

Development and evaluation of statistical habitat suitability models: an example based on juvenile spotted seatrout *Cynoscion nebulosus*

Sven Kupschus*

Fish and Wildlife Conservation Commission, Florida Marine Research Institute, 100 8th Avenue SE, St. Petersburg, Florida 33701, USA

*Present address: The Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

ABSTRACT: Conservation and fisheries managers require models that describe the abundance and distribution of fishes in order to protect and manage exploited fish stocks in the face of anthropogenically induced habitat loss and exploitation. In this study, models describing environmental preferences were developed for juvenile spotted seatrout *Cynoscion nebulosus* within 3 Florida estuaries: the Indian River Lagoon, Tampa Bay, and Charlotte Harbor. A generalized additive model (GAM) was developed to describe the environmental preferences, spatial distribution, and temporal fluctuations of juvenile seatrout in each estuary based on a 4 yr time-series of fisheries-independent catches. All 3 models indicated similar environmental preferences for all populations examined and were also qualitatively consistent with the findings from other studies. Consequently, habitat-preference models based on GAMs are useful tools to predict fish abundances in estuaries lacking fisheries-independent data, given knowledge of the local environmental conditions. These initial findings were further supported by quantitative analyses of the models' abilities to predict abundance in independent datasets, despite complications with multicollinearity of independent variables and temporal differences in the recruitment periodicity between the 3 Florida populations. Estimates of environmentally based habitat value on a relative scale will aid conservation managers in protecting vital nursery habitats in estuaries currently lacking fisheries-independent information and in predicting the effects of future environmental change. Finally, assuming similar environmental conditions between years, the year effect from these models can serve as an index of relative abundance for fisheries managers to predict future recruitment strength and to tune catch-at-age stock-assessment models.

KEY WORDS: *Cynoscion nebulosus* · Fish distributions · Habitat suitability modelling · Generalized additive models (GAM)

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Conservation managers need to predict the abundance and distribution of fish species in estuaries lacking fisheries-independent monitoring data and to assess the possible effects that hydrodynamic changes caused by environmental perturbations may have on fish populations. Both needs can be met by models which accurately describe the relationship between species abundance and the local environment, given

information on environmental conditions or predictions on conditions following hydrological perturbation. Attempts have been made to model fish abundance or distribution quantitatively (parametrically) or qualitatively (or nonparametrically) as a response to environmental conditions (Rubec et al. 1999, Brown et al. 2000). However, these approaches fail to either model the relationship between environment and abundance realistically or in a quantifiable manner (Hastie & Tibshirani 1990).

Generalized additive models (GAMs) combine the flexibility of non-parametric models with the quantifiable statistical evaluation of error structures and model selection criteria. They represent a flexible semi-parametric modelling approach to determining the relationships between dependent and independent variables and, in contrast to traditional parametric habitat preference models, are not restricted to unrealistic monotonic multi-parameter models, or by the usual assumption of normality (McCullagh & Nelder 1989, Neter et al. 1996). The model-free parameter estimation through multiple flexible splines and likelihood-based error estimation covering a multitude of possible error functions, can overcome many of the restrictions of earlier habitat preference models (Hastie & Tibshirani 1990). Spline functions are not monotonic or tied to preconceived notions about the relationship between dependent and independent variables and can therefore more realistically describe the multitude of physiological and behavioural effects that determine the distribution of fishes. The number of possible error structures available due to the use of the log-likelihood approach allows for modelling the non-normal errors typical of survey data. GAMs have already been successfully employed to elucidate complex relationships between fish abundance and environmental conditions (Schwartzman et al. 1994, Bigelow et al. 1999, Maravelias et al. 2000). However, their ability to predict temporal and spatial variations in fish distributions in independent datasets has yet to be rigorously examined.

The aim of this study is not to develop the best possible model describing juvenile fish distributions in all estuaries, rather the aim is to test the ability of GAM models developed based on data in one estuary to describe the spatial and temporal distribution of fishes in other estuaries given local sets of environmental predictors. As a proof of concept of the use of GAMs in modelling fish distributions independent of a particular estuary, this study developed spatially and temporally dynamic models describing the distribution of spotted seatrout (*Cynoscion nebulosus*) based on their habitat preferences in 3 different estuaries. An in-depth investigation of habitat preferences and spatial distribution of juvenile spotted seatrout is fitting for several reasons. First, due to its considerable recreational fisheries value to Florida (Chester & Thayer 1990), much is known about its biology (Moffet 1961, Iversen & Tabb 1962, Tabb 1966, Peebles & Tolley 1988, McMichael & Peters 1989, Nelson & Leffler 2001). Such published information is critical to qualitatively verify the model results in terms of the species ecology and to judge the suitability of models for describing habitat usage and recruitment, even though nursery habitat is poorly defined and its ultimate role in determining recruitment strength is not

yet clear. Second, comparable fisheries-independent data sets are available for 3 Florida estuaries to facilitate quantitative comparison of models derived from fisheries-independent collections made in one estuary with the observed abundances of fishes collected in other estuaries (foreign models). Third, because the high level of fishing pressure and anthropogenically induced environmental change, there is an urgent need for tools to aid conservationists and fisheries managers in protecting these economically important resources (Helser et al. 1993, Rubec et al. 1999). Because of their limited migratory behaviour and their absolute dependence on estuarine habitats (Moffet 1961, Iversen & Tabb 1962, Tabb 1966), seatrout are threatened by localized extinction through naturally or anthropogenically induced habitat perturbations. This paper assesses the ability of GAM model to describe distributions based solely on environmental conditions and to examine where such models might help conservation managers in predicting fish distribution in estuaries lacking fisheries-independent monitoring information. In addition, an assessment was made of the year-effect from such pre-recruit models as a means of predicting future recruitment, which is important to fisheries managers.

MATERIALS AND METHODS

Study sites. Juvenile seatrout in 3 estuaries located at comparable latitudes were examined in this study, 2 populations on the west and 1 population on the east coast of central Florida (Fig. 1). Tampa Bay (TB) and Charlotte Harbor (CH) are bays separated from the Gulf of Mexico by barrier island systems. These 2 estuaries receive significant freshwater inflows from several mainland tributaries, creating a conventional salinity gradient, with salinities decreasing at increasing distances from the Gulf of Mexico. The Indian River Lagoon (IR), adjacent to the Atlantic Ocean, is made up of 2 parallel basins: the more westerly Indian River and the less extensive Banana River. In stark contrast to the tidally driven estuaries on the west coast, the IR hydrography is largely controlled by prevailing wind patterns (Smith 1987). The lagoon receives freshwater, but due to the location of its tributaries and the position of man-made connections to the Atlantic, the lagoon's unconventional salinity gradient is largely perpendicular to its net flow (DaCosta et al. 1987).

Despite hydrographic and zoogeographic differences, species composition in catches and available habitat types were similar in all 3 estuaries. Vegetated habitats included seagrass meadows, made up of *Halodule wrightii*, *Thalassia testudinum* or *Syringodium*

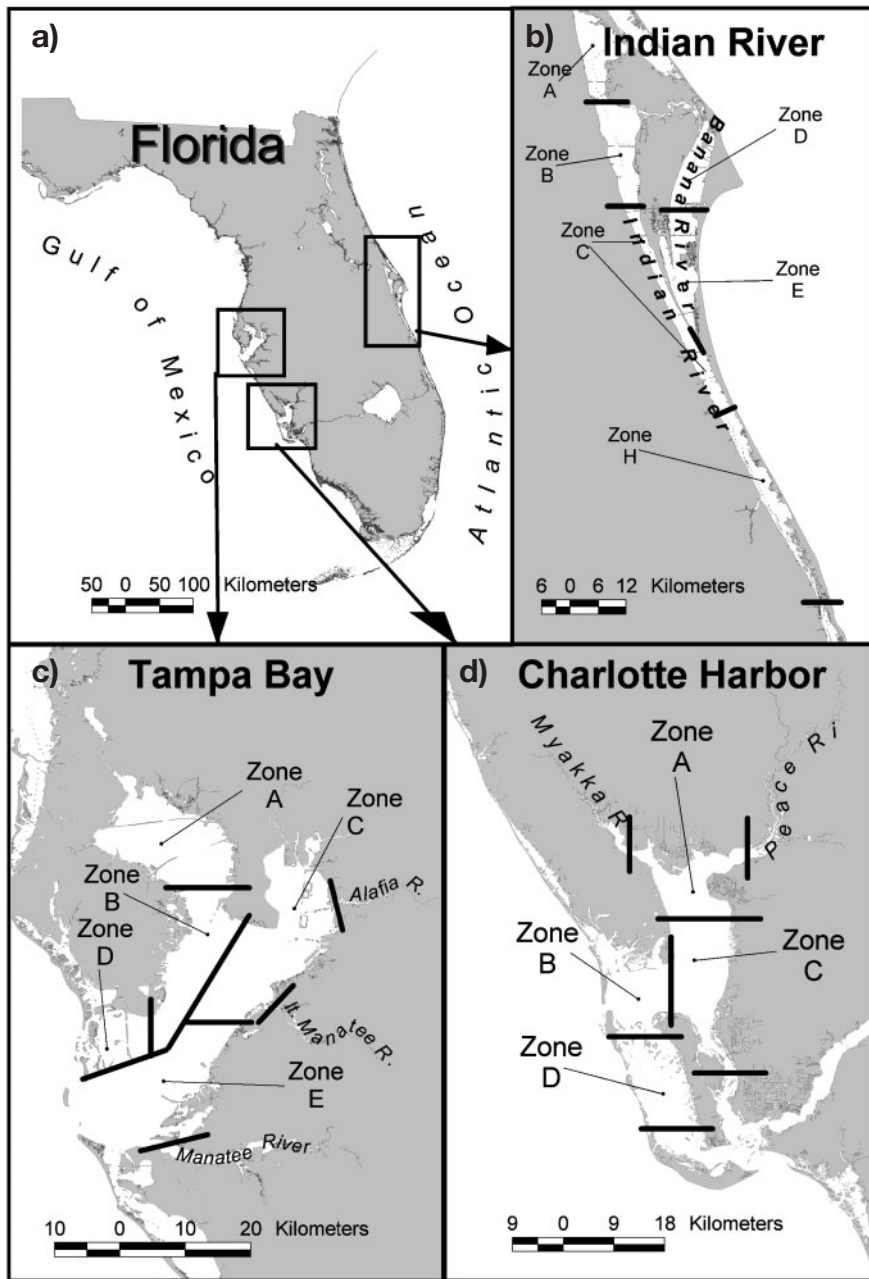


Fig. 1. (a) Locations of the 3 study estuaries in Florida and the proportionally prestratified zones of the sampling designs in (b) the Indian River Lagoon, (c) Tampa Bay and (d) Charlotte Harbor

filiforme, and mangrove fringing forest, predominantly made up of red mangroves (*Rhizophora mangle*).

Sample collection. Sampling in each estuaries was conducted from January 1996 to December 1999 using a monthly, multigear stratified random sampling (StRS) design, stratified spatially by zones (Fig. 1b–d) and proportionally stratified by habitat type within zones. Some changes to the sampling design, were made over the 4 yr period for logistical reasons, slightly compli-

cating the interpretation of the results. The stratified random sampling design in the IR was amended in January 1998 to include the more southerly Zone H and to reduce the monthly sampling effort in Zones A, B and E to a single fall sampling period (Fig. 1b). In TB, sampling in parts of Zone E was eliminated during January to May 1998, after which monitoring in the area resumed at effort levels lower than before (Fig. 1c). In CH, Zone D was dropped from the

sampling design (Fig. 1d) and sampling effort was decreased in all other zones in 1998.

Three different techniques, boat, offshore, and beach sets, were used to deploy a 21.3 m by 1.8 m center-bag seine (3.2 mm #35 knotless nylon Delta mesh) in order to effectively measure fish abundances in a wide variety of habitats (Nelson 2002). Sample sites were chosen randomly without replacement from a sampling universe unique to each deployment technique, consisting of 1' latitude by 1' longitude grids. Boat and offshore sets were proportionally prestratified, and beach sets were unstratified so that it was possible to collapse the stratified design into a simple random design for the purpose of this analysis.

Boat sets were used only in the tributaries of the west coast estuaries to sample steep riverine banks, while riverine habitats were not sampled at all in the IR. The net was set in a semicircle from the shore by boat and manually retrieved to the shore. Beach sets sampled the estuarine shoreline and were set out perpendicular to the shore and dragged for 9.1 m parallel to the shore before being retrieved similar to the boat sets. Offshore sets were used to capture fish away from the immediate shoreline by dragging the net for 9.1 m in water <1.5 m and retrieving it around a pivot-pole to concentrate the sample in the bag. The area swept differed for the 3 techniques, covering 68, 140 and 338 m² for boat sets, offshore sets, and beach sets, respectively.

At least 40 specimens were measured to the nearest mm standard length (SL) from each set, and in case of greater sample numbers the remainder were counted and raised by the length–frequency distribution. Synoptic water-quality data were collected, and additional variables on the spatial, temporal and biological properties of each sampling location were measured for each sample. For a more detailed description of the sampling design, deployment methods and environmental variables collected, a full procedure manual can be obtained from the Fisheries-Independent Monitoring Program.¹

Numerical analysis. Three independent models of the habitat preferences of juvenile seatrout were developed, each based on the data from one of 3 estuaries, CH, IR and TB. Only catch data for seatrout less than 50 mm SL were analysed, because larger seatrout have the ability to effectively avoid the sampling gear, and they tend to occupy alternative habitats as a consequence of ontogenetic behavioural changes such as diet shifts (McMichael & Peters 1989). Additionally, the initiation of schooling behaviour has also been docu-

mented for this species at >50 mm length (Tabb 1961). As larger specimens school, they are encountered more contagiously than would be expected in a random distribution. That is, the mere presence of fish increases the probability of higher abundance, while the absence of fish reduces the probability of high abundance. Schooling fish are better modelled as a negative binomial error distribution. In any case, mixing the assumed Poisson error distribution of the <50 mm seatrout with the negative binomial error distribution of larger schooling individuals would invalidate the error distribution of the model and lead to unrealistic estimates of confidence limits for parameters and hinder model selection through improper evaluation of the deviances.

S-Plus (Statistical analysis software, Insightful) was used to develop a separate GAM (Venables & Ripley 1999) with an assumed Poisson error distribution for each estuary system to describe the abundance of seatrout as a response to day of the year, year, temperature, salinity, water depth at the seine bag, percent seagrass cover, and bottom type (mud, sand, mud–sand mix, oysters with mud, oysters with sand). The choice of the environmental variables was based on either their reported performance as suitable indicators of seatrout abundance or their ability to characterize ichthyofaunal assemblages in estuaries (Kupchus & Tremain 2001). The year effect was parameterised as a classification variable in the model in order to examine interannual changes in the abundance of seatrout in the fisheries independent survey which were not explained by differences in the environmental conditions encountered during sampling.

The differences in effort of each of the deployment techniques associated with differences in the area swept needed to be accounted for in the analysis. Frequently, this is accomplished by dividing the catch by the area to arrive at a catch per unit effort dependent variable. This approach, however, disrupts the error structure, since for very small areas the problem tends towards a binomial problem (a fish caught or not), while on a larger spatial scale the problem tends towards a Poisson problem (how many fish are caught). In this study, to compensate for the differences in effort between the techniques, while maintaining the appropriate error structure the area covered was divided by the largest area (338 m²), and the resulting ratios were used as an offset in the GAM analysis (Venables & Ripley 1999).

A number of the measured environmental variables were highly correlated, because of the interplay of the sampling design and the individual hydrographic characteristics found in each estuary. This multicollinearity can complicate or invalidate the interpretation of the results of any regression analysis, particu-

¹Fisheries Independent Monitoring Program Procedure Manual, available from Robert McMichael, Jr, Florida Marine Research Institute, 100 8th Ave SE, St Petersburg, Florida 33701, USA

larly in linear regression models. To mitigate against this, one of a pair of correlated variables is usually excluded from models (Geary & Leser 1968, Yeo 1984). In the present study correlated variables were retained, because correlation does not make them necessarily redundant (Hamilton 1987). Instead, to avoid including truly redundant variables in the analysis, yet retain all important environmental information in the analysis, the Akaike information criterion (AIC) (McCullagh & Nelder 1989) was used in a bidirectional stepwise procedure to select the most parsimonious model. Implementing this procedure through the S-Plus step.gam function (Venables & Ripley 1999) enabled the objective selection of significant environmental variables and avoided multicollinearity problems in the development of models. The final GAM was of the form:

$$\text{abundance} = \exp\left(\alpha + \sum_{i=1}^n \beta(X_i) + \sum_{j=1}^k f(X_j) + \varepsilon\right)$$

where α is the intercept of the linear predictor, β is the slope of the i th linear component, f is the spline function of the j th smoothed component, and ε is a Gaussian error term (Hastie & Tibshirani 1990).

Originally, the AIC was seen as an objective criterion not only for the selection of independent variables, but also to choose the degrees of freedom (df) for the spline function. However, the resultant models had unjustifiably high degrees of freedom associated with only minimal deviation of the overall trend from spline functions with a lesser degree of freedom. The phenomenon was explained by Hilborn & Mangel (1997), who described the tendency of the AIC to overfit the model beyond the complexity necessary to describe the relationship between the independent variable and its partial effect when the true error distribution was more contagious (contained higher frequencies of small and large values) than expected for a Poisson distribution. To avoid overfitting, the final model was scrutinized by reducing the spline complexity (degrees of freedom for each bicubic spline smoother; Hastie & Tibshirani 1990). If the reduction in the degrees of freedom for the individual splines subjectively changed the overall partial relationship in the new model, the original complexity of the splines was maintained; otherwise the spline complexity was reduced and tested again.

Independent validation and transferability. The flexibility of splines enables GAMs to realistically portray complex relationships between fish and their environment, but this flexibility is also their weakness. When the exact error structure is unknown, the predicted means become susceptible to the effects of outliers. The validity of each model was tested by confronting the model with independent data from

other estuaries (foreign data) to evaluate its predictive utility in other systems. The expected mean abundance for each of the 3 survey data sets was predicted from the environmental conditions by all 3 models developed in this study. The summed deviances between observed and predicted values were used to calculate the explained deviance for all model and estuary combinations, to give an independent measure of the accuracy and the interestuarine transferability of each of the models. Systematic bias in the transferability of models was examined by studying the correlation between the sample predictions from each of the 3 models for a particular estuary.

Spatial comparisons. In contrast to models, maps are static reproductions of a state at a given time. As such, they cannot account for or predict changes in the spatial distribution over time, and hence they underestimate the fluctuations in the distribution over time. This makes them unsuitable as a management tool, or a way of statistically comparing models, however, they represent a useful visual assessment of the ability of the models to capture interannual and intra-annual changes in juvenile seatrout abundance. To examine the spatial and temporal dynamics in years and in months over years, surfaces of the localized, model-predicted abundances were interpolated by inverse distance weighting (IDW) from all collections made in one year and all collections made in May, July, and September in all 4 years (the start, middle, and end of the main recruitment season, respectively). Observed abundance surfaces for the equivalent time periods were also interpolated, but rather than using the IDW method for predictions, observations were kriged (variable radius 12 nearest neighbours) in order to take into account the sample variance associated with observations. The resulting surfaces were standardized by the overall surface mean and variance in order to more easily compare the models and observations in light of differences in the mean abundances and coefficients of variations for the different surfaces.

RESULTS

Three independent GAMs predicting seatrout abundance were developed, each based on the data from one of the 3 Florida estuaries. The variables chosen by the stepwise model selection procedure for inclusion were the same in all models and the degrees of freedom chosen for each spline function were the same and their shapes similar (Figs. 2-4). The models explained 36.64 (IR), 42.54 (CH), and 44.04% (TB) of the deviance in the data from which they were derived (native data) (Table 1).

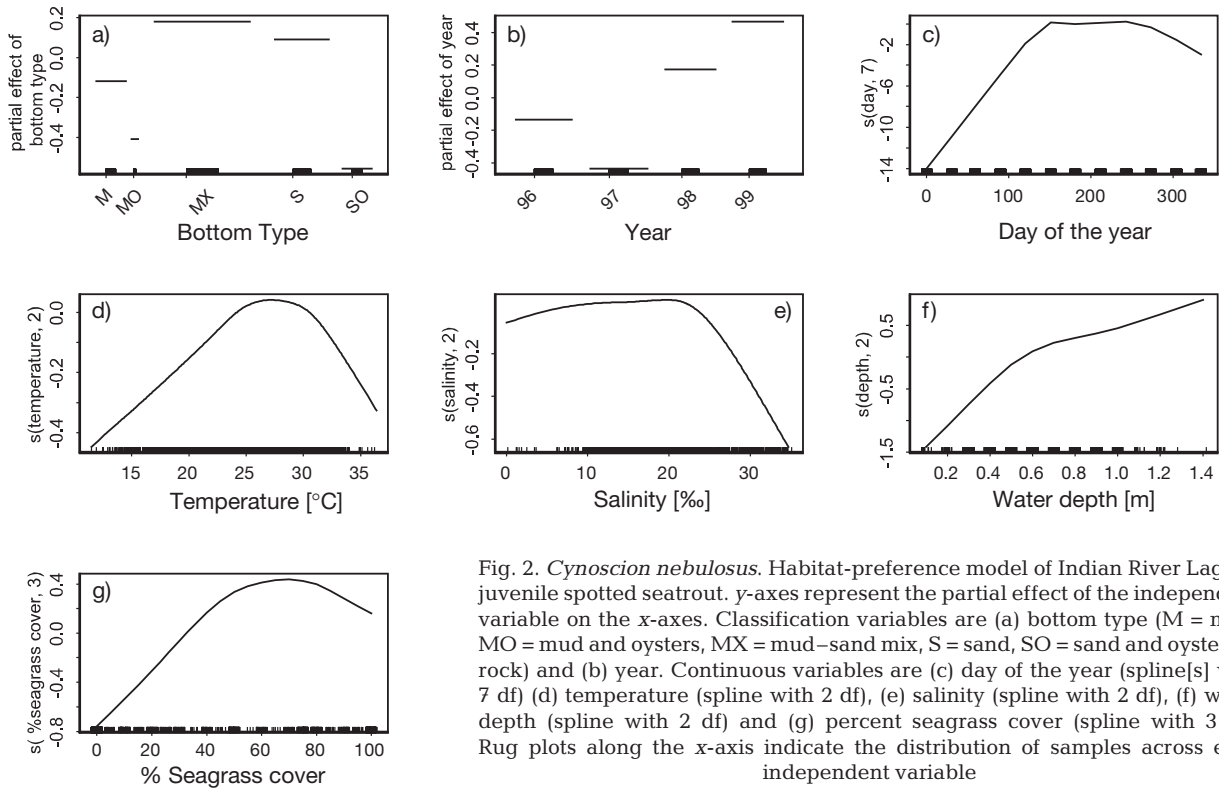


Fig. 2. *Cynoscion nebulosus*. Habitat-preference model of Indian River Lagoon juvenile spotted seatrout. *y*-axes represent the partial effect of the independent variable on the *x*-axes. Classification variables are (a) bottom type (M = mud, MO = mud and oysters, MX = mud–sand mix, S = sand, SO = sand and oysters or rock) and (b) year. Continuous variables are (c) day of the year (spline[s] with 7 df) (d) temperature (spline with 2 df), (e) salinity (spline with 2 df), (f) water depth (spline with 2 df) and (g) percent seagrass cover (spline with 3 df). Rug plots along the *x*-axis indicate the distribution of samples across each independent variable

Environmental variables

The partial effects of temperature indicated an increased abundance of seatrout as water temperatures increased to 28°C for all estuaries, after which abundances declined in the IR (Fig. 2d), continued to increase in TB (Fig. 3d), and remained constant in CH (Fig. 4d). The scale (range on the *y*-axis) of the partial effect, and hence its importance, was smaller in the IR than in TB or CH. The partial effect of salinity implied a unimodal relationship between salinity and seatrout abundance in all estuaries with a maximum of 20 ppt (IR, 20 ‰; TB, 17 ‰; CH, 18 ‰; Figs. 2e, 3e & 4e). As for temperature, the scale of the salinity effect was smaller in the IR than in the other 2 estuaries (Figs. 2e, 3e & 4e). The partial effect of depth was unimodal in TB and CH, rising to a maximum near 1m in both estuaries, whereas in the IR, it continued to rise beyond the range of the sampled depths. The scale of the effect of depth was largest in the IR, smaller in TB, and smallest in CH (Figs. 2f, 3f & 4f). The partial effect of percent seagrass cover rose from 0 to about 60 % cover in all estuaries,

after which it declined in the IR and CH and remained constant in TB. The percent-seagrass-cover partial was comparable in size in all models (Figs. 2g, 3g & 4g). Of the frequently sampled bottom types (sand, sand–mud mix, mud), the sand–mud mix was the most productive

Table 1. Catch statistics, model performance and transferability for the environmental dataset from each of the 3 study estuaries. Numbers in bold indicate characteristics for foreign models

Characteristic	Indian River	Tampa Bay	Charlotte Harbor
Data			
Number of samples	1603	2605	1862
Sum of fish caught in all samples	1465	2556	1593
Mean catch (fish haul ⁻¹)	0.91	0.98	0.86
% occurrence	19.15	21.00	19.33
Total deviation	6606.36	10409.35	6969.47
Model			
Indian River model			
Mean predictions (fish haul ⁻¹)	0.91	0.62	0.76
Residual deviance	4185.61	7479.90	5851.97
% deviance explained	36.64	28.14	15.69
Tampa Bay model			
Mean predictions (fish haul ⁻¹)	1.88	0.98	1.35
Residual deviance	6051.40	5824.85	5616.98
% deviance explained	8.40	44.04	19.41
Charlotte Harbor model			
Mean predictions (fish haul ⁻¹)	1.55	0.63	0.86
Residual deviance	6080.55	7932.81	4004.84
% deviance explained	7.96	23.79	42.54

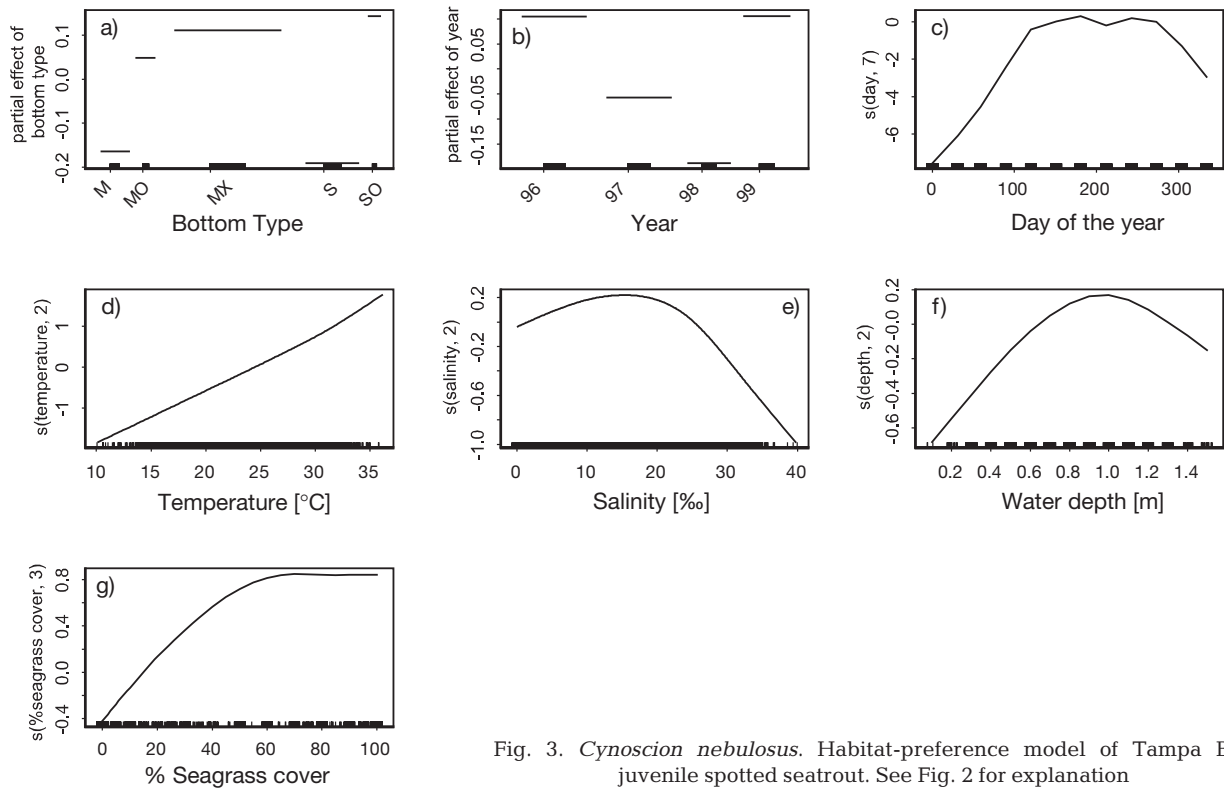


Fig. 3. *Cynoscion nebulosus*. Habitat-preference model of Tampa Bay juvenile spotted seatrout. See Fig. 2 for explanation

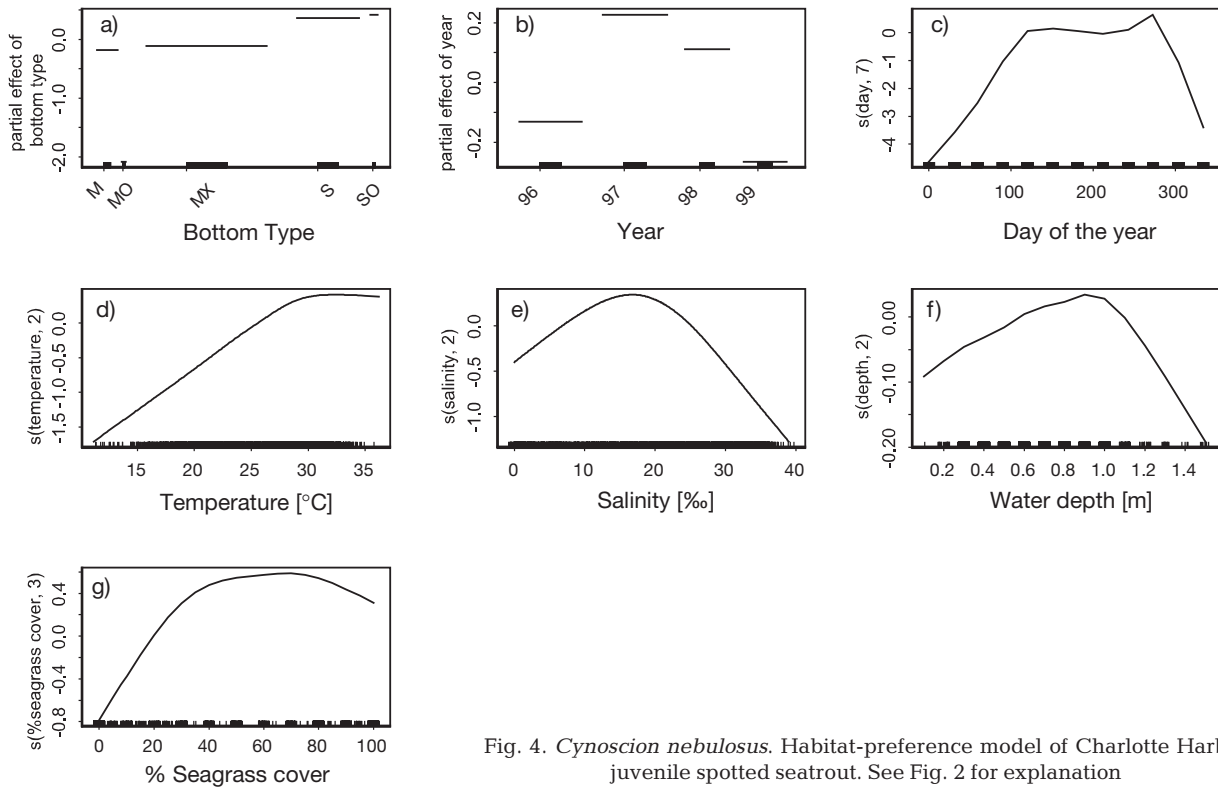


Fig. 4. *Cynoscion nebulosus*. Habitat-preference model of Charlotte Harbor juvenile spotted seatrout. See Fig. 2 for explanation

bottom type according to both the IR and TB models, whereas sand was the most productive bottom type in CH (Figs. 2a, 3a & 4a). The effect of bottom types containing rock or oysters, irrespective of the presence of mud or sand, fluctuated widely between models. These differences may indicate behavioural differences between populations, but it is likely that the overall rarity of the latter 2 substrates and the differences in their prevalence in each estuary account for the disparity. The relative importance of bottom type was significant, but small in all models.

Temporal variables

In all models, the partial effect of day (January 1 = 1 through December 31 = 365) implied a general seasonal increase in abundance of seatrout from May through August (Figs. 2c, 3c & 4c). The timing and exact form of the effect however was different between the models. The IR model indicated a plateau of greatest abundance lasting from Day 125 to Day 250 of the year, with abundances declining sharply outside this period (Fig. 2c). CH and TB models both indicated longer periods of maximum abundance, lasting from Day 100 to Day 280 (Fig. 3c & 4c). The effect of day was bimodal in both estuaries, with local maxima around Day 180 and Day 230, but in CH, the second peak was larger than the first, whereas in TB the peaks were symmetrical. The day partial represented the most important variable describing the abundance of seatrout in each model, with the greatest importance implied for the IR model.

The interannual patterns of abundance were specific to each estuary. IR abundances declined from 1996 to 1997, but have risen since (Fig. 2b), whereas TB seatrout abundances continued to decline until 1998, recovering in 1999 to levels seen in 1996 (Fig. 3b). CH abundances increased from 1996 to 1997 but have declined since (Fig. 4b). The partial effect of year was small, yet significant, in all estuaries.

Spatial comparison between catches and predictions

Distinct interannual differences in the spatial distribution of seatrout in each estuary were predicted by the models derived from their respective data (native models), as indicated by the interpolated surfaces produced from observed and predicted point data (Figs. 5–7). Predictions matched not only the spatial pattern in the relative abundance but also the actual magnitude of mean catches (e.g., TB 1996 to 1999: Fig. 5). Intra-annual variations in the spatial distribution of seatrout within CH (Fig. 6) and TB were also

obvious. In CH and TB, these spatial dynamics could largely be explained by native and foreign models as caused by monthly changes in the environmental conditions, although the CH model consistently overpredicted the abundance of seatrout in the lower part of estuary (Fig. 6), due to an insufficient sampling density associated with the changes in sampling design to Zone D. Despite good interannual correlation between the observed and predicted abundances in the IR, monthly spatial comparisons were less consistent. The IR model demonstrated poor predictive ability, particularly in the Banana River region (Fig. 7a,b), indicating that there was some spatially systematic bias not explained by the environmental conditions included in the model. This part of the estuary is a wildlife refuge, and the salinity regime is strongly influenced by the presence of the Canaveral locks in this area resulting in unusual salinity fluctuations, so that the discrepancies are likely a reflection of the unusual conditions experienced in the area.

Validation and transferability

Seatrout abundances predicted for each of the estuaries by their respective native model were significantly correlated ($\alpha = 0.05$) with abundances predicted for the same samples by foreign models (Fig. 8, Table 2). Although the IR model fared worst in explaining deviance in the native data, it transferred best between estuaries, explaining 28.14 and 15.69% of the deviance in the TB and CH data, respectively (Table 1). The CH model only accounted for 7.96% of the deviance in the IR data and 23.79% of the deviance in the TB data. The TB model produced the worst fit for foreign data (IR data 8.40%; CH data 19.41%), but explained the largest amount of deviance of any model for its native data (44.04%).

Comparisons of the standardized kriged model predictions by year and month showed spatially consistent distribution patterns between models (Fig. 7). The model prediction surfaces, although not dissimilar to surfaces derived from observed data from each estu-

Table 2. Pearson's correlation coefficients for all combinations of model to model comparisons of abundance predictions for each sample in each of the 3 datasets from the 3 Florida estuaries. IR: Indian River; TB: Tampa Bay; CH: Charlotte Harbor.

All correlations were significant at the $\alpha = 0.05$ level

	IR-data	TB-data	CH-data
IR-model \times CH-model	0.515	0.589	0.564
IR-model \times TB-model	0.674	0.767	0.751
TB-model \times CH-model	0.609	0.700	0.608

Interpolated relative abundance in standard deviation units

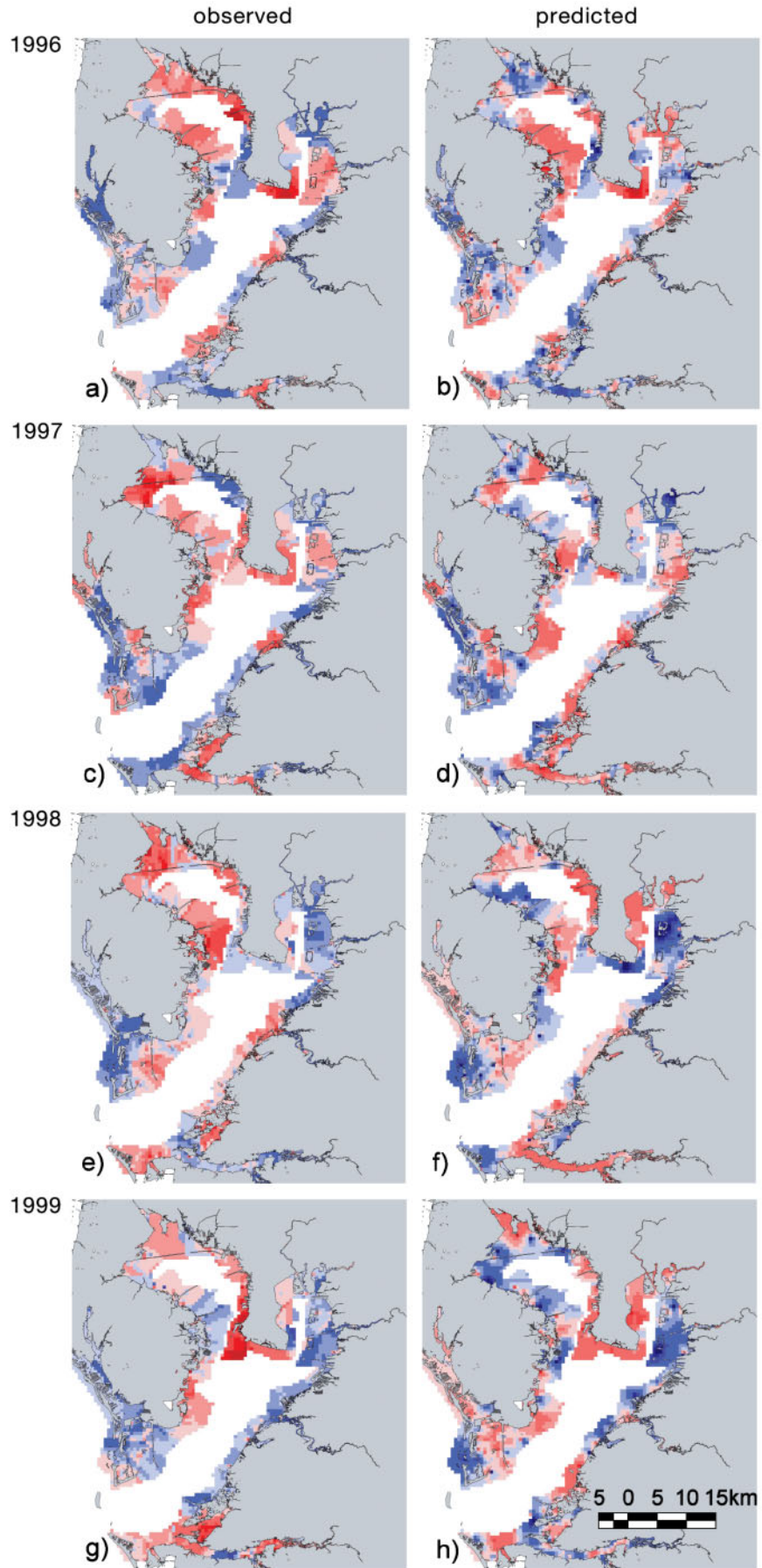
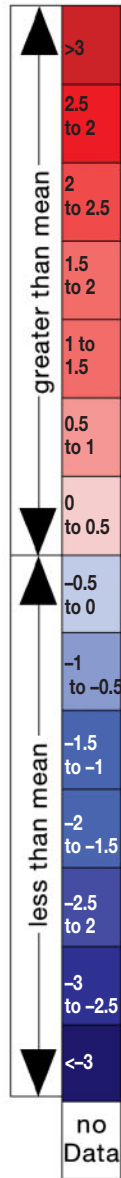


Fig. 5. *Cynoscion nebulosus*. Tampa Bay standardized relative abundance surfaces interpolated from predicted (inverse distance weighting) and observed (kriged) annual point sample abundances and subsequently standardized by mean and variance, indicating interannual differences in the distribution accounted for by the TB model: (a) observed 1996, (b) predicted 1996, (c) observed 1997, (d) predicted 1997, (e) observed 1998, (f) predicted 1998, (g) observed 1999 and (h) predicted 1999

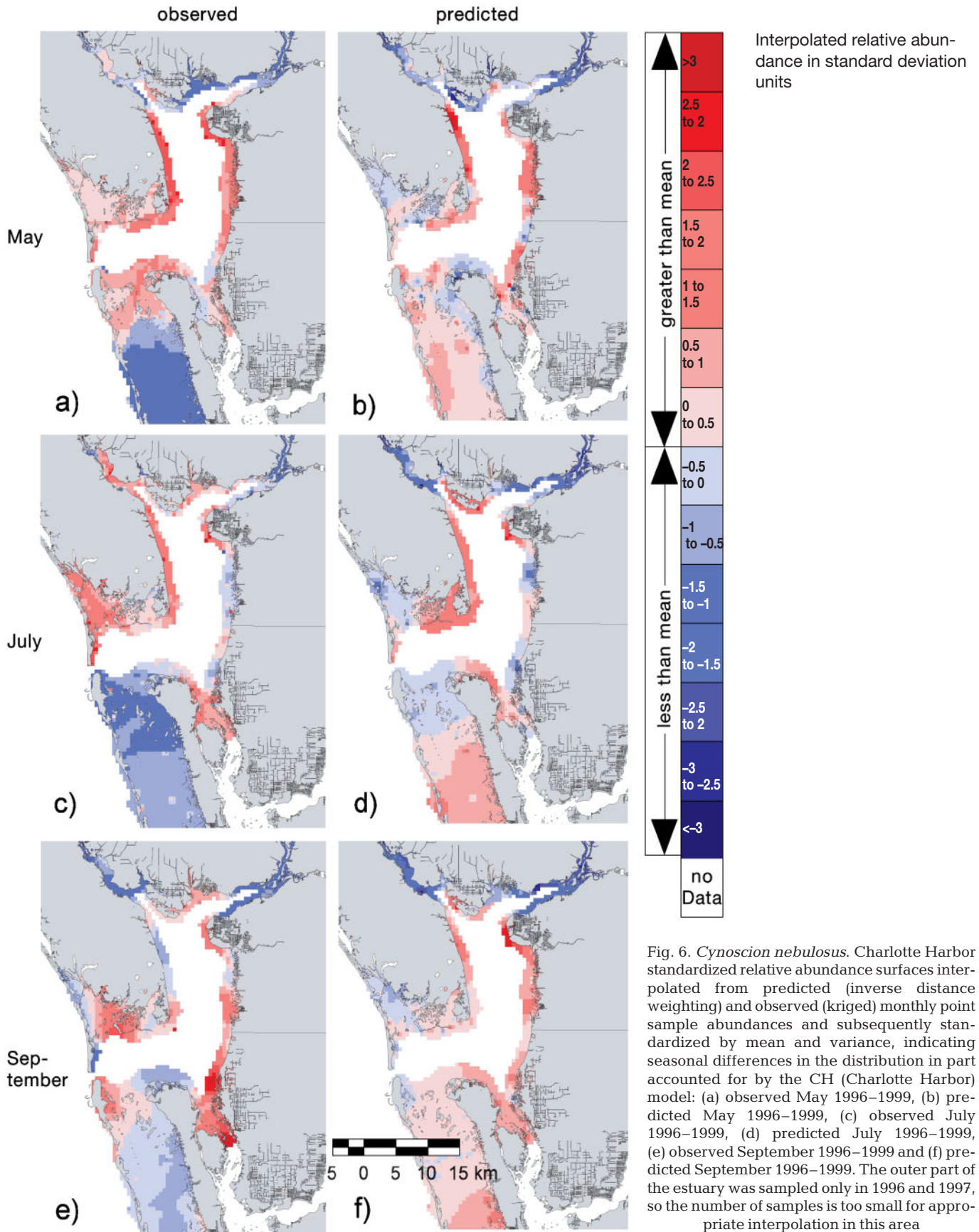


Fig. 6. *Cynoscion nebulosus*. Charlotte Harbor standardized relative abundance surfaces interpolated from predicted (inverse distance weighting) and observed (kriged) monthly point sample abundances and subsequently standardized by mean and variance, indicating seasonal differences in the distribution in part accounted for by the CH (Charlotte Harbor) model: (a) observed May 1996–1999, (b) predicted May 1996–1999, (c) observed July 1996–1999, (d) predicted July 1996–1999, (e) observed September 1996–1999 and (f) predicted September 1996–1999. The outer part of the estuary was sampled only in 1996 and 1997, so the number of samples is too small for appropriate interpolation in this area

Interpolated relative abundance in standard deviation units

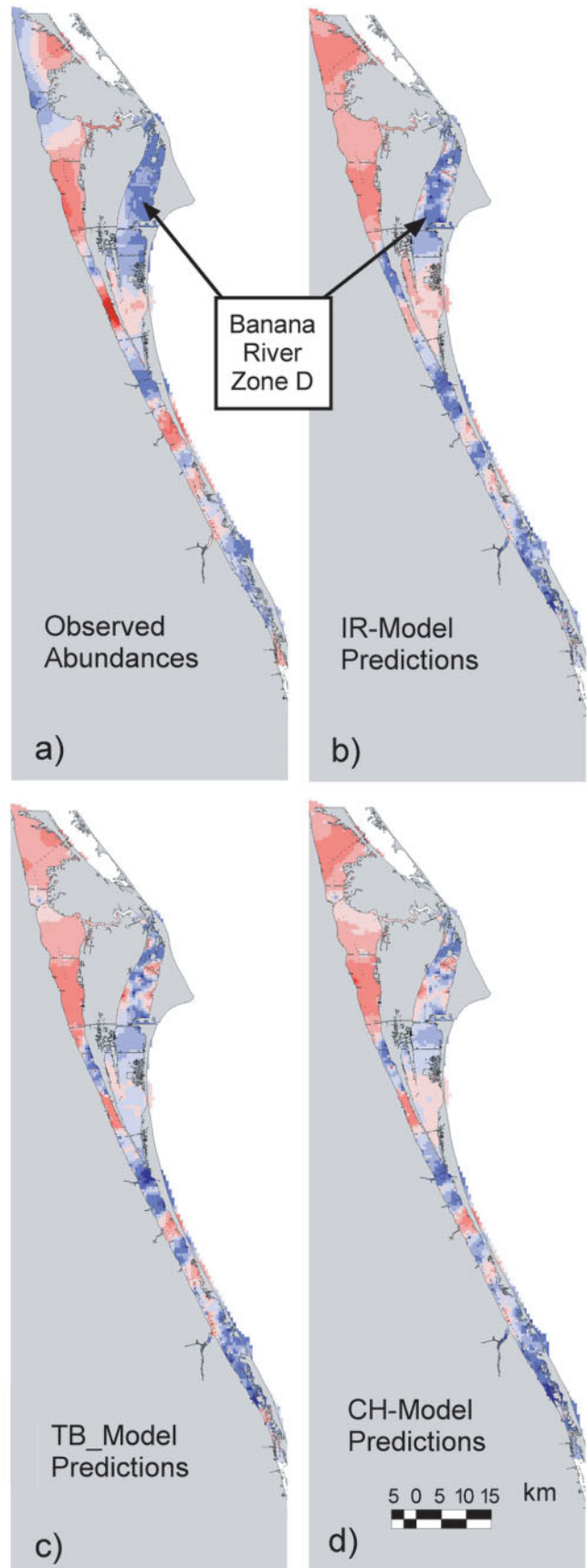
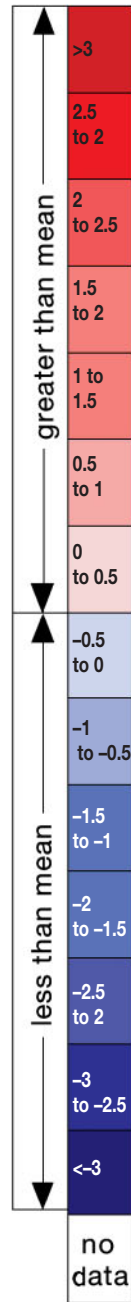


Fig. 7. *Cynoscion nebulosus*. Indian River Lagoon standardized relative abundance surfaces interpolated from the IR (Indian River), TB (Tampa Bay) and CH (Charlotte Harbor) seatrout abundance model predictions (inverse distance weighting) and observed (kriged) 1996 point sample abundances and subsequently standardized by mean and variance. (a) IR observed abundances, (b) IR model predictions, (c) TB model predictions and (d) CH model predictions

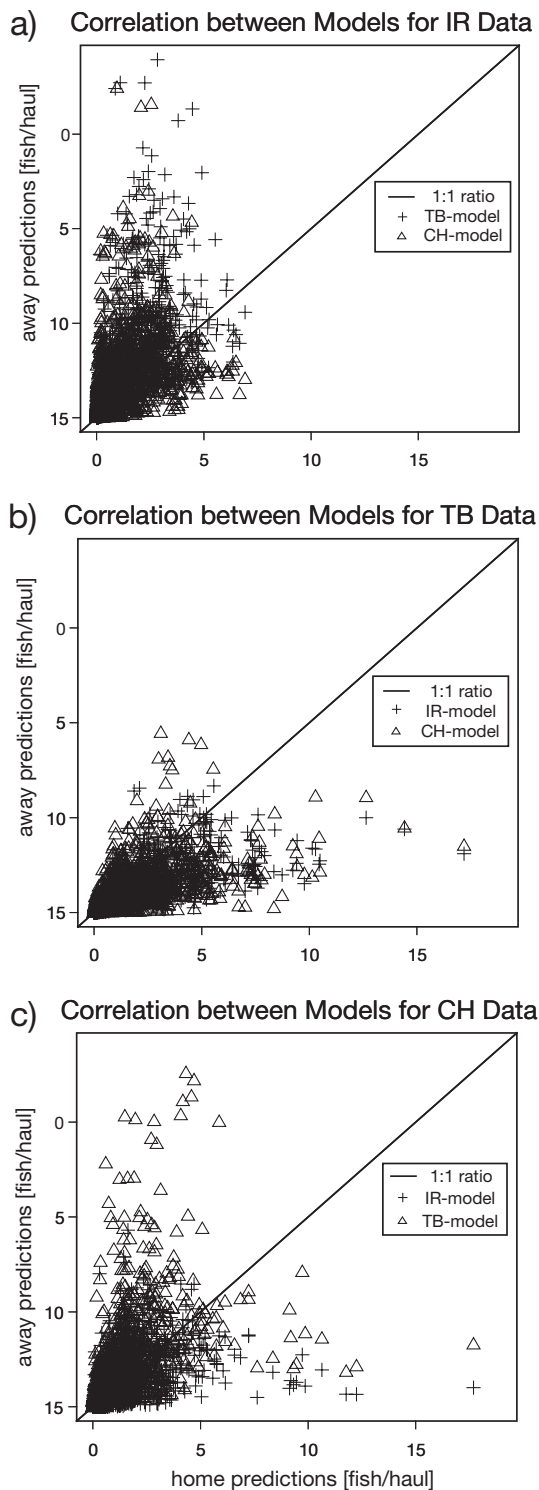


Fig. 8. *Cynoscion nebulosus*. Correlation between the abundances predicted by foreign models with those predicted by the native model for the datasets from (a) Indian River (IR), (b) Tampa Bay (TB) and (c) Charlotte Harbor (CH). Predictions of the native model are always plotted on the x-axis, with the y-axis representing estuary foreign models. The line indicates the expected 1:1 relationship if transferability on the absolute scale were applicable

ary, usually resembled each other more than they resembled the observed data. As with the native model, the foreign models were inconsistent predictors of the spatial distribution of seatrout in the Banana River region of the IR (Fig. 7c,d), again confirming the unusual status of this area.

DISCUSSION

Models of seatrout habitat preferences independently developed for 3 Florida populations showed that juvenile spotted seatrout responded similarly to salinity, temperature, depth, and percent seagrass cover. Habitat preferences were alike for all 3 estuary populations and were similar to documented habitat preferences of this species. The GAMs described the relationships between species abundance and environment and supplied a robust measure of the inter-annual variation in abundance. However, based on published information of the biology and detailed examination of the sampling designs used in this study, it was clear that the models were susceptible to misinterpretation when independent variables were correlated. The predicted spatial distribution of juveniles based on the spatial configuration of habitat characteristics in the estuaries closely matched the observed distribution of catches. What follows is a detailed review of biological inferences from the models and how the results compare with previously reported findings for this species. This review is critical in order to aid in model interpretation, assess transferability, and judge model utility in conserving and managing fish stocks.

Day was a prominent descriptor of seatrout abundance in all models, despite the protracted spawning period reported by Moody (1950), Tucker & Faulkner (1987), Peebles & Tolley (1988), and McMichael & Peters (1989). The CH and TB models showed a bimodal relationship between day and abundance, which was similar to reports of temporal fluctuations in juvenile abundance in Florida by Rutherford et al. (1989) and McMichael & Peters (1989). Seatrout spawning intensity was reported to be similarly bimodal elsewhere along the Gulf Coast (Hein & Shepard 1979). Presumably, the bimodality in juvenile abundance within a year reflected temporal variations in reproductive activity (Nelson & Leffler 2001). In contrast, the IR model indicated a unimodal seasonal pattern of juvenile abundance, concordant with unimodal reproductive patterns reported from the Indian River Lagoon (Crabtree & Adams 1998). The importance of the intra-annual variation in larval supply, whether caused by variation in reproductive effort or differential survival of larvae, greatly limits the abun-

dance of juveniles, which was reflected in the primary importance of day in all models.

The dominance of day in the model was diminished in CH and TB models that were associated with concurrent increased importance of temperature. In addition to the quantitative differences in the importance of temperature, the relationship between temperature and abundance, at first glance, appeared to differ between populations. The optimum temperature for the species is around 28°C (Wohlschlag & Wakeman 1978), consistent with the predicted abundance increases up to this temperature in all models. Higher temperatures were reported to adversely affect the physiology of this species, leading to stress and eventually death (Vetter 1977), so that declining abundances would be expected at such temperatures. The anticipated decline in abundance was only observed in the IR model, with unrealistic increases in abundance predicted by the CH and TB models. The differences in the importance and shape of the temperature partial between the west coast estuaries and the IR can be explained by differences in the degree of correlation between temperature and day within each estuary. The large latitudinal expanse and shallow depth of the IR (Smith 1987) leads to a greater temporal variability in temperatures, resulting in less of a correlation between temperature and the day of the year in the IR than in the other bay systems. Therefore, the IR model can correctly attribute the deviance to day, while the CH and TB models associate some of the deviance due to day with the temperature effect. The predicted high abundances at high temperatures are scaled back by an excessive dip in the day effect during the summer period when temperatures are above 28°C. Multicollinearity between day and temperature, rather than differences in the biology of the populations, was the most likely cause for the difference in the day and temperature effect, particularly given the evidence of fatalities at high temperatures.

Percent seagrass cover was positively correlated to seatrout abundance in all models, second only to day in influence on seatrout distribution after accounting for the collinearity between temperature and day. The prominence attributed to seagrass cover in these models was consistent with reports by McMichael & Peters (1989), Chester & Thayer (1990), and Gilmore (1977). Despite the predicted high-ranking importance of seagrass, juvenile seatrout have also been frequently encountered in areas devoid of seagrass, usually in protected backwaters and tributaries over muddy bottom (Peebles & Tolley 1988, Peebles et al. 1991, Llansó et al. 1998). Even estuaries apparently lacking seagrass altogether have sustained viable populations of seatrout (Darnell 1958). This presents evidence that seagrass itself is not a facultative requirement of

seatrout but rather that percent seagrass cover served as an indirect measure of the ability to find food (Orth et al. 1984) and avoid predators (Hindell et al. 2000) because of habitat complexity. This function may be facilitated in other estuaries by shallow backwaters (Ruiz et al. 1993, Halpin 2000, Paterson & Whitfield 2000). Quantitative measures of food availability and protection from predators offered by different habitats would be necessary in order to accurately predict the distribution of seatrout in estuaries lacking seagrass.

All 3 models largely concur on the shape and the importance of the salinity partial, indicating maximal seatrout abundances at 20 ppt. This value is also close to the physiological optimum of this species (Wohlschlag & Wakeman 1978). The divergence of the IR model from the other estuaries at low salinities results from the lack of sampling in the suboptimal riverine habitats (Peebles et al. 1991) in the IR that are frequently sampled in TB and CH. The few IR samples taken in low-salinity environments are more representative of short-term conditions incidental to sampling and not true long-term oligohaline conditions. Consequently, the IR model overpredicts the abundance of seatrout in mesohaline environments when compared to the other models. In contrast, the CH and TB models correctly predicted few seatrout at low salinities, but unfortunately, the low-salinity river sites in these estuaries also yielded samples from the deepest water depths because of the boat-set technique employed solely in this habitat. Inevitably, this led to the low abundance associated with low salinities also being associated with increasing depth in the west coast models. In contrast, the depth partial in the IR, where rivers were not sampled, continued to increase for the entire length of the depth gradient as reported elsewhere (Rutherford et al. 1989). It is likely that the observed differences between models in the salinity and depth effects were again a result of sampling design induced multicollinearity and not behavioural divergence of the populations. The high degree of correlation of salinity and depth in the CH and TB data, the ecological evidence, and the transferability performance of the IR model all attested to the greater plausibility of the latter model and demonstrated that multicollinearity is hampering the effectiveness of the CH and TB models when predicting abundances in the IR data.

Transferability

Transferability, in this case, is the ability of a model derived from data in one estuary to predict the abundance of seatrout in a second estuary based solely on measured environmental conditions in the latter estu-

ary. The developed models were able to achieve reasonable transferability when confronted with non-native environmental data, explaining significant portions of the total deviance in the data from other estuaries. Observed abundances were significantly correlated to predicted abundances for all foreign models, although the relationship was temporally biased and the overall relationship between model predictions always differed from the expected one-to-one relationship.

The 3 models implied 3 distinct estuary-specific year effects, indicating 3 independent populations, each with its own specific population dynamics. This type of information gleaned from these models has important implications for fisheries managers, who must ultimately consider such effects in producing coast wide stock assessments in order not to allow localized populations to collapse. It might be argued that the sampling-design changes implemented at the end of 1997 invalidate the use of the year effect as a measure representative of estuarine recruitment dynamics. The year effect was smallest in TB, the estuary that has had the most modest sampling design changes, and was largest in the IR, the estuary with the most significant design changes. In CH only small changes in abundance were observed across the sampling design transition—1997 and 1998 data were more similar to each other than to the 1996 and 1999 data. Consequently, differences could not be attributed solely to changes in the sampling design, and abundances must at least in part be controlled by differences in the population dynamics between estuaries. The unique interannual trends meant that models were not useful in predicting the number of seatrout found in an area of a foreign estuary, but could be an invaluable tool in grading the suitability of habitats in a foreign estuary.

The multicollinearity unique to the sampling designs within each estuary represented a hindrance to transferability, as indicated by the better ability of the IR model—the model with the least correlation among independent variables—to explain the most deviance in foreign data sets. Presence of multicollinearity in itself does not seriously impede model predictability, provided that it is constant over the period and area of interest (Neter et al. 1996). In the absence of sampling-design changes, we can assume this is the case within an estuary but certainly not between estuaries, as shown here by the poorer transfer performance of the CH and TB models. It is important to consider the sampling design in light of the species' ecology during the development phase of these complex models and to remember that changes to the design will potentially invalidate the use of the indices of abundance across such changes.

Implications for conservation management

In order to protect and conserve species, conservation managers need information on the resource requirements of all life stages of a species, especially the vulnerable juvenile stages. The models described here allow quantitative evaluation of the important habitats of juvenile seatrout for their native estuaries based on actual environmental data. All models clearly indicated areas of high seatrout abundance (high suitability) as well as areas of low seatrout abundance (low suitability) in non-native estuaries. The suitability of specific estuarine regions was consistent between all models and all estuaries, despite considerable dynamics in space and time. Because the location of optimum conditions (e.g., nursery habitat) varies dynamically in space and time, protection of a specific locale will unlikely suffice to ensure the survival of the species. GAMs can clearly aid conservation managers in reviewing proposed changes to the environment and deciding on options with the least deleterious effects to populations.

For conservation managers it is important to identify the areas of most suitable habitat within an estuary lacking fisheries-independent sampling data. Despite the high residual deviance in these models, it is clear from the maps presented in this paper that, when environmental conditions are sampled in suitable density in space and time, all models were capable of identifying the areas of most suitable habitat in native and foreign estuaries. This justifies the use of low-explanatory-power models by conservation managers. Preference should always be given to those models with least correlation of the independent variables in order to avoid multicollinearity issues during transfer of models, even if it is thought to be at the expense of explanatory power in native estuaries.

Although transferable on a relative scale, the models discussed here were not transferable on the scale of absolute abundances caused by the significant contribution of the population-specific year effect. Therefore, the current models would not be capable of assessing the absolute number of fish affected by environmental alterations or disasters in foreign estuaries. But this might be different for species with a well mixed population or one that recruits from a single-ocean-spawned pool of larval recruits.

Implications for fisheries management

Fisheries managers are interested in obtaining information on the abundance of juveniles before they enter the fishery in order to predict future stock sizes and implement management changes early. In the

past, the most important factor in determining the size of recruitment was thought to be the size of the spawning stock as reflected by the spawning stock biomass (SSB) benchmark currently used for Florida's spotted seatrout stocks. It has, however, become increasingly clear that recruitment can only be poorly linked to SSB, particularly for species restricted to very specific juvenile habitats, because of the variance-dampening effect enforced by its limited availability (Iles & Beverton 2000, van der Veer et al. 2000). A poor spawner–recruit relationship has been observed for Florida spotted seatrout. Therefore, it is expected that the variance-generating function of the reproductive part of the life cycle was overshadowed by the variance dampening function of the other life stages. Assessment of habitat quality and availability should produce better estimates of recruitment than information solely on the spawning stock size. The year effect described by the models represents a robust measure of native juvenile seatrout abundance or future recruitment, assuming a consistent sampling design, stable environmental conditions, and constant niche breadth. Interannual fluctuations were estuary specific, indicating the different seatrout populations are less well mixed than assumed by current stock assessments, and quantitative prediction of recruitment for foreign estuaries is not possible from these models. In contrast to conservation managers, fisheries managers using the models purely in the native estuaries would place heavier emphasis on the explanatory power of the models, rather than their transferability.

Conclusions

The prohibitive costs of intense fisheries-independent surveys has led conservation managers to look for models that can be used to predict fish abundances in systems lacking suitable surveys (Rubec et al. 1999, Brown et al. 2000). This analysis suggests considerable savings can be made by the development of environmentally driven models of species distribution based on modern regression techniques. Generalized additive models are also sure to be invaluable tools for fisheries managers in determining useful recruitment indices for implementation in stock assessments provided that the time series are sufficiently long and uninterrupted by large-scale changes in the sampling design.

The ability of generalized additive models to statistically deal with non-normal data and weighting of different environmental variables as well as the flexibility of splines used to more sensibly model the relationship between abundance and environment means that in data-rich situations these models provide a temporally

and biologically more accurate picture of seatrout distributions than habitat-suitability models (Rubec et al. 1999). However, the susceptibility of the models to multicollinearity issues means special attention needs to be paid to the sampling design under consideration of the biological and ecological characteristics of a species, in order to produce robust distribution models.

Acknowledgements. I thank the field crews in the Indian River Lagoon, Charlotte Harbor and Tampa Bay for collection and processing of samples, and the FWC-FMRI Fisheries Independent Monitoring Program as a whole for making this work possible. The project was supported in part by the funding from the Department of the Interior, US Fish and Wildlife Service, Federal Aid for Sportfish Restoration Project Number F-96 and Florida Saltwater Fishing License Monies.

LITERATURE CITED

- Bigelow KA, Boggs CH, He X (1999) Environmental effects on swordfish and blueshark catch rates in the US North Pacific longline fishery. *Fish Oceanogr* 8:178–198
- Brown SK, Buja KR, Jury SH, Monaco ME (2000) Habitat suitability index models for eight fish and invertebrate species in Casco and Sheepscot Bays, Maine. *N Am J Fish Manage* 20:408–435
- Chester AJ, Thayer GW (1990) Distribution of spotted seatrout (*Cynoscion nebulosus*) and gray snapper (*Lutjanus griseus*) juveniles in seagrass habitats of western Florida Bay. *Bull Mar Sci* 46:345–357
- Crabtree RE, Adams DH (1998) Study V: spawning frequency and fecundity of spotted seatrout, *Cynoscion nebulosus*, in the Indian River Lagoon, Florida. In: Investigations into nearshore and estuarine gamefish abundance, ecology, and life history in Florida (Project F-59). Florida Fish and Wildlife Conservation Commission, Florida Marine Research Institute (FMRI) technical report to the US Department of Interior Fish and Wildlife Service. FMRI, St. Petersburg, FL, p 526–566
- DaCosta SL, Reichard RP, Bhogal VK (1987) Indian River Lagoon circulation and flushing study. Report to Florida Sea Grant College (R/IRL-3) and Florida Department of Environmental Regulation (CM 118). Florida Institute of Technology, Melbourne, FL
- Darnell RM (1958). Food habits of fishes and larger invertebrates of lake Ponchartrain, Louisiana, an estuarine community. *Publ Inst Mar Sci* 5:353–416
- Geary RC, Leser CEV (1968) Significance tests in multiple regression. *Am Stat* 22:20–21
- Gilmore RG (1977) Fishes of the Indian River Lagoon and adjacent waters, Florida. *Bull Fla St Mus Biol Sci* 22(3) 101–147
- Halpin PM (2000) Habitat use by an intertidal salt-marsh fish: trade-offs between predation and growth. *Mar Ecol Prog Ser* 198:203–214
- Hamilton D (1987) Sometimes $R^2 > r^2_{yx1} + r^2_{yx2}$. *Am Stat* 42:90–91
- Hastie T, Tibshirani R (1990) Generalized additive models. Chapman & Hall, London
- Hein S, Shepard J (1979) Spawning of spotted seatrout in a Louisiana estuarine ecosystem. *Proc Annu Conf Southeast Assoc Fish Wildl Agencies* 33:451–465
- Helser TE, Condrey RE, Geaghan JP (1993) Spotted seatrout

- distribution in four coastal Louisiana estuaries. *Trans Am Fish Soc* 122:99–111
- Hilborn R, Mangel M (1997) *The ecological detective, confronting models with data*. Princeton University Press, Princeton, NJ
- Hindell JS, Jenkins GP, Keough MJ (2000) Variability in the abundance of fishes associated with seagrass habitats in relation to diets of predatory fishes. *Mar Biol* 136:725–737
- Iles TC, Beverton RJH (2000) The concentration hypothesis: the statistical evidence. *ICES J Mar Sci* 57:216–227
- Iversen ES, Tabb DC (1962) Subpopulations based on growth and tagging studies of spotted seatrout *Cynoscion nebulosus*, in Florida. *Copeia* 1962:544–548
- Kupschus S, Tremain D (2001) Associations between fish assemblages and environmental factors in nearshore habitats of a subtropical estuary. *J Fish Biol* 58:1383–1403
- Llansó RJ, Bell SS, Vose FE (1998) Food habits of red drum and spotted seatrout in a restored mangrove impoundment. *Estuaries* 21:294–306
- Maravelias CD, Reid DG, Schwarzman G (2000) Modelling spatio-temporal effects of environment on Atlantic herring, *Clupea harengus*. *Environ Biol Fish* 58:157–172
- McCullagh P, Nelder JA (1989) *Generalized linear models*, 2nd edn. Chapman and Hall, London
- McMichael RH, Peters KM (1989) Early life history of spotted seatrout, *Cynoscion nebulosus* (Pisces: Sciaenidae), in Tampa Bay, Florida. *Estuaries* 16:703–717
- Moody WD (1950) A study of the natural history of the spotted trout, *Cynoscion nebulosus*, in the Cedar Key, Florida, area. *Q J Fla Acad Sci* 12:147–171
- Moffet AW (1961) Movement and growth of spotted seatrout (*Cynoscion nebulosus*) in West Florida. *Fla Board Conserv Mar Res Lab Tech Ser* 36
- Nelson GA (2002) Age, growth, mortality, and distribution of pinfish (*Lagodon rhomboides*) in Tampa Bay and adjacent Gulf of Mexico waters. *Fish Bull* 100:582–592
- Nelson GA, Leffler D (2001) Abundance, spatial distribution, and mortality of young-of-the-year spotted seatrout (*Cynoscion nebulosus*) along the Gulf coast of Florida. *Gulf Mex Sci* 19:30–42
- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied linear statistics*, 4th edn. WCB/McGraw-Hill, Boston, MA
- Orth RJ, Heck KL, van Montfrans J (1984) Faunal communities in seagrass beds: a review of the influences of plant structure and prey characteristics on predator–prey relationships. *Estuaries* 7:339–350
- Paterson AW, Whitfield AK (2000) Do shallow-water habitats function as refugia for juvenile fishes? *Estuar Coast Shelf Sci* 51:359–364
- Peebles EB, Tolley SG (1988) Distribution and mortality of larval spotted seatrout, *Cynoscion nebulosus*: a comparison between two adjacent estuarine areas of southwest Florida. *Bull Mar Sci* 42:397–410
- Peebles EB, Flannery MS, Matheson RE, Rast JP (1991) Fish nursery utilization of the Little Manatee River estuary: relationships to physicochemical gradients and the distribution of food resources. In: *Proc Tampa Bay Area Sci Inform Symp Vol 2*. Tampa Bay Regional Planning Council, Tampa, FL, p 341–367
- Rubec PJ, Bexley JC, Norris H, Coyne MS, Monaco ME, Smith SG, Ault JS (1999) Suitability modeling to delineate habitat essential to sustainable fisheries. *Am Fish Soc Symp* 22:108–133
- Ruiz GM, Hines AH, Posey MH (1993) Shallow water as a refuge habitat for fish and crustaceans in non-vegetated estuaries: an example from Chesapeake Bay. *Mar Ecol Prog Ser* 99:1–16
- Rutherford ES, Thomas WS, Tilmant JT (1989) Early life history of spotted seatrout (*Cynoscion nebulosus*) and gray snapper (*Lutjanus griseus*) in Florida Bay, Everglades national park, Florida. *Bull Mar Sci* 44:49–64
- Schwarzman G, Stuetzle W, Kulman K, Powojowski M (1994) Relating the distribution of pollock schools in the Bering Sea to environmental factors. *ICES J Mar Sci* 51:481–492
- Smith NP (1987) An introduction to the tides of Florida's Indian River Lagoon. I. Water levels. *Fla Sci* 50:49–61
- Tabb DC (1961) A contribution to the biology of the spotted seatrout *Cynoscion nebulosus* (Cuvier), of east-central Florida. *Fla Board Conserv Mar Res Lab Tech Ser* 35
- Tabb DC (1966) The estuary as a habitat for spotted seatrout, (*Cynoscion nebulosus*). *Am Fish Soc Spec Publ* 3:59–67
- Tucker JW, Faulkner BE (1987) Voluntary spawning patterns of captive spotted seatrout. *Northeast Gulf Sci* 9:59–63
- van der Veer HW, Berghahn R, Miller JM, Rijnsdrop AD (2000) Recruitment in flatfish, with special emphasis on North Atlantic species: progress made by the Flatfish Symposia. *ICES J Mar Sci* 57:202–215
- Venables WN, Ripley BD (1999) *Modern applied statistics with S-PLUS*, 3rd edn. Springer-Verlag, New York
- Vetter RB (1977) Respiratory metabolism of, and niche-separation between two co-occurring congeneric species, *Cynoscion nebulosus* and *Cynoscion arenarius*, in a south Texas estuary. MSc thesis, University of Texas at Austin, Austin, TX
- Wohlschlag DE, Wakeman JM (1978) Salinity stresses, metabolic responses and distribution of the coastal spotted seatrout, *Cynoscion nebulosus*. *Contrib Mar Sci* 21:171–185
- Yeo GK (1984) A note of caution on using statistical software. *The Statistician* 33:181–184

Editorial responsibility: Otto Kinne (Editor), Oldendorf/Luhe, Germany

*Submitted: December 10, 2002; Accepted: September 23, 2003
Proofs received from author(s): December 19, 2003*