New distributional modelling approaches for gap analysis

A. Townsend Peterson and Daniel A. Kluza

Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, KS 66045, USA

Abstract

Synthetic products based on biodiversity information such as gap analysis depend critically on accurate models of species' geographic distributions that simultaneously minimize error in both overprediction and omission. We compared current gap methodologies, as exemplified by the distributional models used in the Maine Gap Analysis project, with an alternative approach, the geographic projections of ecological niche models developed using the Genetic Algorithm for Rule-Set Prediction (GARP). Point-occurrence data were used to develop GARP models based on the same environmental data layers as were used in the gap project, and independent occurrence data used to test both methods. Gap models performed better in avoiding omission error, but GARP better avoided errors of overprediction. Advantages of the point-based approach, and strategies for its incorporation into current gap efforts are discussed.

INTRODUCTION

Effective biodiversity conservation and natural resource management require accurate information on distributions of species, combined with information on land use and land tenancy (Scott, Tear & Davis, 1996). An early approach to this challenge led to the development of the technique of gap analysis (Scott *et al.*, 1993), which evolved into the US National Gap Analysis Program (Scott *et al.*, 1996). This consists of a federation of state and regional efforts to interpret species' distributions in the context of habitat associations towards more effective management of natural resources.

Methodologies for predicting species' geographic distributions in gap analysis have generally been as follows (Csuti, 1996), although newer efforts have done some experimentation with other methodologies (Scott, Heglund & Morrison, 2002). (1) Published range maps, museum specimen records, and other documented occurrences are used to determine general geographic range, placing species in geographic subunits (e.g. counties). (2) Habitat use and needs of species are evaluated based on the scientific literature, identifying suites of landcover types as suitable for particular species. The intersection of species' presence in broad units (1) and fine-scale habitat requirements (2), modified with additional factors (e.g. elevational limits, proximity to bodies of water) and expert review, yields a predicted distribution for a species.

All correspondence to: A. Townsend Peterson. Tel: 785-864-3926; Fax: 785-864-5335, E-mail: town@ku.edu.

Alternative approaches to modelling species' distributions have been offered (Nix, 1986; Austin, Nicholls & Margules, 1990; Walker & Cocks, 1991; Stockwell, 1999; Stockwell & Peters, 1999), based on statistical modelling of ecological niches from point-occurrence data (hereafter 'point-based approaches'). Curiously, these alternative methods have appeared principally in the scientific/ecological literature, and have received little attention in gap projects (Scott *et al.*, 1996). More surprising, however, is that few head-to-head comparisons have assessed the relative predictive abilities of gap models versus those of alternative methodologies (Krohn, 1996). This 'gap' in the scientific literature is the focus of this paper: we compare the accuracy of gap models and models built using a point-based approach based on geographic predictions of 30 forest passerine bird species in the state of Maine.

ECOLOGICAL NICHE MODELLING AND DISTRIBUTIONAL PREDICTION

Grinnell (1917, 1924) presented an early concept of an ecological niche, defined as the ranges of ecological conditions within which a species is able to maintain populations. This concept contrasts with the more customary Hutchinsonian *N*-dimensional niche concept (Hutchinson, 1957), which focuses more on the role of a species within a local community. These concepts can be refined by distinguishing between the fundamental and realized niches (Hutchinson, 1957), the latter taking into account the effects of interactions among species in reducing habitable areas. Although modelling the fun-

damental niche based on actual distributional data may seem counterintuitive, the broad spectrum of community backgrounds present across a species' geographic distribution allows some degree of insight into this more basic picture of a species' ecological requirements.

Modelling species' ecological niches offers important insights into their distributional ecology. For the vast majority of species, detailed physiological studies of tolerances and requirements are unavailable, making necessary inference of ecological requirements from simpler data. The data that are almost universally available are primary-point occurrences (Peterson, Navarro-Siguenza & Benitez-Diaz, 1998; Peterson, Stockwell & Kluza, 2002*d*): simple records of a particular species at a point in space and time. These data, interpreted in the context of ecological variation across landscapes, can provide detailed hypotheses about geographic distributions of most species.

The *Genetic Algorithm for Rule-Set Prediction* (GARP) includes several algorithms in an iterative, artificial-intelligence-based approach (Stockwell & Noble, 1992; Stockwell, 1999; Stockwell & Peters, 1999). Individual algorithms (e.g. BIOCLIM, logistic regression) are used to produce component 'rules' in a broader rule-set, so portions of the species' distribution can be determined as inside or outside its niche based on different rules. Hence, GARP is a superset of other approaches, and should have greater predictive ability than any one of them. Initial testing of GARP has indicated excellent predictive ability (Peterson & Cohoon, 1999; Peterson, Soberon & Sanchez-Cordero, 1999; Peterson & Vieglais, 2001; Anderson, Laverde & Peterson, 2002*a,b*; Feria & Peterson, 2002; Peterson, Ball & Cahoon, 2002*a*; Peterson *et al.*, 2002*c*; Peterson *et al.*, 2002*d*; Stockwell & Peterson, 2002*a,b*; Anderson, Law & Peterson, in press; Peterson & Robins, in press; A. T. Peterson, R. Scachetti-Pereira & D. A. Kluza, unpubl. data) and insensitivity to BIOCLIM's problems with environmental data density (Peterson & Cohoon, 1999; Stockwell & Peterson, 2002*a,b*).

Within GARP, occurrence points are divided randomly and evenly into training and intrinsic testing data sets. GARP works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection: a method is chosen from a set of possibilities (e.g. logistic regression, bioclimatic rules) and applied to the training data, and a rule is developed or evolved. Predictive accuracy is then evaluated based on 1250 points resampled from the test data and 1250 points sampled randomly from the study region as a whole, as summarized in a rule-significance measure (Stockwell & Noble, 1992; Stockwell, 1999; Stockwell & Peters, 1999). Rules may evolve by a number of means that mimic DNA evolution: point mutations, deletions, crossing over, etc. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 1000 iterations or until convergence.

Two types of error potentially enter into such niche modelling and geographic prediction efforts. *Omission* of areas actually inhabited is a failing of the modelling effort to extend to all ecological conditions under which the species is able to maintain populations. *Commission* error is that of including areas actually uninhabited; this error component includes real error, in which combinations of ecological conditions not actually within the species' niche are included, as well as a historical component. The historical component results from species' absences owing to limited colonization ability, speciation and local extinction, and represents a real feature of species' distributional ecology—not all habitable area is inhabited (Peterson *et al.*, 1999; Peterson & Vieglais, 2001). Most challenging is that measurement of commission error is complicated seriously by the rampant undersampling of biodiversity: many sites apparently not inhabited by species are in actuality inhabited but the species has not been detected there ('apparent error') (Nichols *et al.*, 1998).

METHODS

Through consultation with gap coordinators from around the United States, Maine was chosen as a first test-bed for the gap–GARP comparison. Maine offered the advantages of having (1) a recently completed gap analysis project; (2) appropriate geospatial information available; (3) sufficiently precise distributional information available for birds in the region from time periods that coincided with the provenance of the geospatial information. The overall plan of analysis was that the same environmental data coverages and a subset of the occurrence information were used in GARP analyses as had been used in the Maine Gap Analysis Program; gap and GARP models were evaluated based on the same test data sets.

GARP models were developed based principally on topographic and climatic information, and hence were fairly crudely resolved spatially. Geospatial/environmental data available for Maine (supplied by W. Krohn and colleagues) included the following environmental dimensions: elevation; slope; aspect; annual mean precipitation; average maximum and absolute maximum temperatures for winter, spring, summer, fall and the entire year; and average minimum and absolute minimum temperatures for winter, spring, summer, fall and the entire year. Hence, 26 coverages were used for the ecological modelling efforts. To make Internet-mediated analyses feasible, these data were generalized to a $0.02 \times 0.02°$ pixel resolution (about 1.9×1.9 km in linear distance units) exported as ASCII raster coverages, and deposited on the GARP server at the San Diego Supercomputer Center.[1]

Thirty species of birds were chosen for analysis, under the criteria that they be generally associated with forest

---

[1] http://biodiversity.sdsc.edu/cgi-bin/findir.cgi?screen.cgi. Note, however, that a new version for use on PC desktop computers is now available for public download at http://beta.lifemapper.org/desktopgarp/.

(a habitat type well-sampled at fine scales), and that more than 20 unique latitude–longitude combinations documenting presence be available for analysis. Distributional data were drawn from stop-level data (sample points approximately every 1 km) from the 1990 results of the US Breeding Bird Survey.[2] For each species, we chose 20 occurrence records at random as a testing data set for presences. We also used all sample points from which the species had *not* been recorded as a testing data set for absences ('test data'). (This broader test absence data set is necessary because absences are more difficult to demonstrate than presences, with some dilution of absences by undetected presences (Anderson *et al.*, in press).) The remaining available occurrence points, which ranged from 8 to 541 per species, were used for developing GARP models ('training data').

Training data were analyzed using the Maine coverages on the GARP facility to develop ecological niche models; resulting geographic predictions were exported as ASCII raster grid files. In ArcView (version 3.1), raster grids were imported, and converted to Arc grids. Crude GARP predictions were downscaled to match the resolution of gap models as follows: (1) pixel resolution was artificially increased to $0.001 \times 0.001°$ (90×90 m); (2) a 90 m resolution land-use/land-cover map (also provided by W. Krohn and colleagues) was queried for vegetation types known to be used by each species (DeGraaf *et al.*, 1992) (D. A. Kluza, unpubl. data); (3) the crude prediction was cut so that only those areas predicted present by GARP *and* holding the appropriate land-use/land-cover type were included in the predicted area. This step was appropriate, given that initial predictions based on broad climatic parameters did not include details of local land use patterns, and allowed inclusion of fine-scale phenomena not included in the original modelling effort.

To test predictive efficiency of models, the independent test data points were overlain on the predictions for each species. Predictive ability of the final Maine gap models (Boone & Krohn, 1998; Krohn *et al.*, 1998) was tested using the same test data points. It should be pointed out that whereas measurement of omission error rates is straightforward (= proportion of test presence points incorrectly predicted as absent), commission error calculations are complicated by undetected presences of species (Nichols *et al.*, 1998). However, assuming no spatial bias in detection probabilities for species, these false commission errors should not *bias* results in favour of one method or another.

The pairs of models for each species (GARP and gap) were evaluated using the approaches summarized by Fielding & Bell (1997). These approaches are based generally on a 2×2 'confusion matrix' (Table 1), where $N = a + b + c + d$, c represents omission error, and b represents commission error. These quantities are used to calculate three complementary measures of predictive efficiency (Fielding & Bell, 1997): (1) false positive

Table 1. Confusion matrix summarizing model performance and providing variable names used in summary statistics (Fielding & Bell, 1997).

	Actual	
Predicted	Present	Absent
Present	a	b
Absent	c	d

(commission) rate = $b / (b + d)$; (2) false negative (omission) rate = $c / (a + c)$; and (3) $\kappa = ((a + d) - (((a + c)(a + b) + (b + d)(c + d)) / N)) / (N - (((a + c)(a + b) + (b + d)(c + d)) / N))$. (1) and (2) measure the two components of error described above, whereas (3) offers an overall measure of predictive efficiency.

To deal with 'near misses' caused by imprecision of georeferencing, these measures were further modified using a city-block weighting scheme (Fielding & Bell, 1997), in which c is downweighted by a factor $w_x = 1 - 1/2x$, in which x represents the city-block distance by which a point fell outside of the predicted area; the $w_x > 0.85$ threshold for neglecting near misses suggested by Fielding & Bell (1997) was employed. Relative performance of the two models was compared statistically using the McNemar test described by Fielding & Bell (1997), in which test points classified correctly by both models as present (e) or absent (h), or incorrectly by one method (f) or the other (g), are summarized with the χ-square approximation $(f - g)^2 / (f + g)$, with one degree of freedom. Details of the derivation of each measure are offered in Fielding & Bell (1997).
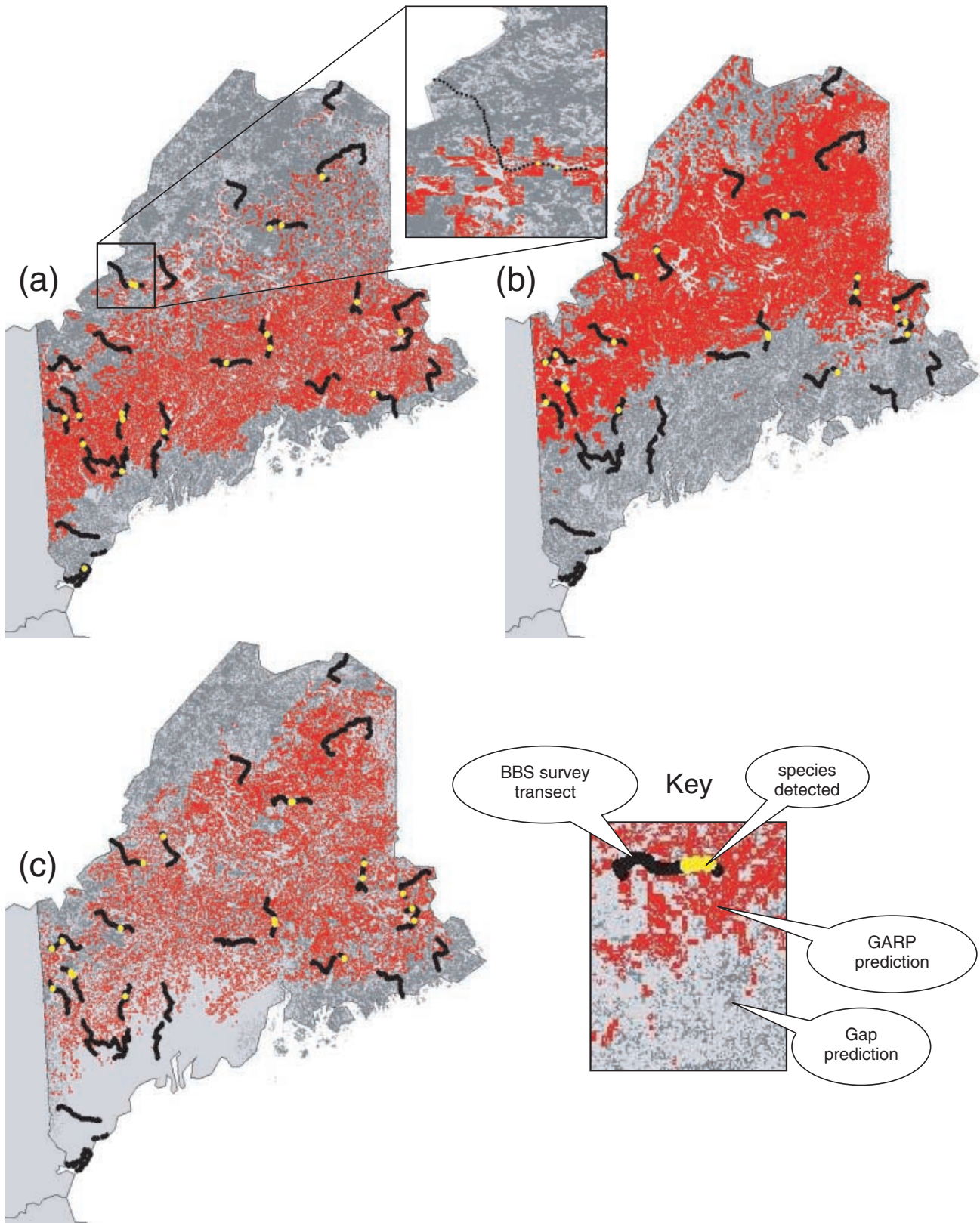
RESULTS

The two approaches to predicting geographic distributions produced results that contrasted sharply (Fig. 1). In general, gap models identified habitats for species, but predictions extended across the entire state; GARP models, on the other hand, homed in on particular regions of the state, and avoided predicting species' presence over the entire state. The initially crude scale of the GARP predictions was refined by use of land-use/land-cover information, providing both fine resolution and regional restriction for predictions.

GARP models were relatively insensitive to training sample size (Fig. 2). Neither false-positive (commission) rates nor false-negative (omission) rates were significantly related to numbers of points used to develop models. This result suggests that the sample sizes available to us for model building in this study did not approach lower limits of inferential power of the GARP algorithm.

Omission error was lower in gap models (Table 1). In 24 of 30 species, gap models had better success in predicting the 20 test presence points, whereas GARP models were better in only three species; the two methods tied for three species. Hence, on the omission criterion, gap models performed better than GARP models. The situation regarding commission error, however, was different. For this error component, GARP models

Fig. 1. Four example species showing results of gap (dark gray) and GARP (red) modelling approaches, with test presence points (in yellow) and all BBS sample points (black) overlaid. Shown are (a) ovenbird, (b) black-throated blue warbler, (c) Blackburnian warbler. A detail of one BBS route is shown for ovenbird. Although patterns of overlap among the two predictions are difficult to display owing to the different levels of detail that each includes, these maps serve to show the differences in regional extent of the predictions, and the coincidence of those differences with patterns of presence and absence of species.
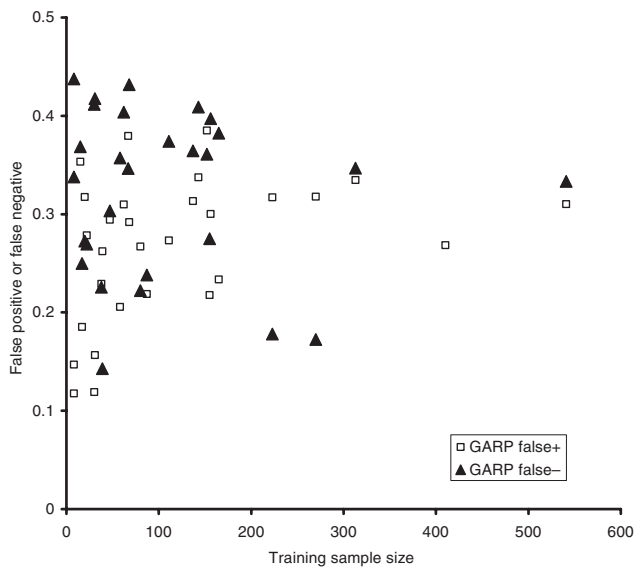
Fig. 2. False-positive and false-negative rates of GARP models as a function of sample size.

outperformed gap models in 28 of 30 species, with substantial differences in commission error (Table 1).

McNemar's tests provided overall tests of model performance, based both on presence and absence test data sets. In all, 28 of 30 species were significantly better predicted by GARP models than by gap models (all $P < 10^{-7}$). The remaining two species, *Regulus satrapa* and *Seiurus noveboracensis*, were predicted significantly better by gap models ($10^{-3} > P > 10^{-5}$). κ scores averaged significantly higher in GARP models than in gap models (0.030 versus 0.009, paired *t*-tests, $P < 10^{-7}$). Both of these tests take into account the overall area predicted present, and examine models' predictive ability above and beyond that expectation.

DISCUSSION

The two approaches produced very distinct types of distributional models. Gap models appear focused on habitat associations, but fail to identify regional patterns of range restriction. GARP models identified regional limitations and managed habitat associations via the fine-grained land-use/land-cover map queries. The fine-grained but geographically extensive gap models contrasted sharply with the less finely resolved, but geographically specific GARP models. This difference in area predicted present (greater in gap, lower in GARP) coincides with differences in error components: gap models omitted fewer points but included many extra areas apparently not inhabited by species. This difference in overall performance, as reflected in the McNemar's test and κ statistic results, indicates that GARP models have overall much greater predictive power than gap models.

An important consideration is that avoiding omission error is quite easy: simply predicting the entire region as 'present' reduces omission error to zero (Fielding & Bell,

1997). For this reason, arguments that accepting higher levels of commission error so as not to omit occurrences lead to an unacceptable conclusion – better just to predict the entire map present for all species – one clearly not desirable for useful products based on distributional models. The true challenge is that of minimizing omission error *and* commission error simultaneously, which removes the easy alternative of predicting ubiquity, and instead involves predicting the smallest possible area without omitting independent test presence points. In this sense – reducing apparent commission error as much as possible without omitting potential habitats for species – GARP models outperformed gap models, reducing commission error substantially, at a relatively small cost of increased omission.

The uniform test presence data sets (20 points) offer a view of the sensitivity of GARP approaches to training sample sizes (Peterson, 2001; Peterson *et al.*, 2002a; Peterson *et al.*, 2002d; Stockwell & Peterson, 2002b). Sensitivity of the method to sample size was minimal: indeed, highly statistically significant models were developed based on as few as eight training points. No significant negative relationship was detected between sample size and either false-positive or false-negative error rates, suggesting that minimum training sample size thresholds for accurate GARP modelling were not approached in this study. Parallel results have been obtained in a more formal test of GARP's sensitivity to sample size in predicting distributions of Mexican birds (Stockwell & Peterson, 2002b).

The GARP methodology is readily incorporated into existing gap efforts. GARP models provide an effective solution to the challenge of modelling species' distributions that forms a major component of gap programmes (Scott *et al.*, 1996). The vertebrate habitat modelling efforts that typify gap applications can identify land-use/land-cover associations for species, and can be used to cut GARP outputs to reduce scales to fine levels. Indeed, GARP models are easily developed on a variety of geographic scales, and can be developed on scales more broad than single states; this method thus offers a clear solution to edge mapping and multistate efforts in regional gap applications. GARP models, given their reduced commission error rates, offer a clear solution to mapping species richness, which has been complicated with gap models (J. A. Hepinstall, unpubl. data). In this way, GARP methodologies can be incorporated directly into ongoing Gap studies, and can greatly improve the predictive accuracy of species' distributional models in gap analyses.

A further feature of GARP modelling is its time-inexpensiveness relative to gap efforts (Scott *et al.*, 1996). After a relatively brief period of development of predictor environmental coverages (3 weeks), modelling 30 species' distributions took one person less than 8 hours. Indeed, with batch implementations of GARP now available, entire faunas of hundreds of species can be modelled in the course of a few days. This time efficiency, which is remarkable in comparisons with the person-years of effort expended in building gap models

Table 2. Model performance for 30 forest bird species for gap and GARP models, including sample size (n); for each model, correct prediction of 20 independent test points at radii of 0, 50, 150 and 250 m, and false-positive, false-negative, and κ statistics (Fielding & Bell, 1997); and a McNemar's χ^2 statistic and associated P-value (Fielding & Bell, 1997). Asterisks indicate the two of 30 species in which gap models outperformed GARP models on the McNemar's test.

Species	A n	Gap 0	Gap 50	Gap 150	Gap 250	Gap B	Gap C	Gap D	GARP A 0	GARP 50	GARP 150	GARP 250	GARP B	GARP C	GARP D	Gap False+	Gap False−	Gap κ	GARP False+	GARP False−	GARP κ	McNemar χ^2	McNemar P
Catharus fuscescens	313	1	2	5	5	679	19	316	8	14	15	18	333	12	662	0.68	1.19	0.014	0.335	0.347	0.022	243.9	5.68E-55
Catharus guttatus	152	16	18	20	20	963	4	193	9	15	18	19	445	11	711	0.83	0.35	−0.009	0.385	0.361	0.015	386.8	4.02E-86
Catharus swainsoni	31	18	20	20	20	521	2	756	5	11	14	18	200	15	1077	0.41	0.33	0.025	0.157	0.417	0.034	243.2	7.73E-55
Contopus virens	62	12	16	20	20	794	8	452	8	9	16	18	386	12	860	0.64	0.37	0.000	0.310	0.404	0.019	261.6	7.83E-59
Corvus corax	80	19	20	20	20	1152	1	76	7	14	16	16	328	13	900	0.94	0.33	−0.013	0.267	0.222	0.027	766.7	9.5E-169
Dendroica caerulescens	20	15	17	19	20	588	5	700	16	16	16	16	409	4	879	0.46	0.36	0.014	0.318	0.273	0.041	59.1	1.48E-14
Dendroica coronata	137	8	13	17	19	547	12	624	8	12	17	18	367	12	804	0.47	0.39	0.007	0.313	0.364	0.021	76.8	1.91E-18
Dendroica fusca	58	11	13	17	20	524	9	726	3	5	13	15	257	17	993	0.42	0.41	0.012	0.206	0.357	0.015	150.1	1.67E-34
Dendroica pinus	8	7	10	13	19	284	13	1016	6	8	16	18	153	14	1147	0.22	0.45	0.028	0.118	0.437	0.056	57.5	3.41E-14
Dendroica virens	68	12	16	18	19	616	8	624	9	11	19	20	362	11	878	0.50	0.35	0.009	0.292	0.432	0.023	129.9	4.31E-30
Empidonax minimus	111	11	13	17	17	588	9	609	12	13	17	19	327	8	870	0.49	0.31	0.010	0.273	0.374	0.038	111.1	5.7E-26
Hylocichla mustelina	156	11	13	19	19	681	8	471	11	13	20	20	346	9	806	0.59	0.37	0.002	0.300	0.397	0.030	215.4	9.1E-49
Mniotilta varia	143	12	14	19	20	500	8	665	10	13	18	20	393	10	772	0.43	0.39	0.014	0.337	0.409	0.021	27.2	1.81E-07
Myiarchus crinitus	38	14	17	20	20	610	6	660	8	12	14	15	291	12	979	0.48	0.36	0.011	0.229	0.226	0.037	217.2	3.64E-49
Nannus troglodytes	165	17	18	20	20	626	3	517	7	9	13	17	267	13	876	0.55	0.35	0.008	0.234	0.382	0.031	234.7	5.67E-53
Parula americana	155	15	16	19	20	451	5	702	9	13	14	16	251	11	902	0.39	0.36	0.024	0.218	0.275	0.047	91.3	1.2E-21
Picoides pubescens	11	11	18	18	20	682	9	615	4	8	10	11	1297	20	804	0.53	0.36	0.006	0.380	*1.00	0.012	63.7	1.45E-15
Piranga olivacea	17	12	15	19	20	667	8	624	7	10	14	15	239	13	1052	0.52	0.38	0.006	0.185	0.250	0.042	300.7	2.29E-67
Poecile atricapilla	270	17	18	19	20	711	3	327	6	12	15	15	330	14	708	0.68	0.35	−0.002	0.318	0.172	0.022	249.8	2.85E-56
Regulus calendula	30	9	14	18	20	328	11	950	5	8	14	17	152	15	1126	0.26	0.41	0.029	0.119	0.412	0.048	113.8	1.45E-26
Regulus satrapa	39	7	11	17	18	264	13	1005	4	8	12	14	333	16	936	0.21	0.38	0.033	0.262	0.143	0.016	*16.4	5.11E-05
Seiurus aurocapillus	410	8	11	17	18	411	12	487	4	6	11	11	241	16	657	0.46	0.38	0.010	0.268	−0.455	0.026	81.0	2.2E-19
Seiurus noveboracensis	15	5	12	17	19	393	15	900	5	10	15	17	457	15	836	0.30	0.44	0.011	0.353	0.368	0.009	*8.9	0.002785
Setophaga ruticilla	223	12	14	16	19	707	8	378	5	11	14	15	344	15	741	0.65	0.37	−0.001	0.317	0.178	0.017	181.6	2.2E-41
Sitta canadensis	47	11	15	20	20	518	9	743	9	11	14	16	371	11	890	0.41	0.38	0.014	0.294	0.303	0.027	49.5	2.01E-12
Sitta carolinensis	8	9	12	19	20	528	11	772	8	11	15	16	191	11	1109	0.41	0.42	0.010	0.147	0.338	0.059	226.7	3.13E-51
Sphyrapicus varius	87	20	20	20	20	786	0	435	12	14	15	15	267	8	954	0.64	0.33	0.002	0.219	0.238	0.059	463.8	7.2E-103
Vireo olivaceus	541	13	17	19	20	455	7	312	9	9	13	16	238	11	529	0.59	0.36	0.004	0.310	0.333	0.038	124.3	7.25E-29
Vireo solitarius	67	17	18	20	20	790	3	451	11	13	17	18	471	9	770	0.64	0.35	0.001	0.380	0.347	0.019	191.0	1.95E-43
Wilsonia canadensis	22	18	20	20	20	888	2	398	7	9	14	15	358	13	928	0.69	0.33	−0.001	0.278	0.270	0.023	417.6	8.07E-93

(Scott *et al.*, 1996), suggests that modelling distributions for gap analysis could be streamlined considerably via GARP approaches.

Finally, GARP models offer numerous capabilities not available with present gap models. For example, integration of current distributional patterns with the effects of global climate change yields new distributional predictions that must be taken into account if management decisions are to be robust to changes expected in coming decades; such modelling efforts are eminently feasible with GARP models (Peterson *et al.*, 2001*b*; Peterson *et al.*, 2002*b*). Species' invasions are also easily incorporated into GARP approaches (Peterson & Vieglais, 2001; A. T. Peterson, M. Papes & D. A. Kluza, unpubl. data; A. T. Peterson, R. Scachetti-Pereira & W. W. Hargrove, unpubl. data; A. T. Peterson, R. Scachetti-Pereira & D. A. Kluza, unpubl. data). Hence, a variety of new functionalities would open for the gap community, as well as simply improved predictive ability in the core mission of the programme.

Acknowledgements

REFERENCES

Anderson, R. P., Laverde, M. & Peterson, A. T. (2002*a*). Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Glob. Ecol. and Biogeog.* **11**: 131–141.

Anderson, R. P., Laverde, M. & Peterson, A. T. (2002*b*). Using niche-based GIS modelling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* **93**: 3–16.

Anderson, R. P., Lew, D. & Peterson, A. T. (In press). Using intermodel variation in error components to select best subsets of ecological niche models. *Ecol. Modell.*

Austin, M. P., Nicholls, A. O. & Margules, C. R. (1990). Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecol. Monogr.* **60**: 161–177.

Boone, R. B. & Krohn, W. B. (1998). *Maine Gap Analysis vertebrate data, Part II: Distribution, habitat relations, and status of breeding birds in Maine*. Orono, ME: Maine Cooperative Fish and Wildlife Research Unit, University of Maine.

Csuti, B. (1996). Mapping animal distribution areas for gap analysis. In *Gap analysis: a landscape approach to biodiversity planning*: 135–145. Scott, J. M., Tear, T. H. & Davis, F. W. (Eds). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

DeGraaf, R. M., Yamasaki, M., Leak, W. B. & Lanier, J. W. (1992). *New England wildlife: management of forested habitats*. General Technical Report NE-144. Radnor, PA: US Department of Agriculture.

Feria, T. P. & Peterson, A. T. (2002). Using point occurrence data and inferential algorithms to predict local communities of birds. *Divers. and Distrib.* **8**: 49–56.

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

Grinnell, J. (1917). Field tests of theories concerning distributional control. *Am. Nat.* **51**: 115–128.

Grinnell, J. (1924). Geography and evolution. *Ecology* **5**: 225–229.

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Sym. Quant. Biol.* **22**:415–427.

Krohn, W. B. (1996). Predicted vertebrate distributions from gap analysis: considerations in the designs of statewide accuracy assessments. In *Gap analysis: a landscape approach to biodiversity planning*: 147–162. Scott, J. M., Tear, T. H. & Davis, F. W. (Eds). Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Krohn, W. B., Boone, R. B., Sader, S. A., Hepinstall, J. A., Schaefer, S. M. & Painton, S. L. (1998). *Maine Gap Analysis: a geographic analysis of biodiversity*. Moscow, ID: US Biological Resources Division, Gap Analysis Program.

Nichols, J. D., Boulinier, T., Hines, J. E., Pollack, K. H. & Sauer, J. H. (1998). Inference models for spatial variation in species richness and community composition when not all species are detected. *Conserv. Biol.* **12**: 1390–1398.

Nix, H. A. (1986). A biogeographic analysis of Australian elapid snakes. In *Atlas of elapid snakes of Australia*: 4–15. Longmore, R. (Ed.). Canberra: Australian Government Publishing Service.

Peterson, A. T. (2001). Predicting species' geographic distributions based on ecological niche modelling. *Condor* **103**: 599–605.

Peterson, A. T., Ball, L. G. & Cohoon, K. C. (2002*a*). Predicting distributions of tropical birds. *Ibis* **144**: e27–e32.

Peterson, A. T. & Cohoon, K. C. (1999). Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol. Model.* **117**: 159–164.

Peterson, A. T., Navarro-Siguenza, A. G. & Benitez-Diaz, H. (1998). The need for continued scientific collecting: a geographic analysis of Mexican bird specimens. *Ibis* **140**: 288–294.

Peterson, A. T., Ortega-Huerta, M. A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R. H. & Stockwell, D. R. B. (2002*b*). Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**: 626–629.

Peterson, A. T. & Robins, C. R. (In press). When endangered meets invasive: ecological niche modelling predicts double trouble for spotted owls, *Strix occidentalis*. *Conserv. Biol.*

Peterson, A. T., Sanchez-Cordero, V., Beard, C. B. & Ramsey, J. M. (2002*c*). Ecologic niche modelling and potential reservoirs for Chagas disease, Mexico. *Emerg. Infect. Dis.* **8**: 662–667.

Peterson, A. T., Sanchez-Cordero, V., Soberon, J., Bartley, J., Buddemeier, R. H. & Navarro-Siguenza, A. G. (2001). Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecol. Model.* **144**: 21–30.

Peterson, A. T., Soberon, J. & Sanchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. *Science* **285**: 1265–1267.

Peterson, A. T., Stockwell, D. R. B. & Kluza, D. A. (2002*d*). Distributional prediction based on ecological niche modelling of primary occurrence data. In *Predicting species occurrences: issues of scale and accuracy*: 617–623. Scott, J. M. (Ed.). Washington, DC: Island Press.

Peterson, A. T. & Vieglais, D. A. (2001). Predicting species invasions using ecological niche modelling. *BioScience* **51**: 363–371.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Anderson, H., Caicco, S. L., D'Ericha, F., Edwards, T. C. Jr., Ullman, J. & Wright, R. G. (1993). Gap analysis: a geographic approach to the protection of biological diversity. *Wildl. Monogr.* **23**: 1–41.

Scott, J. M., Heglund, P. J. & Morrison, M. L. (Eds). (2002). *Predicting species occurrences: issues of accuracy and scale*. Washington, DC: Island Press.

Scott, J. M., Tear, T. H. & Davis, F. W. (Eds). (1996). *Gap analysis: a landscape approach to biodiversity planning*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing,

Stockwell, D. R. B. (1999). Genetic algorithms II. In *Machine learning methods for ecological applications:* 123–144. Fielding, A. H. (Ed.). Boston: Kluwer Academic Publishers.

Stockwell, D. R. B. & Noble, I. R. (1992). Induction of sets of rules from animal distribution data: a robust and informative method of analysis. *Maths. Comput. in Simul.* **33**: 385–390.

Stockwell, D. R. B. & Peters, D. P. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. of Geog. Inf. Syst.* **13**: 143–158.

Stockwell, D. R. B. & Peterson, A. T. (2002*a*). Controlling bias in biodiversity data. In *Predicting species occurrences: issues of scale and accuracy:* 537–546. Scott, J. M. (Ed.). Washington, DC: Island Press.

Stockwell, D. R. B. & Peterson, A. T. (2002*b*). Effects of sample size on accuracy of species distribution models. *Ecol. Model.* **148**: 1–13.

Walker, P. A. & Cocks, K. D. (1991). HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Glob. Ecol. Biogeog. Lett.* **1**: 108–118.