

# Inferring distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa) in geographic and ecological space using ecological niche modeling

Bastian Bentlage\*, A. Townsend Peterson, Paulyn Cartwright

Department of Ecology and Evolutionary Biology, The University of Kansas, 1200 Sunnyside Avenue, Lawrence, Kansas 66045, USA

**ABSTRACT:** Geographic distributions of many marine species are poorly documented or understood, which is particularly true for marine invertebrates. Ecological niche modeling (ENM) offers a means to address this issue, but to date most studies using ENM have focused on terrestrial taxa. In general, ENM relates environmental information to species' occurrence data to estimate the ecological niche of a species, rather than just interpolating a geographic distribution. This process leads to predictions of suitable habitat that generally exceed the range actually inhabited by a single species: such areas of geographic over-prediction (commission) may be inhabited by closely related species, and the model thus offers the inferential power to predict the potential distributions of these species as well. We explored the utility of ENM to investigate potential distributions of chirodropid box-jellyfishes (Cnidaria: Cubozoa), a group of highly toxic invertebrates whose biogeography is poorly understood. We were able to predict reported occurrences of box-jellyfishes throughout the Indo-Pacific from data of closely related species. By doing so, we demonstrate that geographic over-prediction in ENM can be desirable when concerned with predictions beyond current knowledge of species' distributions. Several methods are used for ENM; here, we compared the 2 most commonly used methods, the Genetic Algorithm for Rule-Set Predictions (GARP) and a maximum entropy approach (Maxent). Our comparison shows that Maxent may be more prone to overfitting, whereas GARP tends to produce broader predictions. Transforming continuous Maxent predictions into binary predictions remedies problems of overfitting, and allows for effective extrapolation into unsampled geographic space.

**KEY WORDS:** Ecological niche modeling · Cubozoa · Chirodropida

*Resale or republication not permitted without written consent of the publisher*

## INTRODUCTION

Understanding geographic ranges of species is crucial in studying biogeographic as well as large-scale ecological questions in the marine realm. A particular problem in studying marine species' distributions arises from the paucity of reliable occurrence records that frequently represent only a fraction of the species' actual ranges (Guinotte et al. 2006), providing an incomplete picture of distributions and factors affecting them. Ecological niche modeling (ENM) offers a means of addressing the problem of limited occurrence data on species' distributions by providing predictions

about species potential ranges based on environmental parameters (Soberón & Peterson 2004).

In ENM, known occurrences of species are related to raster environmental data layers in an evolutionary-computing environment to reconstruct species' distributions in environmental dimensions. ENMs can be projected across broader landscapes than the original sampling area to identify a potential geographic distribution for the species. ENMs can then be used for a variety of applications, including design of sampling strategies (e.g. Guisan et al. 2006, Pearson et al. 2007) and identification of areas vulnerable to species invasions (Peterson 2003). ENM is also being employed as

\*Email: bentlage@ku.edu

a forecasting tool for biotic responses to changing climates (e.g. Martínez-Meyer & Peterson 2006) and can be incorporated into analyses of evolution of species' autecology (e.g. Martínez-Meyer et al. 2003).

In particular, theoretical studies predict a high degree of conservatism of ecological niches among closely related species over evolutionary timescales (Holt & Gaines 1992), which has subsequently been demonstrated empirically (e.g. Peterson et al. 1999, Martínez-Meyer & Peterson 2006, Kambhampati & Peterson 2007, Peterson & Nyári 2008, but see Rice et al. 2003 and Graham et al. 2004 for counterexamples). In essence, conservatism is expected since mortality rates are higher under conditions outside the niche (Holt & Gaines 1992). This property of ecological niches may allow prediction of species' distributions and ecology from information regarding their relatives. Specifically, in a phylogenetic context, ENMs can be used to test hypotheses about evolution of species' ecology. For example, investigating patterns of overlap of niche requirements of closely related species may allow untangling the roles of ecology and historical contingencies in shaping of species' distributions and the process of speciation (see Swenson 2008 for a review).

ENM studies to date have focused almost exclusively on terrestrial fauna and flora and, although ENM has been demonstrated to produce robust predictions of geographic distributions of marine species (Wiley et al. 2003, Kluza & McNyset 2005, Guinotte et al. 2006, Bryan & Metaxas 2007, Therriault & Herborg 2008), it has seen only limited application by marine biologists thus far. Our aim was to expand ENM applications to evaluate the utility of these methods for understanding potential distributions and large-scale ecological properties of 3 species of chirodropid cubozoans.

Cubozoa encompasses approximately 40 described species in 2 orders, Carybdeida and Chirodropida (see Daly et al. 2007). Like many other medusozoan cnidarians, cubozoans are characterized by alternation of benthic polyp and free-swimming medusa generations. Here, we focus on potential distributions of the medusa generation in chirodropid box-jellyfishes; polyps have been observed in the wild only once (Hartwick 1991), so knowledge of this life stage is limited. Chirodropid medusae can frequently be encountered in coastal waters in the tropics and subtropics and may display species-specific patterns of seasonal occurrence and abundance.

Many cubozoan species contain potent toxins in their stinging capsules (called nematocysts), and their high toxicity regularly leads to severe stings that can cause cardiac dysfunction or cardiac arrest in humans (reviewed by Winkel et al. 2003). For this reason, cubozoans have received considerable attention from scientists and authorities in Australia, where they represent

a public health threat and cause financial losses in the tourism industry (Bailey et al. 2003). In particular, the Australian chirodropid *Chironex fleckeri* is considered the world's most lethal jellyfish (Fenner & Williamson 1996). Given its practical importance, its distribution in Australian waters is better documented than those of other cubozoan species. Using data from museum records, we investigated the degree to which ENM approaches are able to predict the species' native geographic range.

Outside of Australia, knowledge of chirodropid distributions is mainly anecdotal, but assuming that niche conservatism holds for chirodropids, it should be possible to extrapolate niche predictions based on Australian *Chironex fleckeri* to predict the potential distribution of related species. We also modeled niche distributions of 2 Asian chirodropids (*Chiropsoides buitendijki* sensu lato and *Chironex yamaguchii*), using what limited occurrence information was available. By doing so, we modeled possible geographic ranges of chirodropids throughout Oceania and Asia. We used these models to investigate niche overlap in geographic and ecological space among the 3 chirodropid species. This study demonstrates that ENM can improve understanding of distributions and ecological properties of marine organisms, even when very few distributional data are available.

## MATERIALS AND METHODS

**Study species: taxonomy and nomenclature.** Of the 3 species studied, *Chironex fleckeri* Southcott, 1956 is the only one with a stable taxonomic history. *Chiropsoides buitendijki* (van der Horst, 1907) and *Chiropsoides quadrigatus* (Haeckel, 1880) have seen historical and current taxonomic debate (Mayer 1917, Thiel 1928, Stiasny 1937, Gershwin 2006, Lewis & Bentlage 2009). *C. quadrigatus* was originally described from Burmese waters (Haeckel 1880) and later redescribed based on material from the Philippines (Mayer 1910). The original description, however, was based on a single damaged, juvenile specimen that differs strikingly from Philippine and Japanese chirodropids (Thiel 1928, Stiasny 1937, Gershwin 2006, Lewis & Bentlage 2009). In particular, the few interpretable characters of Haeckel's *C. quadrigatus* closely resemble *C. buitendijki*, which was originally described from Java (van der Horst 1907), suggesting that they may be synonymous (Gershwin 2006). Hence, we included Haeckel's *C. quadrigatus* specimen from Burma in our *C. buitendijki* modeling experiments. The taxonomic confusion surrounding *C. quadrigatus* misled others to identify the chirodropid species that can be found in the Philippines and Japan as *C. quadrigatus*. A recent

study (Lewis & Bentlage 2009) demonstrated this case of mistaken identity, and described the Philippine/Japanese chirodroid as *Chironex yamaguchii* Lewis & Bentlage, 2009.

**Input data.** Occurrence data were accumulated from catalogued museum specimens and the scientific literature (Fig. 1, Table 1). Spatial accuracy of data points varied from source to source; some specimens were georeferenced with reasonably precise latitude and longitude, whereas other locality information consisted of textual descriptions of collection sites only. The latter were georeferenced for the present study, and hence may be subject to greater uncertainty (Table 1).

Environmental data included raster geographic information system grid layers summarizing yearly averages of remotely sensed environmental parameters and ocean colors from the Aqua-MODIS satellite (Table 2; Savtchenko et al. 2004). Aqua-MODIS ocean

color satellite imagery for 2003 to 2006 was averaged across the binned yearly composite images. We also included seafloor topographic grids derived from a combination of satellite measurements of marine gravity field anomalies and ship depth soundings (Smith & Sandwell 1997). All environmental grids were resampled to a resolution of 2.52' for analysis using nearest-neighbor sampling because ENM algorithms require all grids to be at the same resolution and extent. All grids except bathymetry were already on the same resolution and scale, but the native resolution of the bathymetry grid was 2'.

**Ecological niche modeling.** Several approaches have been used and explored for ENM (reviewed by Elith et al. 2006). From these algorithms, we employed 2 commonly used evolutionary-computing algorithms, the Genetic Algorithm for Rule-Set Prediction (GARP; Stockwell & Peters 1999) and a maximum entropy

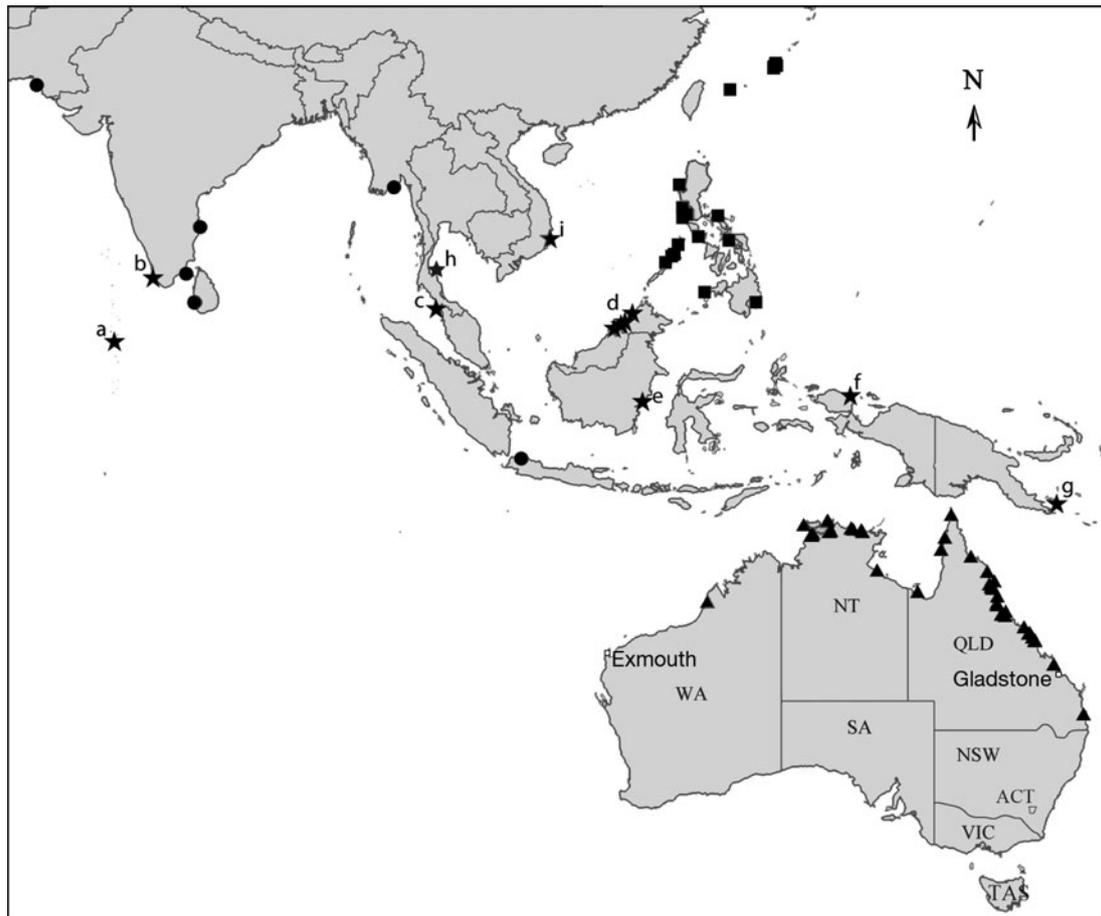


Fig. 1. Occurrence points used in ecological niche modeling, as well as locations from which undetermined chirodroid box-jellyfishes have been reported (the latter were not used in model-building). Reports from (a) the Maldives in Stiasny (1937), (b) India in Nair (1951), (c–g) Thailand (Andaman Sea), Malaysia, Brunei, Sulawesi, Irian Jaya, and Papua New Guinea, respectively, in Fenner & Williamson (1996), (h) Thailand (Gulf of Siam) in Suntrarachun et al. (2001), (i) Vietnam in Dawydoff (1936) and Ranson (1945). ▲: *Chironex fleckeri*; ●: *Chiropsoides buitendijki*; ■: *Chironex yamaguchii*; ★: approximate locations from which indeterminate chirodroids have been reported. ACT: Australian Capital Territory, NSW: New South Wales, NT: Northern Territory, QLD: Queensland, TAS: Tasmania, VIC: Victoria, WA: Western Australia

approach (Maxent; Phillips et al. 2004, 2006). Both GARP and Maxent are designed specifically to manage presence-only data, which offers advantages over other algorithms that require both presence and absence data for model development. Absence information could be used to allow the modeling algorithm to evaluate commission error (i.e. the area predicted present but not actually inhabited by the species). However, absence data are particularly problematic because non-observation of a species at a locality may result from lack of sampling, lack of opportunity for dispersal to sites with suitable habitat, vicariant speciation, or real absence of suitable conditions (Anderson et al. 2003), particularly in poorly-known organisms like chironomids. Below, we outline the implementation of each algorithm, but Stockwell & Peters (1999) and Phillips et al. (2004, 2006) provide more in-depth explanations of the respective software packages.

GARP is a heuristic algorithm that iteratively 'evolves' rule-sets (if-then statements regarding environmental conditions) that are combined into final predictions. First, input occurrence data are divided into training data for model development and extrinsic testing data for model selection. GARP then works in an iterative process, in which it first selects a method from a set of possibilities (logistic regression, bioclimatic rules, range rules, negated range rules), and applies this method to the training data. The resulting rule is evaluated for its significance, based on a set of 1250 presence points generated by resampling from a subset of the input training points with replacement and an equal number of pseudoabsence points (i.e. points generated at random from grid pixels at which the species has not been recorded). In subsequent iterations, rules are changed (evolved) randomly, based on a set of change operators designed to mimic chromosomal evolution (e.g. crossing-over, deletion), thus introducing elements of stochasticity in the rule-changing process. Significance of

Table 1. Sources of occurrence data used to model species' ecological niches. Italicized latitude and longitude values indicate records that were georeferenced by us based on a textual description of the sampling locality. Where applicable, museum catalog numbers or references to the scientific literature are provided. NCL: National Chemical Laboratory India; USNM: US National Museum of Natural History; MAGNT: Museum and Art Galleries of the Northern Territory; QM: Queensland Museum; SAM: South Australian Museum. AUS: Australia; PH: Philippines; JP: Japan; ID: Indonesia; LK: Sri Lanka; IN: India; BUR: Burma; PK: Pakistan

Species Place of origin	Lat	Long	Source
<b><i>Chironex fleckeri</i></b>			
Moreton Bay, QLD, AUS	-27.29	153.26	SAM
Yeppoon, QLD, AUS	-23.14	150.82	SAM
Mackay, QLD, AUS	-21.20	149.27	QM G317061
	-21.12	149.26	SAM
	-20.90	149.05	QM G317015
Repulse Bay, QLD, AUS	-20.58	148.75	QM G3569
Prosperpine, QLD, AUS	-20.50	148.81	SAM
Bowen, QLD, AUS	-20.07	148.35	SAM
Townsville, QLD, AUS	-19.14	146.78	SAM
	-19.02	146.46	SAM
Ross River, QLD, AUS	-18.73	146.84	SAM
Cardwell, QLD, AUS	-18.23	146.09	SAM
Tully Head, QLD, AUS	-18.07	146.09	SAM
Broome, WA, AUS	-17.95	122.21	SAM
Thompson's Creek, WA, AUS	-17.48	146.15	SAM
Sweers Island, QLD, AUS	-17.15	139.60	QM G322298
Cairns, QLD, AUS	-16.86	145.81	QM G4140
	-16.83	145.77	SAM
Machan's Beach, QLD, AUS	-16.80	145.76	SAM
Gladstone, QLD, AUS	-16.68	145.64	SAM
Port Douglas, QLD, AUS	-16.48	145.47	QM G317043
Kurrimine Beach, QLD, AUS	-16.22	145.90	NMM F81653
Townsville, QLD, AUS	-15.46	145.35	SAM
McArthur River, NT, AUS	-15.38	136.25	MAGNT C011739
Bathurst Head, QLD, AUS	-14.22	143.97	SAM
Aurukun, QLD, AUS	-13.60	141.51	SAM
Weipa, QLD, AUS	-12.64	141.84	SAM
Darwin, NT, AUS	-12.49	130.84	SAM
	-12.48	130.84	SAM
	-12.48	130.83	MAGNT C012108
	-12.48	130.84	MAGNT C011617
	-12.48	130.84	MAGNT C005479
	-12.47	130.87	SAM
	-12.47	130.85	MAGNT C014843
	-12.42	130.85	MAGNT C013719
	-12.42	130.83	MAGNT C011147
	-12.40	130.80	MAGNT C003862
	-12.40	130.81	MAGNT C004706
Berry Springs, NT, AUS	-12.35	130.83	MAGNT C011641
Darwin, NT, AUS	-12.35	130.88	MAGNT C011185
	-12.34	130.88	MAGNT C011186
Shoal Bay, NT, AUS	-12.32	130.97	MAGNT C011975
Kakadu National Park, NT, AUS	-12.19	132.34	MAGNT C014953
Milingimbi, NT, AUS	-12.15	135.01	MAGNT C014936
Kakadu National Park, NT, AUS	-12.15	132.24	MAGNT C014934
	-12.06	132.39	MAGNT C014933
Milingimbi, NT, AUS	-12.03	134.93	MAGNT C014952
Maningrida, NT, AUS	-11.97	134.23	MAGNT C014951
	-11.85	134.08	MAGNT C014954
Bathurst Island, NT, AUS	-11.60	130.17	SAM
Port Essington, NT, AUS	-11.27	132.12	MAGNT C015246
Garik Gunak Baru National Park, NT, AUS	-11.20	132.14	SAM
Simpson Point, QLD, AUS	-10.74	142.38	SAM

Table 1 (continued)

Species Place of origin	Lat	Long	Source
<b><i>Chironex yamaguchi</i></b>			
Pujada Bay, Mindanao, PH	6.91	126.24	USNM 27914
Panabutan Bay, Mindanao, PH	7.70	122.00	USNM 27916
Ulugan Bay, Palawan, PH	10.14	118.78	USNM 28691
Malampaya Sound, Palawan, PH	10.80	119.40	USNM 28699
Taytay, Palawan, PH	10.88	119.54	USNM 38016
Malcochin, Linapacan, PH	11.60	119.80	USNM 28692
Cataingan Bay, Masbate, PH	12.02	124.03	USNM 27913
Mansalay Bay, Mindoro, PH	12.30	121.50	USNM 27917
Tilik Bay, Lubang, PH	13.90	123.11	USNM 28698
San Miguel Bay, Luzon, PH	13.90	123.11	USNM 28695
Hamillo Point, Luzon, PH	14.15	120.56	USNM 28700
Subic Bay, Luzon, PH	14.74	120.22	USNM 27911
Bolinao Bay, Luzon, PH	16.60	119.90	USNM 28697
Okinawa Island, JP	26.33	127.75	USNM 1121556
Nakagusuka Bay, Okinawa Island, JP	26.45	127.83	QM G317050
Kana Beach, Okinawa Island, JP	26.47	127.97	QM G317051
Chatan, Okinawa Island, JP	26.30	127.73	QM G317064
Ginowan, Okinawa Island, JP	26.28	127.72	Lewis & Bentlage (2009)
Chatan Beach, Okinawa Island, JP	26.30	127.73	Lewis & Bentlage (2009)
Motobu Port, Okinawa Island, JP	26.65	127.88	Lewis & Bentlage (2009)
Sukuji Beach, Ishigaki Island, JP	24.44	124.12	Lewis & Bentlage (2009)
<b><i>Chiropsoides buitendijki</i></b>			
Jakarta, Java, ID	-6.10	106.87	van der Horst (1907)
Mount Lavinia, Colombo, LK	6.82	79.87	Fernando (1992)
	6.86	79.86	Fernando (1992)
Krusadai Island, Tamil Nadu, IN	9.23	79.22	Menon (1936)
Chennai, Tamil Nadu, IN	13.09	80.33	NCL 106964
<b><i>Chiropsides quadrigatus</i></b>			
Rangoon, BUR	16.37	96.36	Haeckel (1880)
<b><i>Chiropsoides buitendijki</i></b>			
Sandspit, Karachi, PK	24.84	66.92	Tahera & Kazmi (2006)

Table 2. Parameters used for model building, including the loadings of the first 5 principal components (PC) and their importance expressed in cumulative proportions of variance; loadings < 0.1 not shown. All data were continuous. The spatial resolution of topography data was 2'; that of all other parameters was 2.52'. Topography data were acquired from <http://topex.ucsd.edu/> (accessed August 2002); all other data were from <http://oceancolor.gsfc.nasa.gov/> (accessed December 2007). nlw: normalized water-leaving radiance

Parameter	PC 1	PC 2	PC 3	PC 4	PC 5
Seafloor topography	-0.224	0.116	-0.607	0.682	0.329
Sea surface temperature	0.211	0.115	-0.744	-0.419	-0.408
Chlorophyll	0.423			-0.140	
Calcite	-0.365	0.278	0.193		-0.187
Diffuse attenuation	-0.415			-0.193	
nlw at 412 nm	0.386	0.244			0.126
nlw at 443 nm	0.365	0.322			
nlw at 488 nm	0.529	0.529	0.156		
nlw at 531 nm	-0.243	0.485		0.114	-0.505
nlw at 551 nm	-0.183	0.465		-0.491	0.637
Cumulative proportion of variance	0.540	0.790	0.890	0.950	0.990

changes in the predictive values (presence or absence of the species) before and after the application of the new rule is evaluated using a significance parameter outlined by Stockwell & Peters (1999). As GARP is a stochastic approach, alternate runs using the same input data lead to different results. Hence, following best-practices approaches (Anderson et al. 2003), we generated 200 models for each species (previous tests showed that 1000 replicates do not improve the final model compared to 100 replicates). From these 200 models we retained 20 models (in our cross-validation experiments; see below) or 40 models (for our large-scale projections) that showed the lowest omission error, and then discarded the 50% that predicted the most extreme values of areas present ('commission error index' of Anderson et al. 2003). This 'best subset' of replicate models was then summed, pixel by pixel, to create a final prediction.

Maxent addresses the issue of presence-only modeling differently compared to GARP, as it does not rely as explicitly on pseudoabsence data. The basic idea behind Maxent is to model a probability distribution for species occurrence that is most spread out (i.e. exhibits the highest degree of entropy), given certain constraints. In terms of ENM, the constraints are represented by the values of environmental parameters observed at known occurrence points (Phillips et al. 2004, 2006). Maximum entropy is invoked in Maxent because it gives the least biased estimate possible on the given information when characterizing unknown events with a statistical model (Jaynes 1957). In addition to using raw environmental data values (features) as input, Maxent also uses transformations (i.e. feature vectors) of the variables when the input features are continuous variables, which may be the variables squared, multiplications of 2 or more variables, categorizations, and other types of transformations. Feature vectors are included in the modeling process to improve the model-to-data fit. The auto feature option that we used lets Maxent find the feature-vector combina-

tion most suitable to the particular input data set based on sample sizes available (for further information see Phillips et al. 2006, Phillips & Dudík 2008).

Rather than assigning a probability from 0 to 1 to each pixel, the Maxent probability distribution sums to 1 over the entire grid to produce a 'raw' output. Such individual grid cells are assigned very small probability values, making a prediction more difficult to interpret on a map. This problem is addressed by offering 2 additional output formats: cumulative output in which the value assigned to each pixel is the sum of probabilities assigned to all pixels that have equal or smaller raw probability values, and logistic output, which is a logistic transformation of the raw values, providing a probability of species' presence (or habitat suitability) for every grid cell ranging from 0 to 1 (Phillips & Dudík 2008). We use the logistic output format throughout this contribution, but note that all 3 output formats are monotonically related to each other. In contrast to GARP, Maxent is deterministic, so every run with the same dataset returns the same answer.

We performed all experiments using Desktop GARP (ver. 1.1.6; [www.nhm.ku.edu/desktopgarp](http://www.nhm.ku.edu/desktopgarp)) and Maxent (ver. 3.0; [www.cs.princeton.edu/~schapire/maxent](http://www.cs.princeton.edu/~schapire/maxent)). For model building, we considered only spatially unique occurrence points, and models were trained on the same geographic area over which they were projected. Models were built with a random selection of 15% of occurrence points set aside for internal model evaluation by the algorithms. Maxent models were imported into ArcView 3.2 (ESRI) as floating-point grids and then multiplied by 1000 and converted into integer grids, since the former are difficult to manipulate.

**K-fold cross-validation.** To investigate the ability of our ENM techniques to predict known occurrences, we divided the Australian study area (i.e. *Chironex fleckeri*) into 6 sectors containing roughly similar numbers of spatially unique occurrence points. We restricted this cross-validation approach to Australia since the other 2 areas (= species) held only very limited occurrence points. Models were built (1) using occurrence points falling in 5 of the 6 sectors to predict the distribution of points in the sixth, and (2) using points falling in half of the sectors to predict the other half. The significance test we employed to evaluate model performance follows Anderson et al. (2002, 2003): given the proportion of pixels predicted present versus absent in each evaluation sector, we tested whether occurrence points fell into pixels predicted present more often than would be expected at random. Cumulative, 1-tailed binomial probabilities of observed proportions of test points falling in pixels of predicted presence and absence at different threshold levels of model output probabilities were calculated to test for significance. In Maxent, because the area of predicted occurrence be-

comes extremely small in the upper part of the output probability distribution, we reclassified the continuous probability distribution of the output model into a discrete one using several thresholds (i.e.  $>0$ ,  $>0.01$ ,  $>0.1$ ,  $>0.3$ ,  $>0.5$ , and  $>0.8$ ) for each of which we calculated binomial probabilities. For GARP, binomial probabilities were calculated for each congruence level of the 10 best models.

**Large-scale predictions and visualizations in ecological space.** To generate broad-scale predictions of predicted distributions for each species, we set a threshold for prediction of presence as the lowest prediction score, which represents the model output values at points of known occurrence (Pearson et al. 2007). In a few cases, training points fell into grid cells that had been assigned 0 probability of occurrence by the model; as these pixels were adjacent to pixels predicted at a higher threshold, we assigned these occurrences the threshold that was immediately adjacent. To explore more restrictive thresholds, we also calculated the average probability value assigned by the model to grid cells containing occurrence points and set this value as a more stringent threshold for prediction of presence.

Finally, we explored the distribution of each modeled niche in environmental dimensions. We used the 'combine grids' option in the Grid Transformation Tools extension of ArcView to create a composite of all environmental parameter grids with the final ecological niche model grids. We exported the attributes table associated with this grid and then imported a random subset of 100 000 pixel values for visualization in S-Plus (Insightful Corporation). To reduce the dimensionality of this space, we employed principal components analysis based on the correlation matrix. Niche models were transformed into binary predictions using the same thresholds as described above, and areas of modeled niche space were then related to overall availability of environmental conditions in 2-dimensional scatterplots. We also related the environmental conditions within a 250 nautical mile (nmi) (= 463 km) buffer zone around each occurrence point to overall environmental conditions. By this means, we could evaluate which environmental conditions are likely to be encountered by each species in the immediate vicinity of its known populations.

## RESULTS

### Effects of missing data and sampling biases

Models generated performed well in validation exercises in which distributional data from 5 regions were used to train a model that predicts distributions in a

sixth region: in all cases, occurrence points set aside during model building were predicted significantly better than random ( $p < 0.01$ ; Fig. 2a). In Maxent,

under these conditions, the high-probability values clustered around occurrence points used in model building (Fig. 2b). Maxent models nonetheless per-

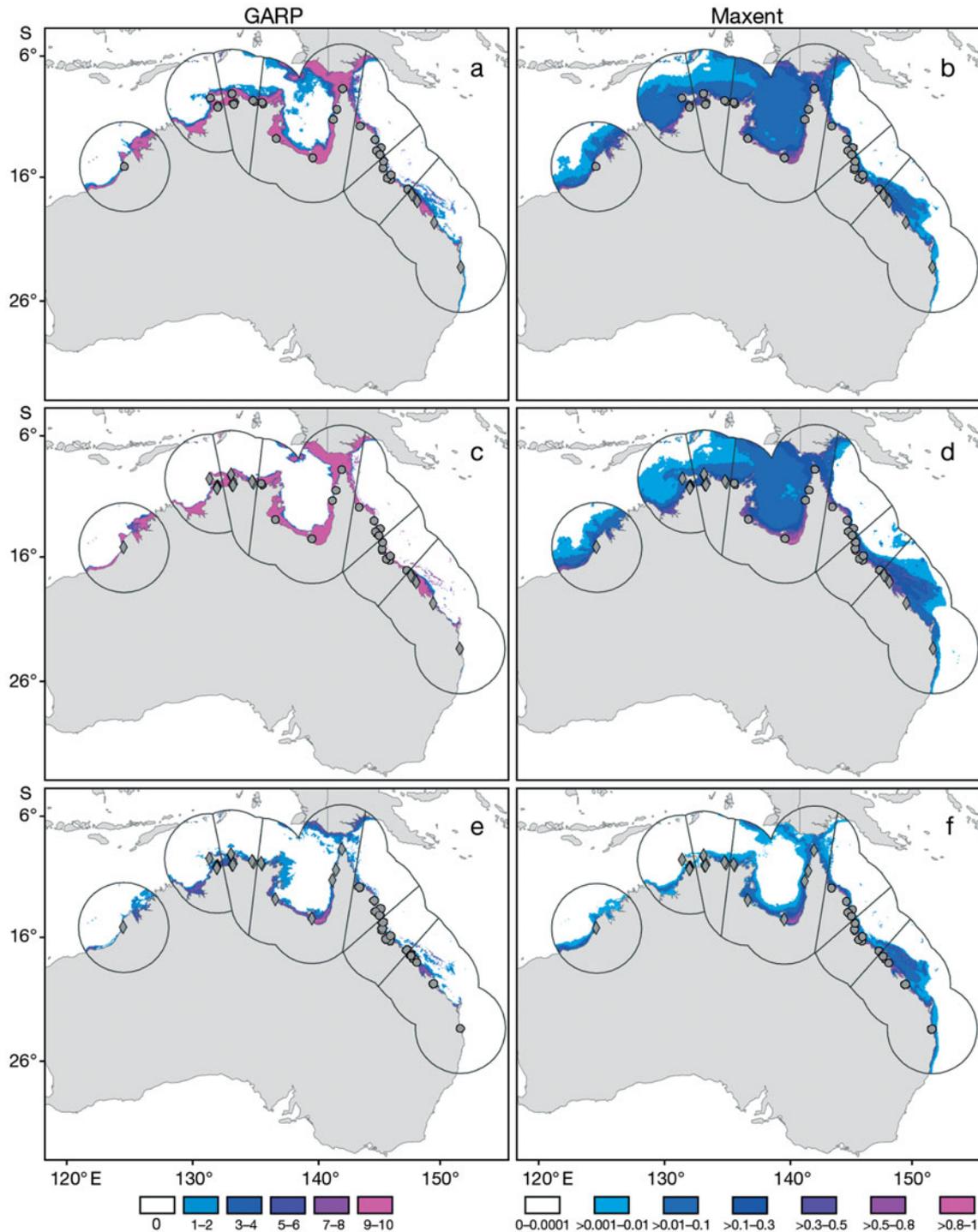


Fig. 2. Representative results from model-validation experiments with Australian *Chironex fleckeri*. Niche models were generated by GARP and Maxent within 5 or 3 of the 6 sectors (sectors are delineated by straight lines) of the buffer zone that encompasses an area of 250 nautical miles (= 463 km) around each occurrence point. Strength of prediction is indicated according to the color gradient (see key). The GARP scores from 1 to 10 represent congruence among the 10 best models; Maxent output represents a probability distribution of occurrence. ●: occurrence records used for model building; ◆: occurrence points used for model evaluation

formed well in extrapolating into 'unsampled' sectors at lower thresholds (i.e. around 0.8; all  $p < 0.01$ ).

More extreme validation experiments, in which half of the regions were set aside for model testing (a total of 15 experiments in each GARP and Maxent), revealed similar differences between GARP and Maxent (Fig. 2c–f). GARP performed well in most cases, predicting points falling into sectors left out from model building with high precision ( $p < 0.01$ ). However, in 2 cases presenting the most extreme spatial biases (i.e. training points all from 1 extreme of the geographic distribution), GARP failed to predict points in the 'unsampled' sectors (both  $p > 0.05$ ; e.g. Fig. 2e). Maxent models behaved similarly and failed in the same 2 scenarios (both  $p > 0.05$ ; e.g. Fig. 2f).

### Large-scale predictions and niche overlap among species

Although GARP and Maxent models differed in several aspects, they displayed largely congruent patterns (Fig. 3). In general, GARP predicted broader areas for *Chiropsoides buitendijki* and *Chironex yamaguchii* than the corresponding Maxent models, but more stringent GARP thresholds and less stringent Maxent thresholds yielded geographic predictions that were closely similar. The GARP model differs from the Maxent model in this comparison in predicting areas offshore of southwestern and southern Australia, as well as near New Zealand. Niche models for *C. fleckeri* and *C. buitendijki* were largely congruent using both algorithms (Fig. 3). The models for *C. yamaguchii*, however, showed only limited overlap with the other 2 species, regardless of modeling method: in fact, the models for this species versus the other 2 seemed to be almost the 'negative' of one another.

Considering only areas of agreement between the 2 modeling techniques, predictions for each species were as follows. Niche models for *Chironex fleckeri* predicted suitable habitat along the northern Australian coast from approximately Exmouth, Western Australia, to central and southern Queensland, including southern Papua New Guinea. In addition, large coastal areas in Southeast Asia are predicted, ranging roughly from Vietnam to Pakistan. The prediction for *Chiropsoides buitendijki* is very similar as regards the geographic extent of the prediction, but with more area predicted in Asia and less in Australia. In both cases, the Philippines are predicted as unsuitable, except for a few pixels. By contrast, niche models for *C. yamaguchii* are distinct from those of the other species, and Philippine distributional areas, and areas of Sulawesi and the lesser Sunda Islands are predicted as suitable. As noted above, this model is nearly the

inverse of the other 2, predicting coasts opposite to the ones predicted by the other models (e.g. the northern coast of Papua New Guinea).

In ecological space, 79% of the variation in the 10-dimensional environmental dataset are explained by its first 2 principal components (Table 2). Projecting the niche models for each species from geographic space into environmental dimensions makes the overlap and differentiation among the 3 niche models especially apparent (Fig. 4). The pattern is similar to that observed in geographic space: overlap among *Chironex fleckeri* and *Chiropsoides buitendijki* niche models is extensive, but the *C. yamaguchii* model occupies a region of environmental space adjacent to, but distinct from, the other 2 species. The only conditions under which *C. yamaguchii* overlaps extensively with the other 2 species is when the lenient GARP prediction is used, which appears to show excessive commission error in both geographic and environmental space. This differentiation is mirrored in the environments that are 'accessible' (i.e. with a 250 nmi buffer zone around each occurrence point) for each species. Whereas the buffered regions for the former 2 appear quite similar, the environment surrounding *C. yamaguchii* in the Philippines and Japan overlaps only on one extreme with those of *C. fleckeri* and *C. buitendijki*.

## DISCUSSION

### Modeling techniques

ENM offers a promising avenue by which to investigate the biogeography of marine invertebrates for which distributional data are often depauperate. In particular, the ready availability of ENM algorithms facilitates exploration of the technique for the GIS user. However, some degree of care is necessary when interpreting ENM results. For example, when challenged to predict species' occurrences across broad, unsampled regions, differences between GARP and Maxent emerge. In general, both algorithms predict testing data considerably better than random expectations. The major difference between GARP and Maxent predictions parallel those documented previously for terrestrial taxa (Peterson et al. 2007): GARP models are more general than those generated by Maxent, thus performing better in extrapolation challenges like the partitioning experiment we conducted, but are more prone to overprediction. In particular, the highest probabilities of predicted occurrence in Maxent closely associate with the points on which the model was trained, which leads to predictive failures of the higher probability levels of many Maxent models. In essence, the number of pixels that are assigned higher

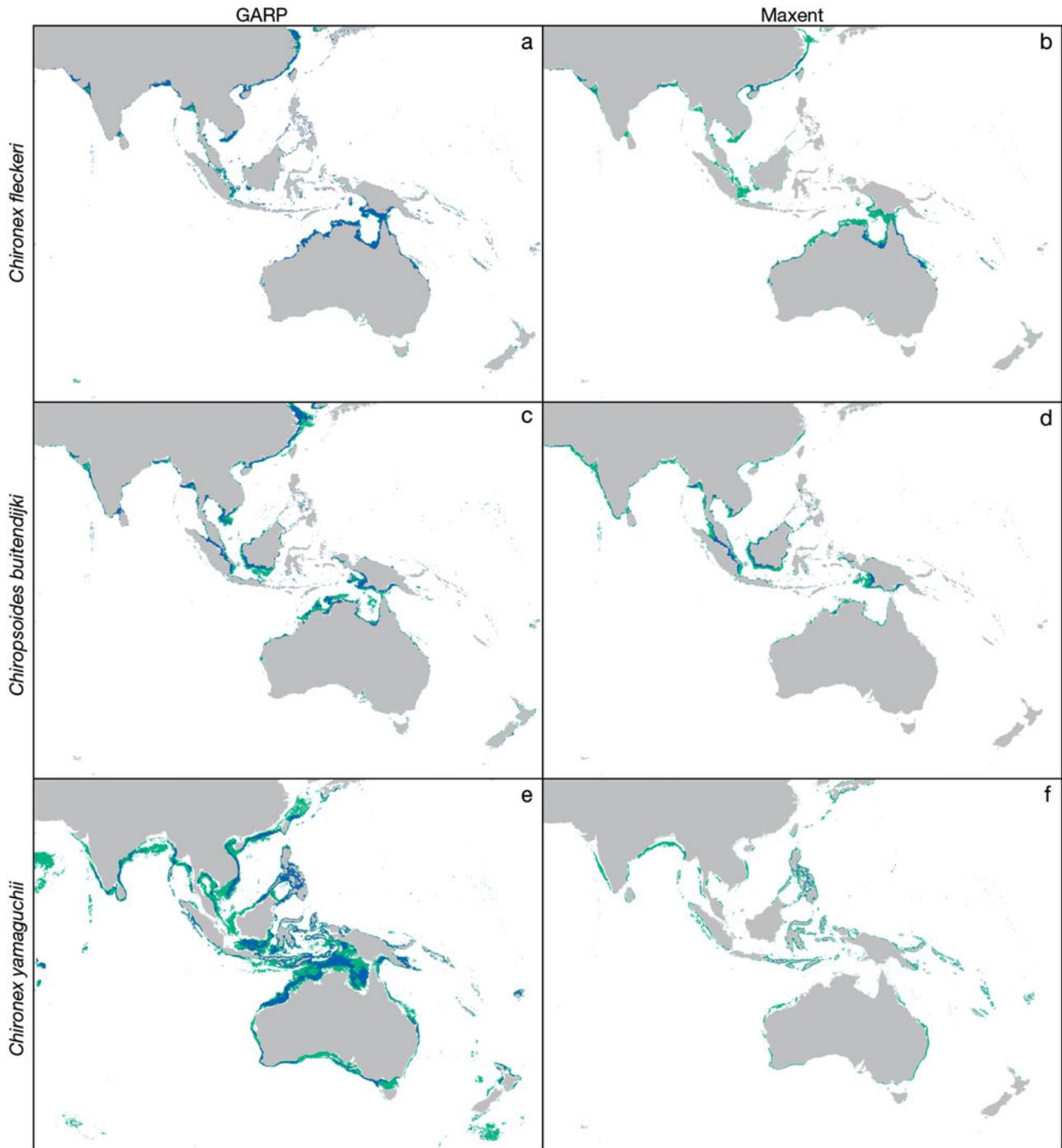


Fig. 3. Geographic distribution predicted for the 3 chirodropid species throughout Oceania and Asia, as inferred by GARP and Maxent. Color indicates the threshold applied to the broad-scale predictions. Green: lenient threshold; blue: stringent threshold

suitability scores is so small that numbers of occurrence data points omitted from predictions become extremely large. As such, thresholding becomes a critical step in considering predictions derived from Maxent models: above a certain probability level, Maxent models tend to reflect overfitting (i.e. they appear to be

more accurate in fitting known data than predicting new data, thus hampering extrapolation; Peterson et al. 2007). Recent changes to Maxent output (i.e. the 'logistic' output), which we employed here, represent a rescaling of the probability distribution to deemphasize this distributional skew (Phillips & Dudík 2008) but

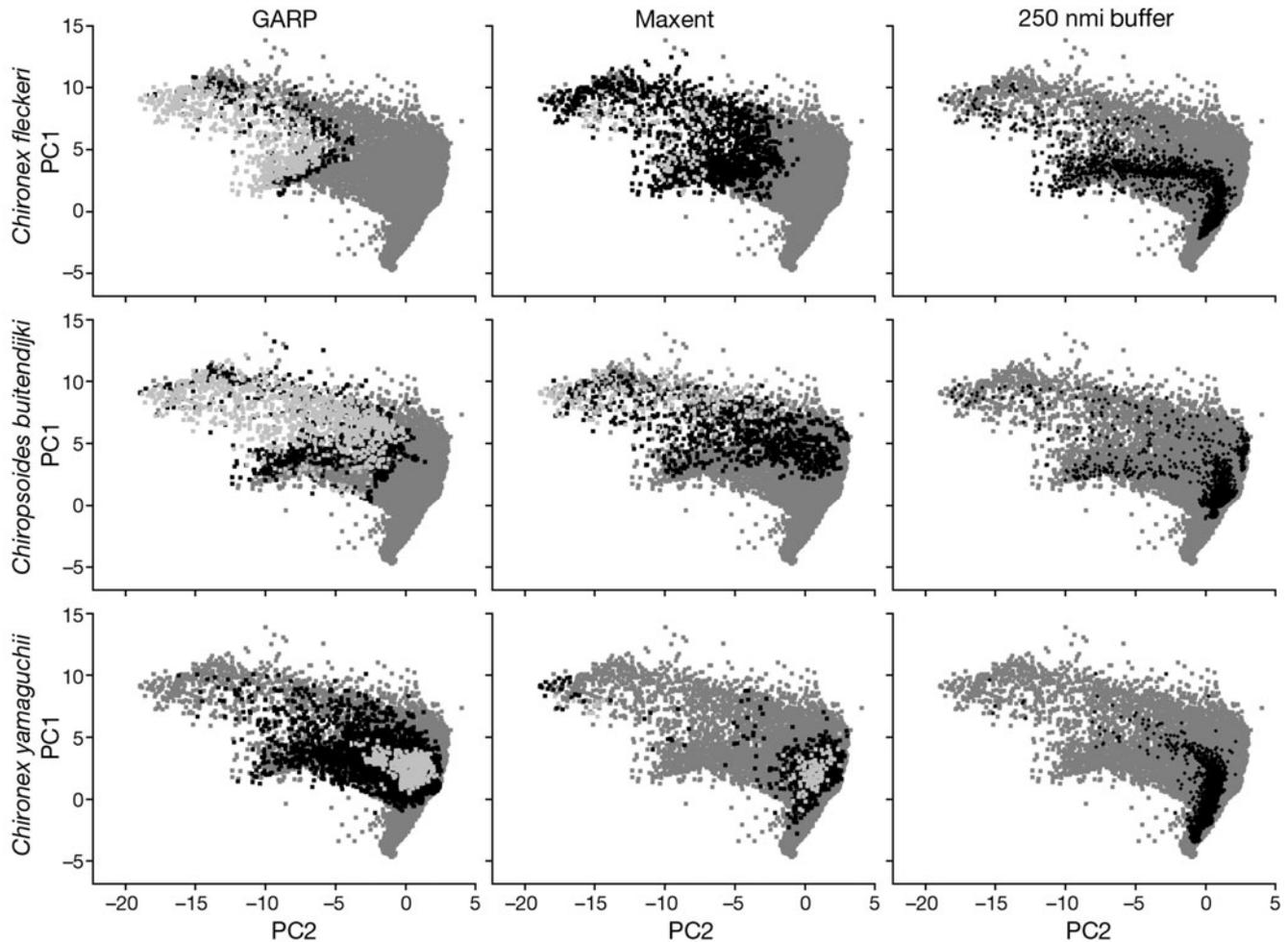


Fig. 4. Distribution of the first 2 principal components (PC) of each final niche model in the entire environmental space of the study area and proportion of the entire environmental space 'accessible' to each species. Dark gray squares: environmental space available in the study region; black squares (left and middle column): area occupied by models under lenient threshold; light gray squares: area occupied by models under stringent threshold; black diamonds (right column): environmental conditions present in a 250 nautical mile (nmi) (= 463 km) buffer around occurrence data points

do not seem to completely remedy the problems pointed out by Peterson et al. (2007).

Our results support the idea that the broader 'ramp' of probability values in Maxent models is less meaningful, and should be interpreted with caution (Peterson et al. 2007). Since all Maxent outputs are related to one another monotonically (Phillips & Dudík 2008) this is true for both output options. A post hoc transformation of continuous Maxent results into binary values by thresholding, however, seems to produce reasonable predictions that allow effective extrapolation. Interestingly, although application of different strategies for thresholding niche models produces results that are much more congruent between GARP and Maxent, GARP models generally cover broader areas when applying the same thresholding criteria to both GARP and Maxent models. Prediction beyond present knowl-

edge of species' distributions, in general, is desirable in ENM, but the question is how much of such overprediction or commission 'error' is biologically informative? Here, ENM 'overprediction' permitted us to predict distributional patterns of species across broad, unsampled areas with considerable accuracy, and to compare patterns of ecological niche occupancy. We were able to predict among species, as well as predict chirodripid occurrences throughout the Indo-Pacific. In this sense, as has been pointed out previously (Soberón & Peterson 2005), commission 'error' is quite desirable in ENM exercises. It is important to note, however, that no algorithm is capable of effective extrapolation if the input data for modeling are biased. In our most extreme spatial stratification experiments (Fig. 2e–f), extrapolation exercises failed most likely because the occurrence data for model training were

sampled from one extreme of an ecological gradient (see Stockwell & Peterson 2002, Peterson 2005).

### Ecological niche comparisons

One may argue that, instead of modeling ecological niches, it may be easier and more desirable simply to interpolate among known occurrence points (Bahn & McGill 2007). For well-sampled species, this approach might be a viable means of estimating realized distributions. ENM, as noted earlier, goes beyond reconstructing species' distributions to assemble a predictive model of their ecological requirements, and as such offers considerably more inferential power.

Among the 3 species analyzed, the geographic distribution of the Australian *Chironex fleckeri* is best known (i.e. from Exmouth, Western Australia, to Gladstone, Queensland). However, current distributional maps (e.g. Fenner & Williamson 1996) do not account for the possibility that the width of the species' distribution along the coastline is unlikely to be uniform. ENMs offer predictions that create a more detailed and explicit picture. For example, 2 occurrence points off the coast of Queensland were not included in any prediction from either algorithms (not discernible in our figures); these specimens may indeed have been sampled that far offshore and identified correctly, but the models suggest that those sites are atypical of the species' occurrence. In particular, these records may represent samples from sink populations, which would be consistent with the idea that *C. fleckeri* does not venture far off the coast into deep waters (e.g. Hooper 2000). In other areas, however, the models suggest that the species may occur farther offshore than the actual occurrence points would indicate (e.g. at the Northern Territory and Queensland boundary), which is a hypothesis that can be tested via additional sampling efforts.

Further, we obtained a single record from Moreton Bay, Queensland, for *Chironex fleckeri*, some 400 km south of Gladstone. In almost all modeling experiments, Moreton Bay was predicted suitable for the species. The Moreton Bay area predicted by the ENM is disjunct from the rest of the *C. fleckeri* potential distribution in Maxent models, and only loosely connected in GARP models (Fig. 3). Hence, it may well be that the bay represents a place to which individuals are dispersed by the predominantly southward-flowing coastal currents, and then may be viable. Clearly, these predictions await additional sampling effort to be iteratively confirmed, rejected, refined, and reinterpreted. Nonetheless, ENMs present a means of offering up hypotheses of geographic distributions of species within their native ranges, which can be tested through further sampling.

That ENM captures the ecological niche, rather than simply the geographic distribution, leads to predictions of suitable habitat for a given species in geographic space that generally exceeds the range it actually inhabits. In some cases, given prevalent niche conservatism (Peterson et al. 1999), sampling areas of 'overprediction' will lead to discovery of related species (Raxworthy et al. 2003), as shown by broad niche overlap between *Chironex fleckeri* and *Chiropsoides buitendijki* throughout Oceania and Asia. Indeed, in several places predicted by these models, chirodropid occurrences have been noted (Fig. 1). Only in a few of these cases has the animal been captured (Dawydoff 1936, Stiasny 1937, Ranson 1945, Nair 1951); at most of these sites, however, characteristic sting marks and clinical symptoms provide the only evidence of chirodropid occurrences (Fenner & Williamson 1996, Suntrarachun et al. 2001). Similarly, the *C. yamaguchii* ENM predicts several areas from which chirodropids have been reported, but for which no specific determination has been possible (Fenner & Williamson 1996).

### Niche differentiation

Somewhat surprisingly, the *Chironex yamaguchii* niche model was distinct from the models for *C. fleckeri* and *Chiropsoides buitendijki*. In particular, ecological theory would lead one to expect that the congeners *C. fleckeri* and *C. yamaguchii* should share more ecological characteristics than either does with *C. buitendijki*. It could be argued that the taxonomic framework is wrong, but this seems unlikely (Lewis & Bentlage 2009, B. Bentlage, P. Cartwright & A. G. Collins unpubl.). Niche conservatism is an expectation, but its violation does not imply that model fit is erroneous. Rather, organisms evolve new characteristics, and instances in which the expectation of discovering niche conservation is not met represent interesting avenues for further inquiry.

Niche differentiation (sensu Kambhampati & Peterson 2007) can be real (i.e. 2 species inhabit different niches, despite having access to the same set of environmental conditions) or apparent (i.e. 2 species inhabit different niches because the environments accessible to each are different). Given these considerations, it may be argued that the differentiation we observe is apparent rather than real. The environment accessible to *Chironex yamaguchii* is simply very different from the environments surrounding the other 2 species (Fig. 4; 250 nmi buffer). Similar conditions do exist in all 3 environments, but the area of overlap is limited to a few pixels (Fig. 4).

Niche differentiation may thus have taken place by ancestral species establishing themselves in the rare,

shared environments (which may have been more extensive at the time) with dispersal into more common environments occurring subsequently. Alternatively, all environments were more equal in the past, and changing climates affected different localities in different ways, leading to niche differentiation. A comprehensive study integrating ENMs with phylogenetic analyses and paleoceanography will allow testing such scenarios.

The exact ecological nature of this differentiation is somewhat complex since a full interpretation of the loadings of each principal component does not seem feasible at this point. Ocean color, such as water-leaving radiances at different wavelengths, which strongly contributes to principal components 1 and 2, has been used previously to distinguish among marine habitats (e.g. Mishra et al. 2005). We suggest that the ocean color can be interpreted as a proxy for habitat structure. However, large-scale ground-truthing will be required to fully interpret which habitats the models predict.

### CONCLUDING REMARKS

We have outlined the types of questions that can be addressed with ENM concerning the distribution, large scale ecology, and evolution of marine taxa. ENMs offer the opportunity to address questions on broad spatial scales that may not be easily addressed using other approaches. Undoubtedly, the predictions derived through modeling approaches will benefit much from evaluation using experimental approaches to test the models' ecological components as well as sampling programs to evaluate their geographic predictions.

*Acknowledgements.* Funding was provided through National Science Foundation Assembling the Tree of Life grant EF-0531779 to P.C. and PADI Foundation grant 173-2007 to B.B. For help with satellite imagery processing and GIS, we thank M. Papeş, and we thank A. G. Collins and J. Soberón for insightful discussions. P. Alderslade, M. Evans, and T. Laperousaz made access to various museum databases possible. Two anonymous reviewers provided critical and insightful critiques that improved this manuscript considerably.

### LITERATURE CITED

- Anderson RP, Gómez-Laverde M, Peterson AT (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Glob Ecol Biogeogr* 11:131–141
- Anderson RP, Lew D, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol Model* 162:211–232
- Bahn V, McGill BJ (2007) Can niche-based distribution models outperform spatial interpolation? *Glob Ecol Biogeogr* 16:733–742
- Bailey PM, Little M, Jelinek GA, Wilce JA (2003) Jellyfish envenoming syndromes: unknown toxic mechanisms and unproven therapies. *Med J Aust* 178:34–37
- Bryan TL, Metaxas A (2007) Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific continental margins of North America. *Mar Ecol Prog Ser* 330:113–126
- Daly M, Brugler MR, Cartwright P, Collins AG and others (2007) The phylum Cnidaria: a review of phylogenetic patterns and diversity 300 years after Linnaeus. *Zootaxa* 1668:127–182
- Dawydoff C (1936) Observations sur la faune pelagique des eaux indochinoises de la mer de Chine meridionale (note preliminaire). *Bull Soc Zool Fr* 61:461–484
- Elith J, Graham CH, Anderson RP, Dudík M and others (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Fenner PJ, Williamson JA (1996) Worldwide deaths and severe envenomation from jellyfish stings. *Med J Aust* 165: 658–661
- Fernando M (1992) Some hazards of diving. *Ceylon Med J* 37:72–80
- Gershwin LA (2006) Comments on *Chiropsalmus* (Cnidaria: Cubozoa: Chirodropida): a preliminary revision of the Chiropsalmidae, with descriptions of two new genera and two new species. *Zootaxa* 1231:1–42
- Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781–1793
- Guinotte JM, Bartley JD, Iqbal A, Fautin DG, Buddemeier RW (2006) Modeling habitat distribution from organism occurrences and environmental data: case study using anemonefishes and their sea anemone hosts. *Mar Ecol Prog Ser* 316:269–283
- Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE (2006) Using niche-based models to improve the sampling of rare species. *Conserv Biol* 20:501–511
- Haeckel E (1880) System der Acraspeden—Zweite Hälfte des Systems der Medusen. *Denkschriften der Medizinisch-Naturwissenschaftlichen Gesellschaft zu Jena, Jena*
- Hartwick RF (1991) Distributional ecology and behavior of the early life stages of the box-jellyfish *Chironex fleckeri*. *Hydrobiologia* 216-217:181–188
- Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evol Ecol* 6:433–447
- Hooper JNA (2000) Box jellyfish. In: Ryan M, Burwell C (eds) *Wildlife of tropical north Queensland*. Queensland Museum, Brisbane, p 6–7
- Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:620–630
- Kambhampati S, Peterson AT (2007) Ecological niche conservation and differentiation in the wood-feeding cockroaches, *Cryptocercus*, in the United States. *Biol J Linn Soc* 90:457–466
- Kluza DA, McNyset M (2005) Ecological niche modeling of aquatic invasive species. *Aquat Invaders* 16:1–7
- Lewis C, Bentlage B (2009) Clarifying the identity of the Japanese Habu-kurage, *Chironex yamaguchii*, sp. nov. (Cnidaria: Cubozoa: Chirodropida). *Zootaxa* 2030:59–65
- Martínez-Meyer E, Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *J Biogeogr* 33:1779–1789

- Martínez-Meyer E, Peterson AT, Navarro-Sigüenza AG (2003) Evolution of seasonal ecological niches in *Passerina* buntings (Aves: Cardinalidae). *Proc R Soc Lond B Biol Sci* 271:1151–1157
- Mayer AG (1910) *Medusae of the world, Vol III. The Scyphomedusae*. Carnegie Institution of Washington, Washington, DC
- Mayer AG (1917) Report upon the Scyphomedusae collected by the US Bureau of Fisheries steamer 'Albatross' in the Philippine and Malay Archipelago. *Bull US Natl Mus* 100: 175–233
- Menon MGK (1936) Scyphomedusae of Krusadai Island. *Bull Madras Gov Mus. Nat Sci Nat Hist Sect* 1:1–9
- Mishra DR, Narumalani S, Rundquist D, Lawson M (2005) High-resolution ocean color remote sensing of benthic habitats: a case study at the Roatan Island, Honduras. *IEEE Trans Geosci Rem Sens* 43:1592–1604
- Nair KK (1951) *Medusae of the Trivandrum coast. Part 1. Systematics*. *Bull Cent Res Inst Univ Travancore Ser C Nat Sci* 2:47–75
- Pearson RG, Raxworthy CJ, Nakamura M, Peterson AT (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J Biogeogr* 34:102–117
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol* 78: 419–433
- Peterson AT (2005) Predicting potential geographic ranges of invading species. *Curr Sci* 89:9
- Peterson AT, Nyári AS (2008) Ecological niche conservatism and Pleistocene refugia in the Thrush-like Mourner, *Schiffornis* sp., in the neotropics. *Evolution* 62:173–183
- Peterson AT, Soberón J, Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science* 285:1265–1267
- Peterson AT, Papeş M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 4:550–560
- Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175
- Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. *Proc 21st Int Conf Machine Learning*. ACM, Banff, AB, p 83
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190:231–259
- Ranson G (1945) Les scyphomeduses de la collection du Museum National d'Histoire Paris. II. Catalogs raisonne; origine des recoltes. *Bull Mus Natl Hist Nat Paris* 2:312–320
- Raxworthy CJ, Martínez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA, Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426:837–841
- Rice NH, Martínez-Meyer E, Peterson AT (2003) Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. *Biol J Linn Soc* 80:369–383
- Savtchenko A, Ouzounov D, Ahmad S, Acker J, Leptoukh G, Koziana J, Nickless D (2004) Terra and Aqua MODIS products available from NASA GES DAAC. *Adv Space Res* 34:710–714
- Smith WHF, Sandwell DT (1997) Global seafloor topography from satellite altimetry and ship depth soundings. *Science* 277:1956–1962
- Soberón J, Peterson AT (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philos Trans R Soc Lond B* 359:689–698
- Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inf* 2:1–10
- Stiasny G (1937) Scyphomedusae. *John Murray Expedition 1933–1934. Sci Rep* 4:203–242
- Stockwell D, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 13:143–158
- Stockwell D, Peterson AT (2002) Controlling bias in biodiversity data. In: Scott JM, Heglund PJ, Morrison ML (eds) *Predicting species occurrences: issues of scale and accuracy*. Island Press, Washington, DC, p 537–546
- Suntrarachun S, Roselieb M, Wilde H, Sitprija V (2001) A fatal jellyfish encounter in the Gulf of Siam. *J Travel Med* 8: 150–151
- Swenson NG (2008) The past and future influence of geographic information systems on hybrid zone, phylogeographic and speciation research. *J Evol Biol* 21:421–434
- Tahera Q, Kazmi QB (2006) New records of two jellyfish medusae (Cnidaria: Scyphozoa: Catostylidae: Cubozoa: Chirodropidae) from Pakistani waters. *J Mar Biol Assoc UK* 86:1482
- Therriault TW, Herborg LM (2008) Predicting the potential distribution of the vase tunicate *Ciona intestinalis* in Canadian waters: informing a risk assessment. *ICES J Mar Sci* 65:788–794
- Thiel ME (1928) Die Scyphomedusen des Zoologischen Staatsinstituts und Zoologischen Museums in Hamburg: I, Cubomedusae, Stauromedusae und Coronatae. *Mitt Zool Mus Hamburg* 43:1–34
- van der Horst R (1907) On a new cubomedusa from the Java-Sea: *Chiropsalmus buitendijki*. *Notes Leyden Mus* 29: 101–106
- Wiley EO, McNyset KM, Peterson AT, Robins CR, Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography (Wash DC)* 16: 120–127
- Winkel KD, Hawdon GM, Fenner PJ, Gershwin LA, Collins AG, Tibballs J (2003) Jellyfish antivenoms: past, present, and future. *J Toxicol Toxin Rev* 22:115–127

Editorial responsibility: Hans-Heinrich Janssen, Oldendorf/Luhe, Germany

Submitted: September 25, 2008; Accepted: March 16, 2009  
Proofs received from author(s): May 19, 2009