealand. Squares represent the 2005 validation field sampling plots in which brassicas were recorded as either present (filled) or absent (open). Inset panel: the cumulative probability distribution for *Brassica* presence in the original 2003 survey model prediction surface (solid line) and the driven track recorded between survey plots (stepwise function).

literature on predicting the landscape-scale distribution or presence of both plant and animal species (e.g. Austin 2002; Thuiller *et al.* 2003; Rushton, Ormerod & Kerby 2004), few of these models account for variation in distribution through time or validate models with independent data. For example, many ecological studies using CART analyses usually partition a large data set into one portion used to create the model, and other portions to validate the model (e.g. Bourg *et al.* 2005). Our results demonstrate that between-year differences in both species distribution and the factors that predict presence or absence can limit the ability of models to be applied across years, and hence their generality. More generally, these models may be used to predict 'hot-spots' where populations are likely to persist but not for detecting rare events (e.g. identifying the distribution and contact zones between crop species and conspecific relatives). Hence, these models may have limited utility for predicting the risk of gene-flow. This raises important issues not only for predicting the distribution of species

through time, but also for risk analysis of weed distribution or gene-flow at the landscape-scale.

A related issue is how robust models are to changes in species prevalence. For example, the ideal data for modelling binomial data is a equal mix of presence–absence data (Harrell 2001); as prevalence (and hence observed presence) increases, statistical power declines (see also Allouche, Tsoar & Kadmon 2006). Here, our 2003 field survey data was an excellent data set for constructing GLM models. However, increasing prevalence and hence proportion of plots containing brassicas in the 2005 resurvey decreased the statistical power of the model and its ability to predict distributions. This is not necessarily a problem for risk assessment or rapid assessment of distribution *per se*, but highlights a potential limitation of this modelling approach to predicting distributions.

Our presence–absence models agreed on *Brassica* presence for large populations, but were not concordant for absences (Fig. 3). Put another way, despite constructing robust models

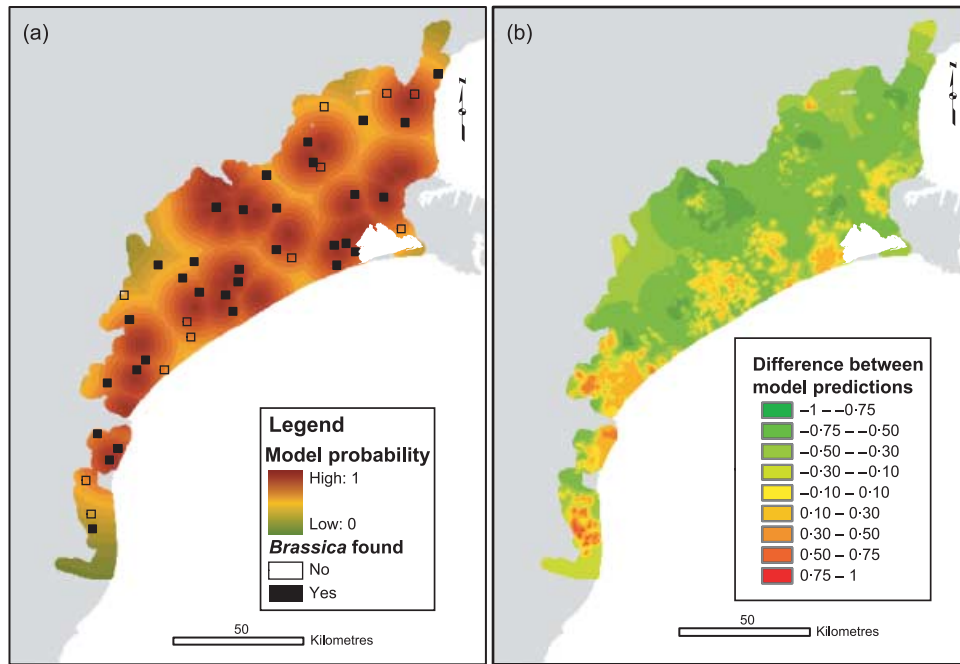


Fig. 3. (a) Left panel: the predicted probability of naturalized *Brassica* presence (red) or absence (green) from the 2005 validation field survey plot model (left panel). The scale is identical to that used in Fig. 1 for the 2003 model output. Squares represent the 2005 validation field sampling plots in which brassicas were recorded as either present (filled) or absent (open). (b) Right panel: absolute difference between predicted probability of *Brassica* presence between the (2003) and 2005 models generated from field survey plot data. Green represents greater predicted presence by the 2005 model, whereas red represents greater predicted presence by the (2003) model. See Methods section for further details of model development.

to predict the distribution of brassicas, we could not identify areas that consistently lacked *Brassica* populations between years. A related, unresolved issue is how much system-specific information is needed to assess risk? For example, the lack of New Zealand-specific information on the distribution and abundance of brassicas motivated this study and justified using a rapid field assessment in the first instance. Thus, in the absence of system-specific information, we suggest that the precautionary principle must be used for managing naturalized population, i.e. that wild relatives or feral populations may be present anywhere in the study system. This result supports an approach to gene-flow risk which anticipates that gene flow will happen in agricultural systems, and puts effort into managing wild relatives or naturalized populations rather than allocating resources to understand the magnitude of these risks.

An alternative interpretation of our results for variability in the ability to predict *Brassica* distribution among years is that the abundance and distribution of *Brassica* are variable through time and thus difficult to predict between years (see also Wolfenbarger & Phifer 2000; Légère 2005). Does the distribution and abundance of *Brassica* populations change so much year-to-year or from stochastic factors that no matter what sampling or modelling approach is used, predictions are difficult? If both models (2003 and 2005) are an accurate representation of the actual distribution of *Brassica* populations, then the distribution changed markedly between 2003 and 2005; *Brassica* populations were much more widespread

in 2005 than in 2003 (Figs 1 and 2). Because naturalized *Brassica* populations have been present in the study system for decades we do not believe that this represents an invading front or range expansion, but rather that seed bank dynamics and climatic factors may control this year-to-year variation in abundance or distribution; such dynamics cannot be captured by single season survey data. Similarly, brassicas and other weedy or fast-growing species may have larger interannual variation in abundance or seed production than perennial species (e.g. Franco & Silvertown 1996; Grime 2001). Alternatively, management regime might overwhelm any climatic or environmental effects, and changes in management regimes through time should play a bigger role than biological or environmental limits, at least in these arable systems (e.g. Freckleton & Watkinson 2002). Surveys could not be made over fewer years for our current study (only two time steps), but more generally, one would expect to detect variation in gene flow and population sizes more quickly with annual or short-lived species than with perennial or long-lived species. The implications of this are that gene-flow is detected more quickly for short-lived species; however, the general approach of rapidly assessing the landscape-scale distribution of species can be applied easily to longer-lived species. These issues can be resolved either by using mechanistic models that incorporate population processes and response to interannual factors (e.g. Claessen *et al.* 2005) or by using longer-term data; both approaches are less cost-effective and require more detailed data to generate parameters (in the case of mechanistic models)

or repeated measures to capture interannual variation in abundance (long-term monitoring).

Although we cannot distinguish *de novo* *Brassica* populations derived from fresh seed spillage from either crop escapes (i.e. from fields to verges) or from persistent seed bank populations directly, three lines of evidence suggest that fresh seed spillage is unlikely to have caused the differences in *Brassica* abundance observed between 2003 and 2005. First, the total area of *Brassica* seed production between 2002 and 2005 in the study system varied by < 20% between years (seed certification statistics 2005–06; Agriquality Limited, Lincoln, New Zealand). Secondly, persistent feral populations are known to persist in other areas of New Zealand, even after more than a decade has passed since *Brassica* cultivation in the area (E. Johnson, Agriquality New Zealand, personal communication). Thirdly, a seed bank study confirmed that where *Brassica* populations were observed in the field a seed bank was present, but in plots in which we did not observe brassicas no seed bank was observed, with a single exception (D. Peltzer, unpublished data). This evidence suggests, at least over the course of this study, that most *Brassica* populations observed are probably persistent rather than *de novo*. Furthermore, this persistence through time has important implications for gene flow risk and field evaluations of brassicas or other species with persistent propagules.

Distribution models may be most appropriate for species having distribution limits in the study system. Our study system, the Canterbury Plains of New Zealand, is a relatively uniform environment and may represent suitable habitat for naturalized *Brassica* species through most, if not all, the system. Similarly, the large (3 × 3 km) survey plots used are probably sufficient in size to contain some suitable habitat for naturalized brassicas based on the wide distribution of seed and forage crop cultivation (seed certification statistics 2005–06; Agriquality Limited, Lincoln, New Zealand) and botanical expert opinion and experience (P. Heenan, personal communication). Although some suitable habitat is available within plot, habitat quality and the abundance of *Brassica* populations may vary widely at the landscape scale. Previous studies also highlight variation in *Brassica* presence among habitats; for example, *B. rapa* is common on river banks (except in Scotland, Wilkinson *et al.* 2003a) but infrequent in ruderal habits elsewhere (Gwynne & Murray 1985; Rich 1991). On the other hand, limits for wild relatives may be greater than that of crops, so in areas relevant for crop production (and in the absence of a current invasion by a relative) weed distribution limits will not be reached. In contrast, habitat selection models or distributional models may depend upon environmental or other factors limiting the range of species in the system (e.g. Austin 2002; Bourg *et al.* 2005). These results imply that models calibrated for one species or system may not be applied safely to other systems or naturalized populations of a species.

These results have important implications both for developing models of species distributions and for predicting gene-flow risk at the landscape scale. Despite constructing a model that was validated subsequently from presence-only data in one

year, it did not predict *Brassica* distribution in a subsequent year; this suggests that between-year or stochastic factors operate strongly and are a major consideration for risk assessment. Models used to predict species distributions in general need to account for these between-year effects either by incorporating mechanistic processes or long-term data.

Acknowledgements

We thank J. Overton and two anonymous referees for their thoughtful comments on the manuscript, E. Johnson of Agriquality New Zealand and A. Stewart of PGG Wrightson seeds for helpful discussions, and W.G. Lee, L. Newstrom and P. Heenan for collegial support. This research was supported by the New Zealand Foundation for Science Research and Technology contract CO9 X0601.

References

- Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.
- Andow, D.A. & Zwahlen, C. (2006) Assessing environmental risks of transgenic plants. *Ecology Letters*, **9**, 196–214.
- Armstrong, T.T., FitzJohn, R.G., Newstrom, L.E., Wilton, A.D. & Lee, W.G. (2005) Transgene escape: what potential for crop–wild hybridization? *Molecular Ecology*, **14**, 2111–2132.
- Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.
- Beckie, H.J., Warwick, S.I., Nair, H. & Séguin-Swartz, G. (2001) Gene flow in commercial fields of herbicide-resistant canola (*Brassica napus*). *Ecological Applications*, **13**, 1276–1294.
- Bourg, N.A., McShea, W.J. & Gill, D.E. (2005) Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology*, **86**, 2793–2804.
- Claessen, D., Gilligan, C.A., Peter, J.W. & van den Bosch, F. (2005) Which traits promote persistence of feral GM crops? Part 1: implications of environmental stochasticity. *Oikos*, **110**, 20–29.
- Drake, J.M., Randin, C. & Guisan, A. (2006) Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, **43**, 424–432.
- Elliott, L.J., Mason, D.C., Wilkinson, M.J., Allainguillaume, J., Norris, C., Alexander, M. & Welters, R. (2004) The role of satellite image-processing for national-scale estimates of gene flow from genetically modified crops: rapeseed in the UK as a model. *Journal of Applied Ecology*, **41**, 1174–1184.
- Ellstrand, N.C., Prentice, H.C. & Hancock, J.F. (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539–563.
- FitzJohn, R.G., Armstrong, T.T., Newstrom, L.E., Wilton, A.D. & Cochrane, M. (2007) Hybridisation within *Brassica* and allied genera: evaluation of potential for transgene escape. *Euphytica*, **158**, 209–230.
- Flannery, M., Meade, C. & Mullins, E. (2005) Employing a composite gene-flow index to numerically quantify a crop's potential for gene flow: an Irish perspective. *Environmental Biosafety Research*, **4**, 29–44.
- Franco, M. & Silvertown, J. (1996) Life history variation in plants: an exploration of the fast–slow continuum hypothesis. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **351**, 1341–1348.
- Freckleton, R.P. & Watkinson, A.R. (2002) Are weed population dynamics chaotic? *Journal of Applied Ecology*, **39**, 699–707.
- Godfree, R.C., Young, A.G., Lonsdale, W.M. & Burdon, J.J. (2004) Ecological risk assessment of transgenic pasture plants: a community gradient modelling approach. *Ecology Letters*, **7**, 1077–1089.
- Gray, A.J. (2004) Ecology and government politics: the GM crop debate. *Journal of Applied Ecology*, **41**, 1–10.
- Grime, J.P. (2001) *Plant Strategies, Vegetation Processes, and Ecosystem Properties*, 2nd edn. Wiley, Chichester.
- Gwynne, D.C. & Murray, R.B. (1985) *Weed Biology and Control in Agriculture and Horticulture*. Batsford Press, London.
- Harrell, F.E. (2001) *Regression Modeling Strategies*. Springer, Berlin.
- Heenan, P.B., Dawson, M.I., FitzJohn, R.G. & Stewart, A.V. (2007) Experimental hybridisation of *Brassica* species in New Zealand. *New Zealand Journal of Botany*, **45**, 53–66.
- Heenan, P.B., FitzJohn, R.G. & Dawson, M.I. (2004) Diversity of *Brassica*

- (Brassicaceae) species naturalised in Canterbury, New Zealand. *New Zealand Journal of Botany*, **42**, 815–832.
- Leathwick, J., Wilson, G., Rutledge, D., Wardle, P., Morgan, F., Johnston, K., McLeod, M. & Kirkpatrick, R. (2003) *Land Environments of New Zealand*. David Bateman Ltd, Auckland.
- Légère, A. (2005) Risks and consequences of gene flow from herbicide-resistant crops: canola (*Brassica napus* L.) as a case study. *Pest Management Science*, **61**, 292–300.
- Levine, J.M., Montserrat, V., D'Antonio, C.M., Dukes, J.S., Grigulis, K. & Lavorel, S. (2003) Mechanisms underlying the impacts of exotic plant invasions. *Proceedings of the Royal Society, Series B, Biological Sciences*, **270**, 775–781.
- Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, J., Clout, M. & Bazzaz, F.A. (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689–710.
- Pessel, E.D., Lecomte, J., Emeriau, V., Krouti, M., Messean, A. & Gouyon, P.H. (2001) Persistence of oilseed rape (*Brassica napus* L.) outside of cultivated fields. *Theoretical and Applied Genetics*, **102**, 841–845.
- Pilson, D. & Prendeville, H.R. (2004) Ecological effects of transgenic crops and the escape of transgenes into wild populations. *Annual Review of Ecology, Evolution and Systematics*, **35**, 149–174.
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R 2.3.1. R Development Core Team, Royal Foundation for Statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org>.
- Raybould, A.F. & Gray, A.J. (1993) Genetically modified crops and hybridization with wild relatives: a UK perspective. *Journal of Applied Ecology*, **30**, 199–219.
- Rentch, J.S., Fortney, R.H., Stephenson, S.L., Adams, H.S., Grafton, W.N. & Anderson, J.T. (2005) Vegetation–site relationships of roadside plant communities in West Virginia, USA. *Journal of Applied Ecology*, **42**, 129–138.
- Rich, R.G.C. (1991) *Crucifers of Great Britain and Ireland*. Botanical Society of the British Isles, London.
- Rieger, M.A., Lamond, M., Preston, C., Powles, S.B. & Roush, R.T. (2002) Pollen-mediated movement of herbicide resistance between commercial canola fields. *Science*, **296**, 2386–2388.
- Rieger, M.A., Preston, C. & Powles, S.B. (1999) Risks of gene flow from transgenic herbicide-resistant canola (*Brassica napus*) to weedy relatives in southern Australian cropping systems. *Australian Journal of Agricultural Research*, **50**, 115–128.
- Royal Commission on Genetic Modification (2002) *Report of the Royal Commission on Genetic Modification*. Wellington, New Zealand.
- Rushton, S.P., Ormerod, S.J. & Kerby, G. (2004) New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**, 193–200.
- Scheffler, J.A. & Dale, P.J. (1994) Opportunities for gene transfer from transgenic oilseed rape (*Brassica napus*) to related species. *Transgenic Research*, **3**, 263–278.
- Snow, A.A., Andow, D.A., Gepts, P., Hallerman, E.M., Power, A., Tiedje, J.M. & Wolfenbarger, L.L. (2005) Genetically engineered organisms and the environment: current status and recommendations. *Ecological Applications*, **15**, 377–404.
- Thuiller, W., Aratijo, M.B. & Lavorel, S. (2003) Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.
- Traxler, G. (2006) The GMO experience in North and South America. *International Journal of Technology and Globalisation*, **2**, 46–64.
- de Vries, F.T., van der Meijden, R. & Brandenburg, W.A. (1992) A study of the real chances for spontaneous gene flow from cultivated plants to the wild flora of the Netherlands. *Goteria Supplement*, **1**, 1–100.
- Wilkinson, M.J., Elliott, L.J., Allainguillaume, J., Shaw, M.W., Norris, C., Welters, R., Alexander, M., Sweet, J. & Mason, D.C. (2003a) Hybridization between *Brassica napus* and *B. rapa* on a national scale in the United Kingdom. *Science*, **302**, 457–459.
- Wilkinson, M.J., Sweet, J. & Poppy, G.M. (2003b) Risk assessment of GM plants: avoiding gridlock? *Trends in Plant Science*, **8**, 208–212.
- Wolfenbarger, L.L. & Phifer, P.R. (2000) The ecological risks and benefits of genetically engineered plants. *Science*, **290**, 2088–2093.

Received 29 March 2007; accepted 16 August 2007

Handling Editor: Robert Freckleton