Using delta generalized additive models to produce distribution maps for spatially explicit ecosystem models☆

Arnaud Grüss [a,b,*,1], Michael Drexler [c,1], Cameron H. Ainsworth [c]

[a] Southeast Fisheries Science Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, FL 33149-1099, USA
[b] University of Miami, Rosenstiel School of Marine and Atmospheric Science, Cooperative Institute for Marine and Atmospheric Studies, 4600 Rickenbacker Causeway, Miami, FL 33149, USA
[c] University of South Florida, College of Marine Science, 140 7th Avenue South, St. Petersburg, FL 33701, USA

ABSTRACT

Spatial ecosystem models, such as OSMOSE, have become integral tools in achieving ecosystem-based management for their ability to thoroughly describe predator–prey dynamics in a spatially explicit context. Distribution maps, which define the initial spatial allocation of functional groups abundance, can have a large effect on the predator–prey dynamics that spatially explicit ecosystem models simulate. Here, we introduce the delta GAM approach we developed to be able to produce distribution maps for an OSMOSE model of the West Florida Shelf (Gulf of Mexico), OSMOSE-WFS. This delta GAM approach predicts the spatial distribution of different life stages of the multiple functional groups represented in OSMOSE-WFS ('life-stage groups') at different seasons, over the entire Gulf of Mexico (GOM) shelf including areas where abundance estimates do not exist, using different research survey datasets and regional environmental and habitat features. Our delta GAM approach consists of fitting two independent models, a binomial GAM and a quasi-Poisson GAM, whose predictions are then combined using the delta method to yield spatial abundance estimates. To validate delta GAMs, bootstraps are used and Spearman's correlation coefficients (Spearman's ρ's) between predicted and observed abundance values are estimated and tested to be significantly different from zero. We use pink shrimp (Farfantepenaeus duorarum) to demonstrate our delta GAM approach by predicting the summer distribution of this species over the GOM shelf and the West Florida Shelf. Predictions of the delta GAM reflect existing empirical research related to pink shrimp habitat preferences and predictions of a negative binomial GAM previously designed for the GOM. We find that using a delta rather than a negative binomial GAM saves significant computation time at the expense of a slight reduction in GAM performance. A positive and highly significant Spearman's ρ between observed and predicted abundance values indicates that our delta GAM can reliably be used to predict pink shrimp spatial distribution. Spearman's ρ was also positive and highly significant in every life-stage group represented in OSMOSE-WFS and season, though often low. Therefore, delta GAMs fitted for the different life-stage groups and seasons correctly predict qualitative differences between low- and high-abundance areas and are deemed appropriate for generating distribution maps for OSMOSE-WFS. The delta GAM approach we developed is a simple, convenient method to create distribution maps to be fed into spatially explicit ecosystem models, where wide spatial and taxonomic coverage is desired while benefits of high precision estimates are lost at run-time.

Published by Elsevier B.V.

1. Introduction

Ecosystem-based management (EBM) of marine systems has become a central paradigm throughout the world (Pikitch et al., 2004; Garcia and Cochrane, 2005; McLeod and Leslie, 2009). Spatial ecosystem models, such as Ecospace (Christensen and Walters, 2004; Walters et al., 2010), Atlantis (Fulton et al., 2004, 2007) and OSMOSE (Shin and Cury, 2001, 2004), have become integral tools in achieving EBM for their ability to thoroughly describe predator–prey dynamics in a spatially explicit context. Within a spatially explicit context, the predation mortality of functional groups (i.e., groups of species sharing similar ecological niches and life-history traits) becomes influenced by their degree of spatial overlap with potential predators. Therefore, distribution maps, which define the spatial allocation of functional groups abundance or biomass at the start of simulations or at the beginning of each time step, can have a large effect on the predator–prey dynamics that spatially explicit ecosystem models simulate.

Fisheries management in the Gulf of Mexico (GOM) has recently begun to embrace EBM through the development of a comprehensive Integrated Ecosystem Assessment program (Levin et al., 2009, 2013; Samhouri et al., 2013). Within the GOM Integrated Ecosystem Assessment program, several spatially explicit ecosystem models have been designed, including an Atlantis model for the entire Gulf of Mexico (Atlantis-GOM) and an OSMOSE model for the West Florida Shelf (OSMOSE-WFS), to simulate predator–prey dynamics and the potential impacts of management measures in the GOM. One major challenge for developing Atlantis-GOM and OSMOSE-WFS has been the construction of distribution maps despite the availability of several large fisheries-independent survey databases.

We developed delta generalized additive models (GAMs; Hastie and Tibshirani, 1990; Guisan et al., 2002) for predicting the spatial distribution of different life stages of multiple functional groups over the entire GOM shelf, which can then be used to define distribution maps for the OSMOSE-WFS model. In the following, we: (1) briefly review how distribution maps have been produced and utilized in spatially explicit ecosystem models to date, focusing on the widely used OSMOSE and Atlantis models; (2) discuss the use of GAMs to construct distribution maps for spatial ecosystem models; and (3) give an outline of the content and objectives of the present study.

1.1. Distribution maps in spatially explicit ecosystem models

Spatial ecosystem models usually explicitly consider a large number of functional groups whose spatial distribution in the real world can vary substantially within life stages and between seasons. However, the generation of distribution maps for spatially explicit ecosystem models is usually constrained by the availability of survey data and the total number of samples available. Spatial abundance and biomass data are typically collected for species of high economic importance. Moreover, available spatial data generally cover only a fraction of the total habitat area used by functional groups of interest. Given these limitations, a single, simple framework is desired to produce distribution maps for the multiple functional groups represented in spatial ecosystem models. The spatial distribution of functional groups in an ecosystem model at the start of simulations or at the beginning of each time step constitutes an important step in capturing ecosystem dynamics. Fish movements implemented after this step ultimately affect the predation mortality rates and diet composition of functional groups.

OSMOSE is a two-dimensional, individual-based, multi-species model whose basic units ('super-individuals') are schools, which consist of organisms belonging to the same functional group and the same life stage, which have the same length, weight, and, at a given time step, the same spatial coordinates (Shin and Cury, 2004, 2001). Schools are distributed in space at each time step using a set of maps of presence/absence or density maps created for specific functional groups, life stages and seasons. When the distribution of schools remains static (within a season or if the distribution is constant throughout the year), schools move to immediately adjacent cells within their distribution area following a random walk. Maps of presence/absence used in OSMOSE have generally been produced from the literature and experts' opinion (e.g., Shin et al., 2004; Travers et al., 2009; Brochier et al., 2013) or directly from research surveys and commercial fisheries data (Fu et al., 2013). Marzloff et al. (2009) created density maps for the eight functional groups represented in their OSMOSE model directly from acoustic and trawl survey data.

Atlantis is a sophisticated biogeochemical marine ecosystem model (Fulton et al., 2004, 2007). This model integrates ecological, fisheries, physical and chemical dynamics in a three-dimensional, spatially explicit domain. Distribution maps are employed in Atlantis for defining the allocation of functional groups biomass over space at the start of simulations. These distribution maps are habitat preference maps or density maps, produced directly from research survey data or the literature (e.g., Fulton et al., 2007; Horne et al., 2010; Kaplan et al., 2010). Ainsworth et al. (2011) took a different approach and used a habitat similarity matrix to extrapolate research survey data to the entire spatial footprint of their Atlantis model for the northern Gulf of California.

All the above-mentioned methods used to generate distribution maps for spatial ecosystem models are convenient ways to create maps rapidly. However, these methods limit the spatial coverage of spatially explicit ecosystem models to those areas that are consistently sampled. Furthermore, with the exception of Ainsworth et al. (2011), extrapolations are not an objective process and instead rely heavily on individual opinion.

1.2. Using GAMs to produce distribution maps for spatially explicit ecosystem models

Data-driven statistical models such as GAMs offer a valuable, objective way to predict abundance and biomass over extensive geographical regions. However, the potential of GAMs for producing distribution maps for spatially explicit ecosystem models has not been exploited until recently (Drexler and Ainsworth, 2013). GAMs relate an ecological response to a suite of predictors using non-linear smoothing functions (Hastie and Tibshirani, 1990; Guisan et al., 2002). One major disadvantage of GAMs is that they require a relatively large amount of data to have the high degrees of freedom that guarantee their flexibility (Wood, 2006). One of their major advantages is that they can be used to estimate spatial patterns of abundance or biomass over a broad geographic region *spanning both sampled and unsampled areas* (e.g., Koubbi et al., 2006; Vaz et al., 2006; Loots et al., 2007; Planque et al., 2007).

Drexler and Ainsworth (2013) developed a GAM approach to predict the relative abundance of multiple functional groups in shelf areas across the entire Gulf of Mexico, based on abundance estimates coming from a fisheries independent dataset, the SEAMAP groundfish/trawl dataset (GSMFC, 2011; SEDAR, 2011), and regional environmental and habitat features. There is an overdispersion of abundance estimates in the SEAMAP groundfish/trawl dataset linked to the high number of zero data, as is frequently the case with research survey data and other ecological data (e.g., Barry and Welsh, 2002; Koubbi et al., 2006; Vaz et al., 2006; Loots et al., 2010). To account for this, Drexler and Ainsworth (2013) used a negative binomial GAM approach (Barry and Welsh, 2002; Zeileis et al., 2008). The authors utilized this approach to generate distribution maps for 40 functional groups represented in the Atlantis-GOM model.

A negative binomial GAM approach is similar to a delta GAM approach combining the predictions of a binomial GAM and those of a quasi-Poisson GAM (Welsh et al., 1996; Barry and Welsh, 2002; Zeileis et al., 2008). The negative binomial GAM approach entails the fitting of a single model and the estimation of a shape parameter, θ, which is an index of overdispersion. The delta GAM approach relies on the fitting of two independent models, one model predicting presence/absence and another one estimating the magnitude of the abundance where present (Welsh et al., 1996; Barry and Welsh, 2002). Therefore, the delta GAM approach offers a perspective on spatial patterns of abundance different from that offered by the negative binomial GAM approach, by identifying factors influencing the presence of species and environmental conditions favoring high abundances (Barry and Welsh, 2002; Koubbi et al., 2006).

A major advantage of the delta GAM approach over the negative binomial GAM approach is the computation time needed for model fitting. The estimation of the θ parameter in negative binomial GAMs is an iterative process which takes a relatively long computation time (Barry and Welsh, 2002). The speed at which delta GAMs are fitted is valuable for generating seasonal distribution maps for different life stages of the multiple functional groups represented in spatially explicit ecosystem models (hereafter usually simply referred to as 'life-stage groups'). Even though GAM approaches can be partially automated, attention must be paid to the GAMs fitted for each individual life-stage group and season.

1.3. Content and objectives of the present study

In the present paper, we introduce the parsimonious delta GAM approach we developed for predicting the spatial distribution of multiple life-stage groups over the entire GOM shelf including Mexican and Cuban waters and US areas where abundance and biomass estimates do not exist, using different research survey datasets and regional environmental and habitat features. This approach was primarily developed to allow an initial spatial allocation of life-stage groups at different seasons in the OSMOSE-WFS ecosystem model.

In the following, we: (1) describe the study areas and study functional groups; (2) detail the research survey, environmental and habitat data that were inputted into GAMs; (3) give details of the delta GAM approach we devised for estimating spatial distributions in the shelf areas of the GOM, before evaluating and validating this approach; (4) implement a negative binomial GAM approach in order to make comparisons between our GAM approach and that applied by Drexler and Ainsworth (2013); (5) employ the GAMs we developed to predict spatial distributions over the entire GOM shelf and the West Florida Shelf and to construct distribution maps for the OSMOSE-WFS and Atlantis-GOM ecosystem models; and (6) use pink shrimp (*Farfantepenaeus duorarum*) as a case study to demonstrate our delta GAM approach. We thoroughly analyze the influence of environmental and habitat parameters on the spatial distribution of pink shrimp during the summer season, and predict the summer distribution of the species over the GOM shelf and the West Florida Shelf. Pink shrimp was chosen as an example because this species is very well represented in available research survey data and its spatial distribution is strongly correlated to environmental and habitat features.

2. Materials and methods

2.1. Study areas

The Gulf of Mexico is a Large Marine Ecosystem enveloped by the economic exclusive zones of the United States (US), Cuba and Mexico (NOS, 2008). Five US states border the GOM: Florida,

Alabama, Mississippi, Louisiana and Texas. The continental shelf of the GOM extends to water depths of around 200 m (Becker et al., 2009).

The West Florida Shelf is one of the main subregions of the GOM Large Marine Ecosystem under high and increasing fishing and environmental pressures (Coleman et al., 2004; Okey et al., 2004; Steidinger, 2009; Karnauskas et al., 2013). The region considered in the present paper to generate distribution maps for the OSMOSE-WFS model corresponds to the West Florida Shelf excluding the Florida Keys and Dry Tortugas (Fig. 1).

2.2. Study functional groups

2.2.1. Functional groups represented in OSMOSE-WFS

Functional groups considered in the present study are the twelve high trophic level groups of fish and invertebrate species that are explicitly considered in the OSMOSE-WFS ecosystem model (Table 1). These twelve species groups made a substantial contribution to total biomass and fisheries economy in the West Florida Shelf region during the 2000s, and/or are important components of the West Florida Shelf food web. Species of a given functional group exhibit similar life history traits, body size ranges, diets and exploitation patterns. Some individual species (e.g., gag grouper, *Mycteroperca microlepis*) constitute their own group, as they are emblematic to the West Florida Shelf and of high economic importance (Table 1).

2.2.2. Pink shrimp

Pink shrimp (all life stages) is illustrated in the present paper in detail as a case study. Shrimps are economically some of the most important marine resources in the GOM. The value of their combined landings was approximately $368 million in 2010 (NOAA, 2010). Around 94% of landings of pink shrimp occur in West Florida (NOAA, 2010), where the species is particularly abundant. Pink shrimps are also found in large numbers in the area around the Dry Tortugas, as well as in the northern Gulf of Campeche and especially along the coast of Yucatan (Costello and Allen, 1970; Bielsa et al., 1983).

Hotspots for pink shrimp are usually located between 9 and 44 m of water depth (GMFMC, 2006). Mulholland (1984) derived a habitat suitability index model for pink shrimp. The author reports pink shrimp catches over a large range of temperatures (5–38 °C), with the highest catches realized at temperatures ranging from 20 to 38 °C. Moreover, according to Mulholland's (1984) habitat suitability index model, sandy-silt and silty-sand are the sediment types with the highest suitability for pink shrimp, followed by hard bottom, whereas soft bottom is the sediment type with the lowest suitability for the species.

2.3. Research survey datasets

Three research survey datasets were used for the present study: the SEAMAP (Southeast Area Monitoring and Assessment Program) groundfish/trawl, the NMFS BLL (National Marine Fisheries Service bottom longline) and the SEAMAP reef fish video datasets (Table 2 and Fig. 2). There is an overdispersion of abundance and biomass estimates in these three datasets related to the high number of zero data (Table 3).

Prior to fitting GAMs so as to be able to construct distribution maps for study functional groups, it was necessary to: (1) identify the life-stage groups and seasons for which distribution maps should ideally be produced for the OSMOSE-WFS ecosystem model, from information in the literature; and (2) determine the distribution maps that can be effectively produced, from examination of the number of data points in the SEAMAP groundfish/trawl, NMFS BLL and SEAMAP reef fish video datasets. After consideration of

Table 1
Functional groups explicitly considered in the OSMOSE-WFS ecosystem model. The reference species of each functional group is indicated in bold.

Functional group	Species
King mackerel	**King mackerel (*Scomberomorus cavalla*)**
Amberjacks	**Greater amberjack (*Seriola dumerili*)**, banded rudderfish (*Seriola zonata*), lesser amberjack (*Seriola fasciata*)
Red grouper	**Red grouper (*Epinephelus morio*)**
Gag grouper	**Gag grouper (*Mycteroperca microlepis*)**
Red snapper	**Red snapper (*Lutjanus campechanus*)**
Sardine-herring-scad complex	**Scaled sardine (*Harengula jaguana*)**, Spanish sardine (*Sardinella aurita*), Atlantic thread herring (*Opisthonema oglinum*), round scat (*Decapterus punctatus*)
Anchovies and silversides	**Bay anchovy (*Anchoa mitchilli*)**, striped anchovy (*Anchoa hepsetus*), silversides (Atherinidae spp.), alewife (*Alosa* sp.)
Coastal omnivores	**Pinfish (*Lagodon rhomboides*)**, spottail pinfish (*Diplodus holbrooki*), orange filefish (*Aluterus schoepfii*), fringed filefish (*Monacanthus ciliatus*), planehead filefish (*Monacanthus hispidus*), orangespotted filefish (*Cantherhines pullus*), honeycomb filefish (*Acanthostracion polygonius*), Atlantic spadefish (*Chaetodipterus faber*), scrawled cowfish (*Lactophrys quadricornis*), pufferfish (Tetraodontidae spp.)
Reef carnivores	**White grunt (*Haemulon plumieri*)**, black sea bass (*Centropristis striata*), rock sea bass (*Centropristis philadelphica*), belted sandfish (*Serranus subligarius*), longtail bass (*Hemanthias leptus*), butter hamlet (*Hypoplectus unicolor*), creole fish (*Paranthias furcifer*), slippery dick (*Halichoeres bivittatus*), painted wrasse (*Halichoeres caudalis*), yellowhead wrasse (*Halichoeres garnoti*), bluehead (*Thalassoma bifasciatum*), reef croaker (*Odontoscion dentex*), jackknife-fish (*Equetus lanceatus*), leopard toadfish (*Opsanus pardus*), scopian fish (Scorpaenidae spp.), bigeyes (Priacanthidae spp.), littlehead porgy (*Calamus proridens*), jolthead porgy (*Calamus bajonado*), saucereye progy (*Calamus calamus*), whitebone progy (*Calamus leucosteus*), knobbed progy (*Calamus nodosus*), French grunt (*Haemulon flavolineatum*), Spanish grunt (*Haemulon macrostomum*), margate (*Haemulon album*), bluestriped grunt (*Haemulon sciurus*), striped grunt (*Haemulon striatum*), sailor's grunt (*Haemulon parra*), porkfish (*Anisotremus virginicus*), neon goby (*Gobiosoma oceanops*)
Reef omnivores	**Doctorfish (*Acanthurus chirurgus*)**, other surgeons (Acanthuridae spp.), blue angelfish (*Holacanthus bermudensis*), gray angelfish (*Pomacanthus arcuatus*), cherubfish (*Cantropyge argi*), rock beauty (*Holacanthus tricolor*), cocoa damselfish (*Pomacentrus variabilis*), bicolor damselfish (*Pomacentrus partitus*), beau gregory (*Pomacentrus leocostictus*), yellowtail damselfish (*Microspathodon chrysurus*), seaweed blenny (*Parablennius marmoreus*), striped parrotfish (*Scarus croicensis*), bibled goby (*Coryphopterus glaucofraenum*), Bermuda chub (*Kyphossus sectarix*)
Shrimps	**Pink shrimp (*Farfantepenaeus duorarum*)**, brown shrimp (*Farfantepenaeus aztecus*), white shrimp (*Litopenaeus setiferus*), other shrimp species
Large crabs	**Blue crab (*Callinectes sapidus*)**, stone crabs (*Menippe mercenaria* and *Menippe adina*), horseshoe crab (*Limulus polyphemus*), hermits crab (e.g., *Pylopagurus operculatus* and *Clibanaris vittatus*), spider crabs (e.g., *Stenocionops furcatus*), arrow crabs (e.g., *Stenorynchus seticornis*)

Table 2
Research survey data used to fit generalized additive models in the present study.

Research survey dataset	SEAMAP (Southeast Area Monitoring and Assessment Program) groundfish/trawl dataset	NMFS BLL (National Marine Fisheries Service bottom longline) dataset	SEAMAP reef fish video dataset
Data source	Public SEAMAP database (Rester, 2011)	Private NMFS database (Grace, NMFS Pascagoula, Mississippi, pers. comm.)	Private NMFS database (Campbell and Noble, NMFS Pascagoula, Mississippi, pers. comm.)
Sampling locations	Randomly chosen locations over the entire continental shelf of the Gulf of Mexico up to 200 m depth	Randomly chosen locations over the entire continental shelf of the Gulf of Mexico	Randomly chosen locations over the entire continental shelf of the Gulf of Mexico
Seasons of data collection	Summer (July–August) and/or fall (September–December), and, sometimes, spring (April–June) and winter (January–March)	Summer (usually July–September)	Spring–summer (May–August)
Sampling effort	Area swept by each tow	100 hooks fished for 1 h (soak time)	20-min record from one of the four camcorders used at a given station
Does sampling effort vary from one sampling event to the next?	Yes	No	No
Life-stage groups for which data are extracted from the dataset	All the life-stage groups represented in the OSMOSE-WFS model, except the different life stages of red grouper (*Epinephelus morio*) and gag grouper (*Mycteroperca microlepis*)	Older juvenile and adult red groupers and older juvenile gag grouper	Adult gag grouper

Table 3
Data used to construct seasonal maps of environmental and habitat parameters, resolution of these data, and manipulations required on these data to attain a contiguous surface with which to make predictions with generalized additive models.

| Environmental/habitat parameter	Data source	Resolution	Manipulations
Depth	DOC/NOAA/ESDISNGDC – Gulf of Mexico Data Atlas (Becker et al., 2009)	1.85 km	No manipulation
Surface chlorophyll a (Chl a)	MODIS-Terra 4 km Satellite-NASA Giovanni Portal (http://daac.gsfc.nasa.gov/giovanni/)	4 km	No manipulation
Bottom temperature	NODC/GOM Regional Climatology (Boyer et al., 2009)	0.1°	Spline interpolation
Bottom dissolved oxygen (DO)	NODC/GOM Regional Climatology (Boyer et al., 2009)	1.0°	Spline interpolation
Sediment type	dSEABED – Gulf of Mexico Data Atlas (Buczkowski et al., 2006; Jenkins, 2011)	Low	Nearest neighbor interpolation
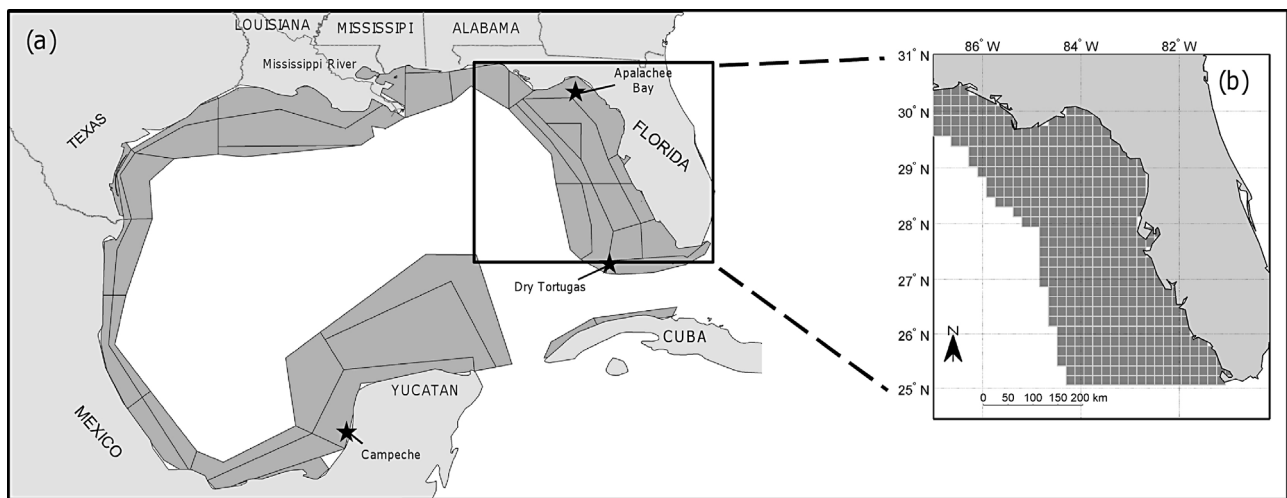
Fig. 1. Study areas. (a) Map of the Gulf of Mexico showing the spatial polygons of the Atlantis-GOM ecosystem model (filled in dark gray). Black stars indicate the location of the Dry Tortugas, Apalachee Bay and Campeche. (b) Map of the West Florida Shelf in the Gulf of Mexico showing the spatial cells of the OSMOSE-WFS ecosystem model (filled in dark gray).

the GAMs we wanted to fit (see Eq. (1) in Section 2.5.1), we estimated that the maximum number of coefficients that a GAM could include is equal to 41. The number of coefficients in a GAM cannot be larger than the number of data used to fit the model (Wood, 2006). Therefore, we considered that delta GAMs could be fitted for a specific life-stage group and season only if more than 40 of the abundance or biomass estimates available for this life-stage group and season were positive. Our preliminary analysis revealed that, given available research survey data, a total of 27 density maps can be created for the OSMOSE-WFS model using a GAM approach (Appendix A). In particular, it was decided that, due to a lack of data in the SEAMAP groundfish/trawl dataset for spring and fall, observations from spring and summer and from fall and winter would usually be lumped together.

2.3.1. SEAMAP groundfish/trawl dataset

SEAMAP groundfish/trawl data were used to fit GAMs for all study functional groups (including pink shrimp) except red grouper (*Epinephelus morio*) and gag grouper; these two species are scarce in the SEAMAP groundfish/trawl dataset.

SEAMAP groundfish/trawl surveys are carried out every year, in summer (July–August) and/or fall (September–December), and also, rarely, in spring (April–June) and winter (January–March), over the entire continental shelf of the GOM up to 200 m depth, but only in US waters (GSMFC, 2011; SEDAR, 2011). SEAMAP sampling stations are chosen along the US GOM shelf using a random design with proportional allocation by bottom area within statistical zones. SEAMAP groundfish/trawl surveys do not cover some historical fishing grounds of the US GOM shelf, including the area surrounding the Dry Tortugas. SEAMAP stations are typically sampled using a 40-feet otter trawl 24-h a day, with a tow time (bottom time) of 30 min per station, and at a speed of 2.5–3.5 knots (SEDAR, 2011).

Abundance and body size data collected during SEAMAP groundfish/trawl data surveys over the period 2000–2010 were extracted from the public SEAMAP database (Rester, 2011). Abundance data were aggregated by life-stage groups and seasons, using survey body size estimates and size benchmarks (e.g., size at maturity) from FishBase (Froese and Pauly, 2010). Sampling effort for the SEAMAP groundfish/trawl survey is considered equal to the area swept by each tow, which was calculated using the Euclidian distance between start and end points assuming a 40-feet trawling width.

2.3.2. NMFS BLL dataset

NMFS BLL data were used to fit GAMs for older juvenile and adult red groupers and older juvenile old gag grouper. NMFS BLL surveys are conducted each year in summer (usually July–September) throughout the US GOM shelf, using a commercial-type longline gear (Ingram et al., 2005; Henwood et al., 2006). Survey sites for NMFS BLL effort are selected at random within three depth strata (9–55 m, 55–183 m and >183 m). The number of randomly selected survey sites occupied within sampling zones is proportionally allocated based on the amount of continental shelf area within each of the depth strata (Grace, NMFS Pascagoula, Mississippi, pers. comm.). Sampling effort for the NMFS BLL survey is 100 hooks fished for 1 h (soak time). Fish caught by the longline gear are brought aboard or held at surface for estimating body size and weight or for tagging.

Biomass and body size data collected during the NMFS BLL surveys over the period 2001–2010 were extracted from a database made available by NOAA/NMFS (Grace, pers. comm.). Biomass data were aggregated by life-stage groups and seasons, using survey body size estimates and size benchmarks coming from FishBase.

2.3.3. SEAMAP reef fish video dataset

SEAMAP reef fish video data were used to fit GAMs for adult gag grouper. Video surveys take place each year in spring–summer (May–August) at different sites located throughout the US GOM shelf (Gledhill et al., 2005; Campbell et al., 2013). These surveys use a two-stage sampling design. The first stage uses stratified random sampling to select 10′ × 10′ blocks, defined by a geographic location and a total reef habitat area. A 0.1 by 0.1 mile grid is overlaid onto the reef area contained within blocks selected during the first stage, and ultimate survey sites are randomly selected from that grid (Campbell et al., 2013). SEAMAP reef fish video surveys use a fixed stereo video camera system comprising four camcorders, which is deployed to a maximum depth of 150 m for 40 min. One video tape from each survey station out of four is selected, from which a 20 min segment is viewed to obtain estimates of abundance ('mincount' estimates). 'Mincount' is the greatest number of fish that appears on screen at one time for a given 20-min video record.

SEAMAP reef fish video mincount data for gag grouper collected over the period 2001–2010 were made available by NOAA/NMFS (Campbell and Noble, NMFS Pascagoula, Mississippi, pers. comm.).
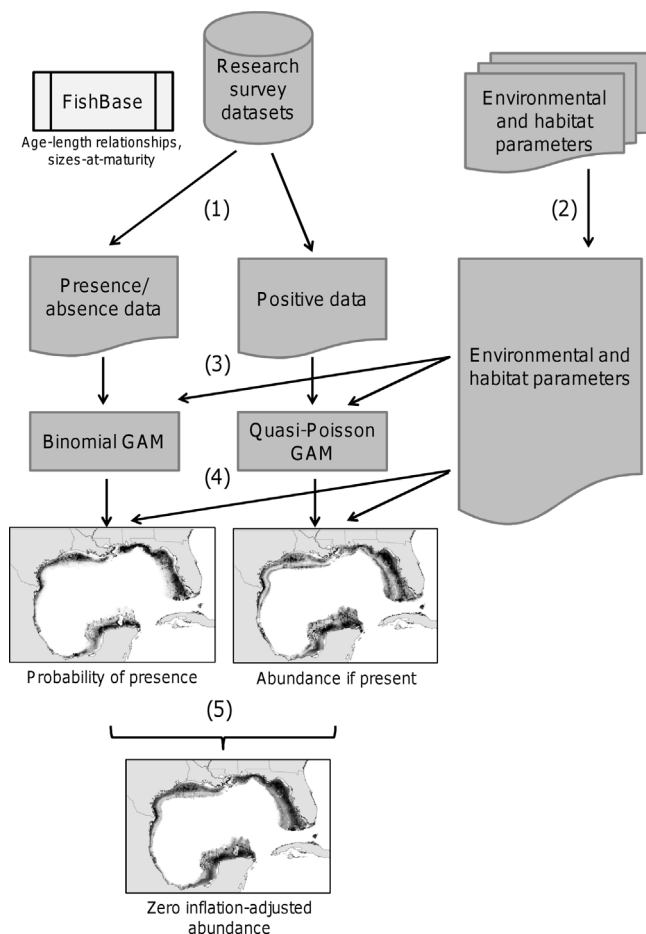
Fig. 2. Schematic of the different steps followed to predict spatial distributions over the Gulf of Mexico (GOM) shelf using a delta generalized additive modeling approach. (1) Extraction of presence/absence data and positive abundance/biomass data from survey datasets, and aggregation of these data by life-stage groups and seasons, using survey body size estimates and size benchmarks coming from Fish-Base. (2) Identification of pertinent environmental and habitat parameters to feed into generalized additive models (GAMs) and manipulations on these data to obtain a contiguous surface from which to fit GAMs and make predictions with these GAMs. (3) Fitting of binomial GAMs and quasi-Poisson GAMs using presence/absence data and positive abundance/biomass data, respectively; and environmental and habitat parameters. (4) Predictions of probabilities of a non-zero event and abundances/biomasses given a non-zero event over the GOM shelf using binomial GAMs and quasi-Poisson GAMs, respectively; and environmental and habitat parameters. (5) Predictions of spatial distributions over the GOM shelf through the combination of binomial and quasi-Poisson GAMs according to the delta method.

Given the depths at which SEAMAP reef fish video surveys take place, gag groupers identified on video records consist only of adult individuals.

2.4. Environmental and habitat predictors

Environmental predictors included in GAMs consist of depth (in m), surface chlorophyll a (Chl a; in $mg\,m^{-3}$), bottom dissolved oxygen (DO; in $mL\,L^{-1}$) and bottom temperature (in °C). A habitat predictor, sediment type, is also included in GAMs, and is divided into four categories: mud, sand, gravel and rock. The five aforementioned environmental and habitat predictors were chosen for the present study for their wide spatial distribution spanning the entire GOM shelf.

A 0.1° gridded map of long-term environmental and habitat parameters was created for each season (spring: April–June; summer: July–August; spring–summer: April–August; fall: September–December; winter: January–March; fall–winter:

September–March) (step (2) in Fig. 2). The data used to generate spatial estimates of environmental and habitat parameters and the manipulations required on these data to attain a contiguous surface with which to make predictions with GAMs are detailed below. For model fitting, the map of long-term environmental and habitat parameters produced for a given season was overlaid with the starting locations of research surveys for that season. Seasonal 0.1° gridded maps of environmental and habitat parameters were used both for (1) fitting GAMs (Section 2.5.1); and (2) predicting spatial distributions over the whole GOM shelf (i.e., including Mexican and Cuban waters and US areas where abundance and biomass survey estimates do not exist) using fitted GAMs (Section 2.5.2).

2.4.1. Spatial estimates of depth

We accessed the SRTM30_PLUS global bathymetry grid from the GOM Coastal Observing System (http://gcoos.tamu.edu/), from which a continuous raster of bathymetry with a resolution of 1.85 km was produced.

2.4.2. Spatial estimates of Chl a

We obtained measurements of Chl a by averaging the seasonal composites MODIS-Terra 4 km measurements and the NASA Ocean Biogeochemical Model from 2005 to 2010 accessed via the GIO-VANNI portal (http://daac.gsfc.nasa.gov/giovanni/).

2.4.3. Spatial estimates of DO and temperature

For each season, measurements of DO and bottom temperature at the maximum depth for each grid point were extracted from the National Oceanographic Data Center (NODC) regional climatology database (Boyer et al., 2009). These measurements were incomplete and the DO data had a low resolution (1.0°). Therefore, DO and bottom temperature data were subjected to a spline interpolation on a 0.1 °C grid using ArcGIS v10.0 (ESRI, 2011) so as to provide a contiguous surface from which to fit GAMs and make predictions with these GAMs.

2.4.4. Spatial estimates of sediment type

The best available data on sediment type, dSEABED2006 (Buczkowski et al., 2006; Jenkins, 2011), does not provide complete coverage for the entire GOM Shelf, though the area sampled by the SEAMAP and NMFS BLL surveys has substantial coverage. Moreover, dSEABED2006 data have a low resolution. Therefore, a nearest neighbor function was executed on a 0.1° grid using the natural neighbor function in ArcGIS v10.0 (ESRI, 2011) in order to provide a contiguous surface from which to fit GAMs and make predictions with these GAMs.

2.5. Delta GAM approach

We developed a GAM approach for predicting the spatial distribution of life-stage groups represented in the OSMOSE-WFS ecosystem model at different seasons, based on abundance or biomass survey estimates and environmental and habitat features at research survey locations. Due to the large number of zero abundance data in the SEAMAP datasets and of zero biomass data in the NMFS BLL dataset, and the need for a single parsimonious framework for predicting the spatial distribution of a large number of life-stage groups at different seasons, we opted for a relatively simple delta GAM approach. Our delta approach involves describing the probability of presence using a binomial GAM, and abundance or biomass if present using a quasi-Poisson GAM, and then multiplying the results together to obtain an overall abundance or biomass (steps (3)–(5) in Fig. 2).

2.5.1. Model fitting

Both the binomial GAM and the quasi-Poisson GAM were fitted using the 'mgcv' package in the R version 2.15.3 environment (Wood, 2006), following the equation:

$$g\ (\eta) = s(\text{depth}) + s(\text{chl a}) + s(\text{temperature}) + s(\text{DO})$$
$$+ factor\ (\text{sediment type}) \tag{1}$$

where η is either the probability of success when given binomial response data, or an estimate of abundance or biomass when given non-zero data; g represents the link function between η and each additive predictor; and s is a thin plate regression spline fit to a given predictor. A logit link function is used in the binomial GAM, and a log link function in the quasi-Poisson GAM. When SEAMAP groundfish/trawl data are provided to the quasi-Poisson GAM, the model includes an offset, with equivalent link function, to account for the fact that SEAMAP groundfish/trawl fishing effort (i.e., the area swept by each tow) can vary from one sampling event to another. (i.e., $offset(g(effort))$ is added at the right side of Eq. (1)).

Smoothness selection for GAMs was fitted using a spline-based penalized likelihood estimation. Weighted penalties in the binomial and quasi-Poisson GAMs were determined, respectively, by Un-Biased Risk Estimator (UBRE) and Generalized Cross Validation (GCV) criteria (Wood, 2006). An extra penalty was applied to each predictor as the smoothing parameter approached zero. This allowed the complete removal of a predictor from a GAM when the smoothing parameter equals zero. The extra penalty results in partially automated GAM selection, which is convenient given that binomial and quasi-Poisson GAMs need to be fitted for a large number of life-stage groups at different seasons. However, the suite of predictors given in Eq. (1) needs to be reduced for some life-stage groups to prevent GAM overfitting or runaway predictions (Koubbi et al., 2006). Therefore, GAM fitting is often an iterative rather than a one-shot process.

2.5.2. Model evaluation

The binomial and quasi-Poisson GAMs that were fitted for the different life-stage groups represented in OSMOSE-WFS at different seasons were evaluated by analyzing the total deviance explained by the models (0–100%; the higher the better). Then, the binomial and quasi-Poisson GAMs were used to predict, respectively, the likelihood of non-zero values and the abundance or biomass across the 0.1° gridded seasonal maps of environmental and habitat parameters in locations of non-zero density. Next, respective cell values from the two GAMs were combined to yield an estimate of mean abundance or mean biomass according to the equation:

$$predicted(y) = p * u \tag{2}$$

where y is the estimate of abundance or biomass; p the probability that $y > 0$ estimated by the binomial GAM; and u the expected abundance or biomass from the quasi-Poisson model. The resulting delta GAM was used to predict abundance or biomass across the 0.1° gridded seasonal map of environmental and habitat parameters for water depths up to 200 m.

2.5.3. Model validation

The delta GAMs that we fitted for the different life-stage groups represented in OSMOSE-WFS at different seasons were validated using the datasets internal to their development where observed and predicted values of abundance or biomass could be compared. From these datasets of observed and predicted densities, 1000 bootstrap datasets were produced by resampling with replacement within the range of observed and predicted densities. Then, Spearman's correlation coefficients (Spearman's ρ's) between the abundance or biomass values predicted by delta GAMs and those observed in research survey datasets were estimated and tested to be significantly different from zero (Vaz et al., 2006; Loots et al., 2007).

We initially wanted to validate delta GAMs using datasets external to their development. To do so, we would have split research survey data into two sets at random, and performed GAM fitting with the first set only and model validation with the second set only (Araújo et al., 2005; Austin, 2007; Drexler and Ainsworth, 2013). It was not possible to proceed this way for some of the life-stage groups and seasons due to a scarcity of research survey data. As mentioned earlier, the number of coefficients in a GAM cannot be larger than the number of data used to fit the model. If we had split up data for validation purposes, we would have ended up in a situation where the number of coefficients in the quasi-Poisson GAM would have been larger than the number of data used to fit the model in several life-stage groups and seasons. Therefore, in order to utilize a single, simple framework to validate the totality of the delta GAMs we fitted, and to prevent lowering the predictive power of GAMs, we chose to use the bootstrap procedure described above.

2.5.4. Model demonstration

We used pink shrimp as a case study to demonstrate our delta GAM approach. We analyzed thoroughly the influence of environmental and habitat parameters on the spatial distribution of pink shrimp during the summer season. Then, we predicted the summer distribution of the species over the entire GOM shelf in summer, using the 0.1° gridded map of environmental and habitat parameters for the summer season described in Section 2.4. Finally, to illustrate the usefulness of the delta GAM approach to generate distribution maps for spatially explicit ecosystem models, the predictions of the delta GAM for the summer distribution of pink shrimp were averaged according to the Atlantis-GOM spatial polygons and the OSMOSE-WFS grid cells (Fig. 1). We also produced density maps for different life-stage groups represented in OSMOSE-WFS model at different seasons from those delta GAMs that were validated.

2.6. Negative binomial GAM for pink shrimp in summer

To allow for comparisons between our delta GAM approach and the negative binomial GAM approach applied by Drexler and Ainsworth (2013), we also developed a negative binomial GAM with a log link function for pink shrimp in summer. This negative binomial GAM was fitted using the 'mgcv' package in the R version 2.15.3 environment (Wood, 2006), following Eq. (1) plus an offset for fishing effort at the right side of Eq. (1). The smoothness selection, weighted penalties (determined here by UBRE) and extra penalty described in Section 2.5.1 were implemented here as well. The θ parameter was determined by UBRE, and its estimation was limited to a range of 1–10 (Wood, 2006). The computation time needed to fit the negative binomial GAM was recorded and compared to the time needed to implement our delta GAM approach for pink shrimp in summer.

A Spearman's ρ between observed and predicted values was estimated for the negative binomial GAM, using the bootstrap procedure described in Section 2.5.3. The negative binomial GAM was then used to predict pink shrimp abundance in summer for water depths up to 200 m from the grid map of environmental and habitat parameters constructed for the summer season. Finally, the predictions of the negative binomial GAM were averaged according to the Atlantis-GOM spatial polygons and the OSMOSE-WFS grid cells.

	Absent	Present	% total
Mud	1080	174	44
Sand	490	196	49
Gravel	17	4	1
Rock	78	22	6

3. Results

3.1. Results for pink shrimp in summer

3.1.1. Evaluation and validation of GAMs

For pink shrimp in summer, the binomial GAM described 16.5% of the deviance in presence/absence data, and the quasi-Poisson GAM 32.5% of the deviance of the non-zero abundance data. Comparing the abundance values predicted by the delta GAM to observed abundance values resulted in a positive and highly significant Spearman's correlation coefficient ($\rho = 0.35$ with 1000 bootstrapped datasets; 95% confidence interval (CI): 0.31–0.38; $p < 0.001$). Depth, Chl a, temperature and DO, as well as sand and rock sediment types were found to be significant predictors ($p < 0.05$) of, respectively, presence/absence and abundance at non-zero locations. All the aforementioned environmental and habitat parameters were therefore retained in the binomial and quasi-Poisson GAMs used for predictions for pink shrimp in summer.

The negative binomial GAM fitted for pink shrimp in summer described 57% of the deviance in abundance data and yielded a Spearman's ρ between modeled and observed abundance values slightly greater than that obtained with the delta GAM approach ($\rho = 0.38$ with 1000 bootstrapped datasets; 95% CI: 0.35–0.42; $p < 0.001$). However, it took around 13 min to implement the negative binomial GAM approach in the R version 2.15.3 environment with a 2.4 GHz Intel Core i7-2760 qm processor vs. 11 s to implement the delta GAM approach.

3.1.2. Influence of environmental and habitat parameters on pink shrimp spatial distribution

Pink shrimp was observed in 396 of the 2061 summer SEAMAP groundfish/trawl locations used to fit GAMs (~19%; Table 4). The temperatures in the observations used to fit the binomial GAM usually range between 10 and 32 °C (Fig. 3a). The probability of pink shrimp presence in summer is negatively affected by temperatures lower than 15 °C, does not fluctuate greatly for temperatures ranging from 15 to 30 °C, and is positively affected by temperatures greater than 30 °C. Temperatures greater than 38 °C likely reflect the error associated with the interpolation of environmental data across the sampling grid. The quasi-Poisson GAM found the abundance of pink shrimp in summer not to fluctuate greatly across the range of temperature values observed, especially for temperatures greater than 20 °C (Fig. 3b).

While sand and mud share an equivalent probability of occurrence in the binomial GAM (49% and 44%, respectively; Table 4), the quasi-Poisson GAM predicts significantly higher densities of pink shrimp on sand and hard bottom than on gravel and mud (Fig. 4a).

Depth has a slight positive effect on the probability of pink shrimp presence in summer up to 30 m; any depth greater than 30 m has a negative effect on the probability of presence of pink shrimp (Fig. 3c). The quasi-Poisson GAM predicts high abundance of pink shrimp at depths between roughly 10 and 40 m, as well as at depths above 75 m, where the species was observed at a few SEAMAP groundfish/trawl stations (Fig. 3d).

Chl a concentrations lower than 5 mg m^{-3} have a very slight positive effect on the probability of pink shrimp presence in summer. The probability of presence of pink shrimp decreases with increasing Chl a density when Chl a concentration is higher than 15 mg m^{-3} (Fig. 3e). Chl a density has a slight negative effect on the abundance of pink shrimp at non-zero locations for Chl a concentrations ranging between 2 and 6 mg m^{-3}. The negative effect of Chl a density on pink shrimp abundance is greater for Chl a concentrations above 6 mg m^{-3}, where pink shrimp was observed at a few SEAMAP groundfish/trawl stations (Fig. 3f).

The probability of pink shrimp presence in summer does not fluctuate greatly across the range of DO values observed when DO concentrations are lower than 4 mL L^{-1}. DO concentrations higher than 4 mL L^{-1} have a slight positive effect on the probability of pink shrimp presence (Fig. 3g). DO concentrations greater than 6 mL L^{-1} likely reflect the error associated with the interpolation of environmental estimates. The quasi-Poisson GAM predicts that abundance of pink shrimp is highest for DO concentrations in the range of 2.5–4 mL L^{-1}, and drops when DO concentrations are >4.5 mL L^{-1} (Fig. 3h).

The predictions of the negative binomial GAM we fitted for pink shrimp in summer and those of the negative binomial GAM fitted in Drexler and Ainsworth (2013) will not be identical since we used all SEAMAP groundfish/trawl pink shrimp data to train the model while Drexler and Ainsworth (2013) used 2/3 of these data for model training and 1/3 for model validation. However, we observed that predictions of the negative binomial GAM we fitted (Figs. 3i–l and 4b) are similar to those of the negative binomial GAM fitted in Drexler and Ainsworth (2013). We identified only one small discrepancy between the negative binomial GAM we fitted and Drexler and Ainsworth (2013)'s negative binomial GAM; at Chl a concentrations higher than 4.5 mL L^{-1}, pink shrimp abundance in summer drops sharply in our negative binomial GAM, while it increases slightly in Drexler and Ainsworth (2013). This small discrepancy has a negligible impact on the spatial distributions predicted by negative binomial GAMs since Chl a concentrations usually do not exceed 4.5 mL L^{-1} in shelf areas of the GOM in summer (http://daac.gsfc.nasa.gov/giovanni/). Insights from our negative binomial GAM into environmental and habitat effects on pink shrimp distribution are similar to those gained by considering the predictions of the binomial and quasi-Poisson GAMs together (compare Fig. 3i to a + b, j to c + d, k to e + f, l to g + h, and Fig. 4b to a). The negative binomial GAM, like the binomial and quasi-Poisson GAMs, predicts depth, Chl a, temperature, DO, and sand and rock sediment types to be significant predictors ($p < 0.05$) of pink shrimp spatial distribution in summer.

3.1.3. Distribution maps

For pink shrimp in summer, the binomial and quasi-Poisson GAMs predict, respectively, areas with high probability of non-zero abundance and areas of high abundance over the entire mid-depth portion of the West Florida Shelf, in the area around the Dry Tortugas, near the Louisiana-Texas border, on the Texas-Mexico border and on the northern Campeche Bank (Fig. 5a and b). The highest probabilities of presence were found in the area surrounding the Dry Tortugas, near the Florida Panhandle and along the coast of Yucatan with values >0.7 (Fig. 5a). Areas of highest abundance if present were found in the area surrounding the Dry Tortugas, near the Florida Panhandle and especially in the Apalachee Bay, and in the northern Gulf of Campeche (Fig. 5b). The delta GAM, predicts a similar distribution to the binomial GAM with a reduced abundance compared to the quasi-Poisson GAM. According to the delta GAM, areas of highest abundance for pink shrimp in summer occur at mid depth over the West Florida Shelf, in the area around the Dry Tortugas, on the northern Campeche Bank and, to a lesser extent, near the Louisiana-Texas border and on the Texas-Mexico border (Fig. 5c). In the quasi-Poisson GAM, pink shrimp is predicted to be relatively abundant at depths above 75 m, where the species was
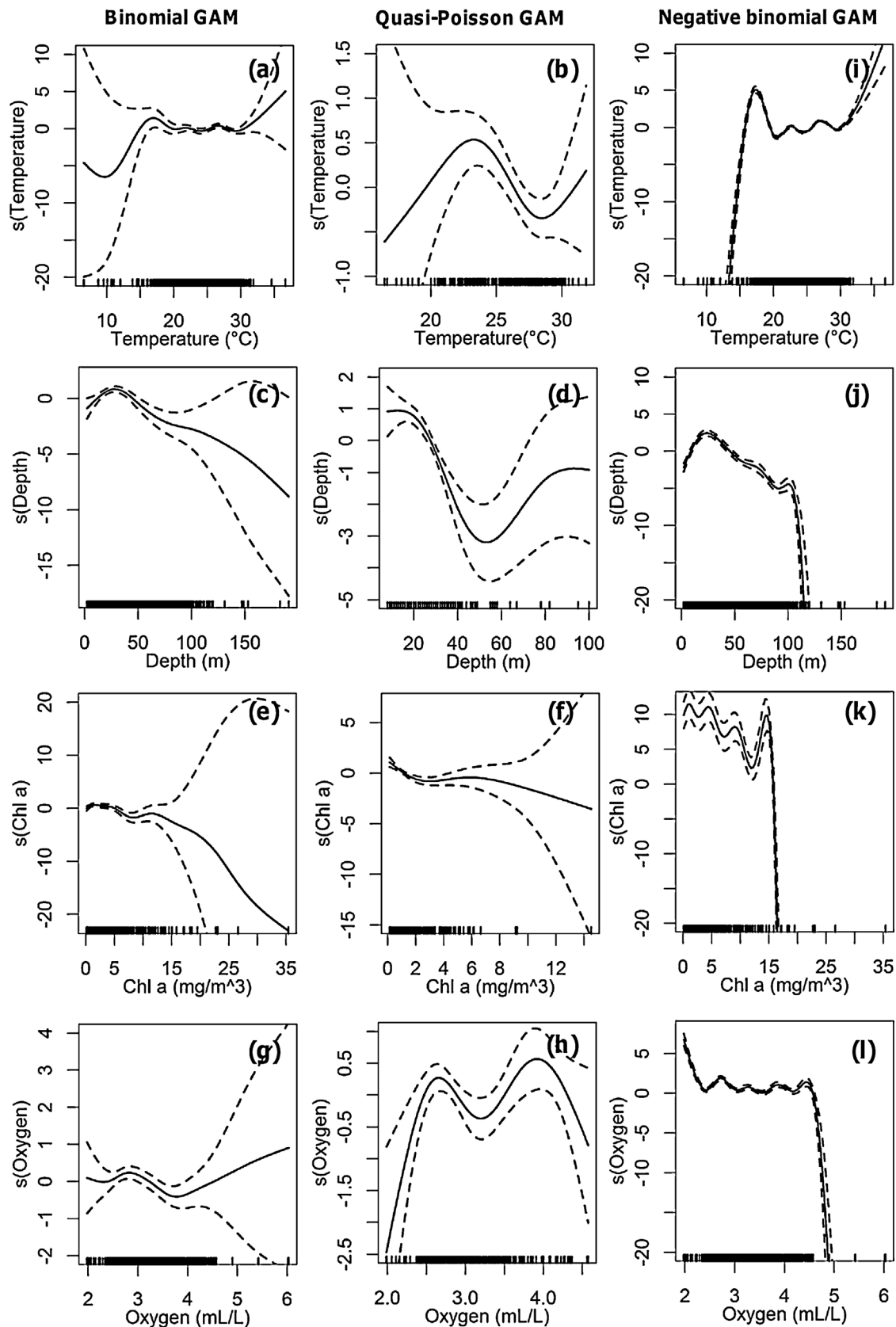
Binomial GAM **Quasi-Poisson GAM** **Negative binomial GAM**

Fig. 3. Fits for the binomial, quasi-Poisson and negative binomial GAMs developed for predicting the spatial distribution of pink shrimp (*Farfantepenaeus duorarum*) in summer. Smoothed curve of the additive effect to the estimated abundance of pink shrimp for the individual environmental parameters considered in the GAMs. Dotted lines represent 95% confidence intervals, and each mark along the *x*-axis a single observation. Note that the scale of the *y*-axis differs from one panel to the next for display purposes.
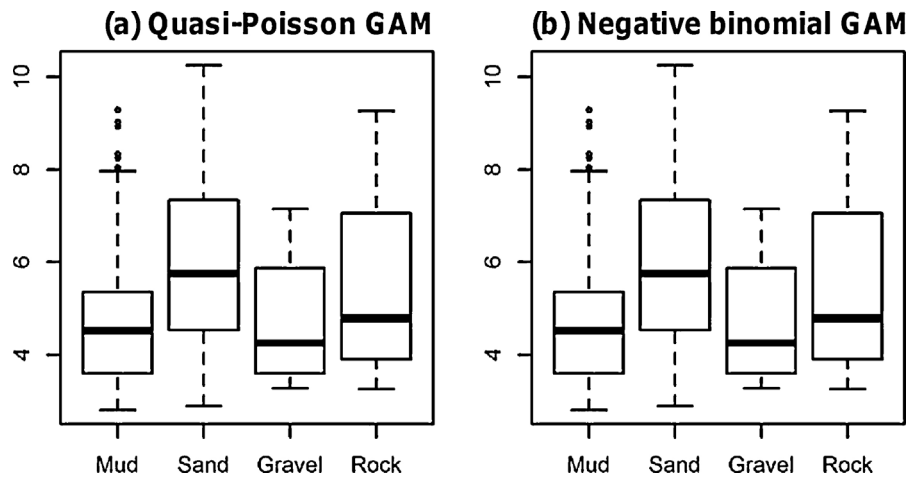
Fig. 4. Distribution of pink shrimp abundance for each sediment type predicted by (a) the quasi-Poisson GAM and (b) the negative binomial GAM. Abundance is expressed here in ln(abundance + 1). Because we are primarily interested in which sediment type is suitable for pink shrimp the zeros have been removed for (b) for display purposes.

observed at a few research survey stations. This is not the case in the delta GAM (Figs. 5a and 6a). The negative binomial GAM concurs with the delta GAM as to hotspots of pink shrimp in summer (Fig. 6b). Note that the absolute abundances predicted by the negative binomial GAM are significantly higher than those predicted by the delta GAM. This makes sense given that the negative binomial GAM approach is a one-step approach, while the delta GAM approach is a two-step approach where predictions are obtained as the combination of a probability of presence and an estimate of abundance (Eq. (2)).

Spatial estimates of abundance for pink shrimp in summer were averaged according to the Atlantis-GOM spatial polygons and the OSMOSE-WFS grid cells, as shown in Fig. 6c–f. Atlantis-GOM polygon spatial distributions reflect spatial distributions directly predicted by the delta GAM over a 0.1° grid (Fig. 6c). Areas of highest abundance for pink shrimp are located in polygons at mid depth over the West Florida Shelf, on the northern Campeche Bank and on the Texas-Mexico border. OSMOSE-WFS grid spatial predictions provide insight into hotspots of pink shrimp in summer in the West Florida Shelf (Fig. 6d). In this area of the GOM, highest abundance of pink shrimp occurs north of 28° N, and especially in the waters off the Florida Panhandle. Atlantis-GOM polygon and OSMOSE-WFS grid spatial distributions constructed from predictions of the negative binomial GAM are similar to those produced
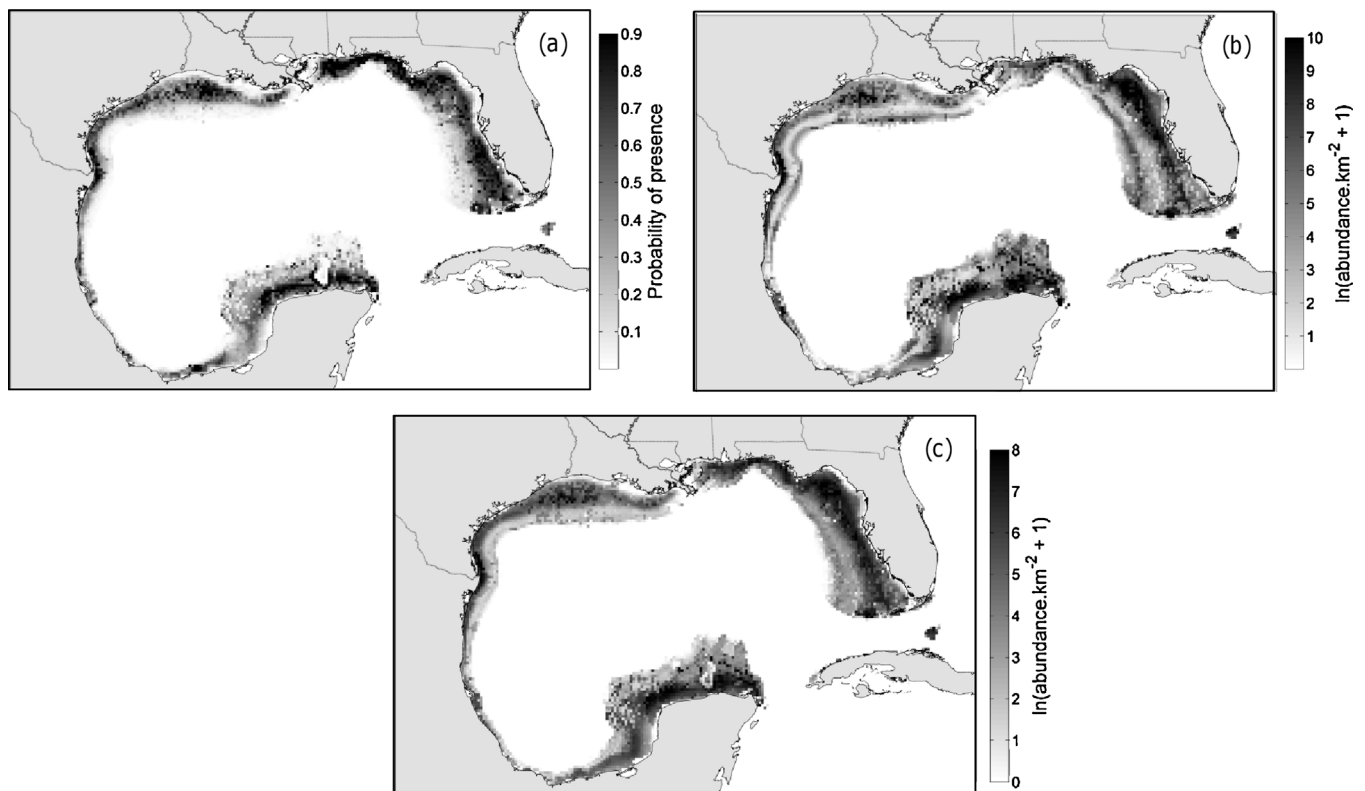

Fig. 5. Results of the (a) binomial, (b) quasi-Poisson and (c) delta GAMs for pink shrimp in summer. (a) Probability of a non-zero event predicted by the binomial GAM. (b) Catch per unit effort (CPUE), expressed as ln(abundance km^{-2} + 1) given a non-zero event predicted by the quasi-Poisson GAM. (c) Combination of the binomial and quasi-Poisson GAM according to the delta method to yield a zero inflation adjusted estimate of CPUE, expressed as ln(abundance km^{-2} + 1).

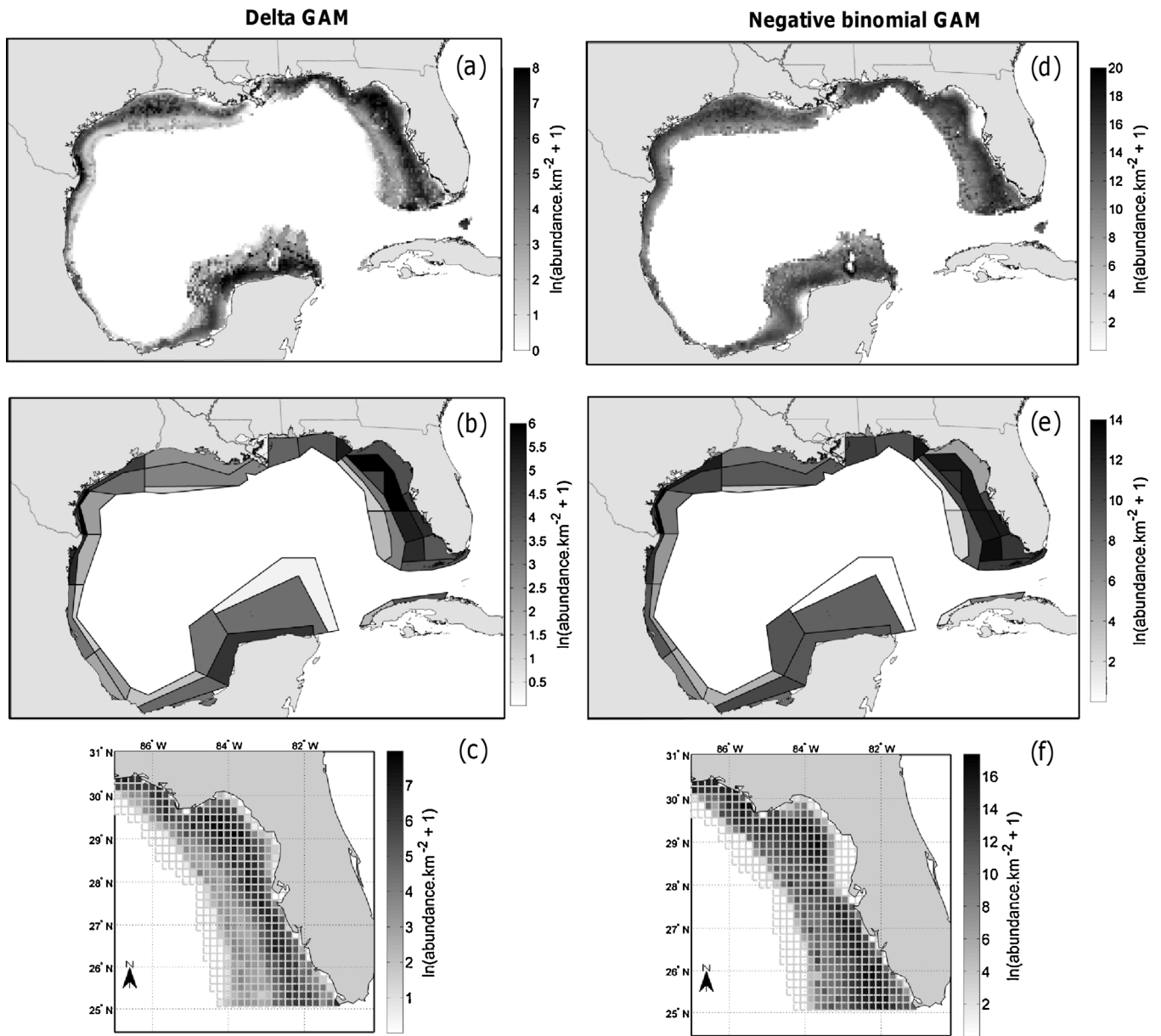**Delta GAM**   **Negative binomial GAM**

Fig. 6. Spatial patterns of pink shrimp abundance in summer predicted by the delta GAM (a, c, d) and the negative binomial GAM (b, e, f) at different spatial resolutions. (a, b) Spatial patterns of catch per unit effort (CPUE), expressed in ln(abundance km^{-2} + 1), for the entire Gulf of Mexico directly predicted by the (a) delta and (b) negative binomial GAMs (0.1° gridded map). (c, e) Spatial aggregation performed for the Atlantis-GOM ecosystem model. CPUE expressed in ln(abundance km^{-2} + 1) derived from (a, b) according to the box geometry of Atlantis-GOM. (d, f) Spatial aggregation performed for the OSMOSE-WFS ecosystem model, which focuses on the West Florida Shelf. CPUE expressed in ln(abundance km^{-2} + 1) derived from (a, b) according to the spatial grid of OSMOSE-WFS.

from predictions of the delta GAM (Fig. 6e and f). However, relative abundance of pink shrimp in the area surrounding the Dry Tortugas and in the Southern West Florida Shelf (around 26° N) is higher with the negative binomial GAM than with the delta GAM, while relative abundance of the species on the northern Campeche Bank is higher with the delta GAM.

3.2. Results for all study functional groups

3.2.1. Evaluation and validation of GAMs

The binomial GAMs fitted for the life-stage groups represented in OSMOSE-WFS at different seasons described between 3.3% and 53.2% of the deviance in presence/absence data with a median value of 12.4%; and the quasi-Poisson GAMs fitted for the different life-stage groups between 3.8% and 86.6% of the deviance of non-zero data with a median value of 23.9% (Table B).

In every life-stage group and season, the Spearman's ρ between the abundance/biomass values predicted by delta GAMs and observed abundance/biomass values was positive and highly significant ($p < 0.001$) and the 95% confidence interval of this Spearman's ρ was positive. Mean Spearman's ρ's ranged between 0.09 and 0.47, with a median value of 0.15 (Table B).

The significance of environmental and habitat predictors for the different life-stage groups and seasons varied greatly. Moreover, GAM fitting for some life-stage groups and seasons was an iterative process, which led to the exclusion of some predictors, so as to prevent runaway predictions or GAM overfitting.

3.2.2. Distribution maps

The density maps we produced for different life stages of the functional groups represented in OSMOSE-WFS ecosystem model using the delta GAM approach are given in Appendix A.

4. Discussion

In the present paper, we developed a parsimonious framework to predict the seasonal distribution of different life stages of the multiple functional groups (life-stage groups) represented in spatially explicit ecosystem models from a relatively limited number of observations. This framework consists in fitting two independent statistical models, a binomial GAM and a quasi-Poisson GAM, whose predictions are then combined using a delta method to yield spatial estimates of abundance or biomass. We implemented a delta GAM approach primarily to produce density maps for the OSMOSE-WFS ecosystem model, which focuses on the West Florida Shelf.

4.1. Demonstration with pink shrimp

We demonstrated our delta GAM approach for pink shrimp in summer. The binomial and quasi-Poisson GAMs fits for pink shrimp in summer were in agreement with previous studies related to the habitat preference of pink shrimp with regard to temperature, sediment type and depth. Predictions of our binomial and quasi-Poisson GAMs on the effects of temperature and sediment type on pink shrimp spatial distribution concurred with Mulholland (1984)'s findings (Figs. 3a, b and 4a). Moreover, water depth ranges associated with highest pink shrimp abundance from the GMFMC (2006) data were in agreement with quasi-Poisson GAM results (Fig. 3d). In addition to reflecting previous research on habitat suitability for pink shrimp, our binomial and quasi-Poisson GAMs provided new insights into the effects of chlorophyll a and dissolved oxygen concentrations on the spatial distribution of the species (Fig. 3e and h).

A positive and highly significant Spearman's correlation coefficient (Spearman's ρ) between observed and predicted abundance values indicated that our delta GAM can reliably be used to predict the spatial distribution of pink shrimp. The spatial distributions predicted by the delta GAM reflect results from the literature (Costello and Allen, 1970; Bielsa et al., 1983) and predictions of the negative binomial GAM designed by Drexler and Ainsworth (2013). In particular, the delta GAM manages to identify hotpots of pink shrimp in areas that were not sampled but where the species is known to be highly abundant and to have been subjected to high fishing pressure (e.g. the area near the Dry Tortugas; Bielsa et al., 1983).

4.2. Delta GAMs vs. negative binomial GAMs

Our delta GAM and negative binomial GAM approaches provide similar insights into environmental and habitat effects on species distribution and into spatial distributions. Nonetheless, for the summer pink shrimp distribution, the percentage of deviance explained by the negative binomial GAM was greater than the percentage of deviance explained by both the binomial GAM and the quasi-Poisson GAM. The Spearman's ρ obtained for pink shrimp in summer with the negative binomial GAM was also slightly greater than the Spearman's ρ obtained with the delta GAM. Moreover, the level of deviance in abundance data explained by the negative binomial GAMs fitted in Drexler and Ainsworth (2013) (between 10 and 83% with a median value of 33.6%) is higher than the level of deviance in presence/absence data and abundance data explained, respectively, by the binomial GAMs and quasi-Poisson GAMs fitted in the present study.

While the negative binomial GAM performs better than the delta GAM, it takes 1/70th of the computation time to implement the delta GAM approach. The efficiency of the delta GAM calculation is critical when one needs to construct a large number of distribution maps for spatial ecosystem models since GAM fitting is often an iterative rather than a one-shot process. In other words,

using a delta rather than a negative binomial GAM saves significant computation time at the expense of a slight reduction in GAM performance.

4.3. Using delta GAMs to produce distribution maps for OSMOSE-WFS

The Spearman's ρ between observed abundance/biomass values and abundance/biomass values predicted by delta GAMs evaluated for the different life-stage groups represented in the OSMOSE-WFS model and seasons was always positive and highly significant. Therefore, the different delta GAMs correctly predict qualitative differences between low- and high-abundance areas, and are all deemed appropriate for generating distribution maps for OSMOSE-WFS.

It must be noted that Spearman's ρ's were often low (Table B), suggesting that predictions of some delta GAMs are not entirely reliable. However, this result does not compromise the use of the entire set of density maps we produced for OSMOSE-WFS, because: (1) distribution maps that are fed into spatially explicit ecosystem models only need to be good enough to guide functional group distributions toward a realistic spatial equilibrium; and (2) our density maps globally agree with spatial distributions emerging from simulations of an Ecospace model of the West Florida Shelf and with information from the literature.

The density maps we generated for OSMOSE-WFS using delta GAMs, regardless of their degree of precision, offer a considerable improvement over the assumption of homogeneous biomass distribution typically made in fisheries stock assessments. Through the use of these density maps, the degree of spatial overlap between prey and their potential predators can be explicitly considered at each model time step, allowing for a more realistic evaluation of predations mortality rates and diet compositions. Furthermore, density maps produced using delta GAMs do not predict stable equilibrium spatial distributions in OSMOSE-WFS, due to (i) the distribution of limited numbers of schools over the OSMOSE-WFS spatial grid at run-time using the density maps; and (ii) the implementation of small-scale random walk movements in OSMOSE-WFS when the distribution of schools remains static. Thus, spatial distributions initialized in OSMOSE-WFS with density maps settle at run-time to a new, but related, stable equilibrium. Therefore, distribution maps fed into OSMOSE-WFS or any other spatially explicit ecosystem model do not need to exhibit high precision since such benefits are lost at run-time.

We observed that the density maps we produced for OSMOSE-WFS are globally in agreement with the spatial distributions emerging from simulations of WFS Reef Fish Ecospace, an Ecospace model of the West Florida Shelf (Chagaris, Fish Wildlife Research Institute, St. Petersburg, Florida, pers. comm.). Furthermore, density maps generated for OSMOSE-WFS can be deemed reliable based on information from the literature, and especially from knowledge of the seasonal movement patterns of functional groups (Appendix A). For example, most coastal omnivores (e.g., pinfish, *Lagodon rhomboides*) are known to move to offshore waters during the late fall and spawn there from late fall to early spring (Darcy, 1985; Hansen, 1969). The distribution maps we obtained with delta GAMs for coastal omnivores for spring–summer and fall–winter reflect these seasonal movement patterns (Appendix A). The density maps provided in Appendix A will be used for the initial spatial allocation of functional groups abundance in the OSMOSE-WFS model until additional abundance and biomass data (e.g., spatial fisheries-dependent data) and environmental data become available to improve the predictions of delta GAMs.

4.4. Limitations of our delta GAM approach and perspectives

The goal of the present study was to introduce a single, simple framework to construct distribution maps for a large number of life-stage groups including both data-rich and data-limited animals. Our work is a first step in predicting the relative abundance of life-stage groups at different seasons. It is clear that improvements could be made to our delta GAM for data-rich life-stage groups. Improvements could include the consideration of spatial autocorrelation, though this may be practical in only a few cases due to limited data availability (see discussions in Drexler and Ainsworth, 2013). Moreover, spatial distributions are not controlled only by environment and physical habitat. Predictors such as population memory, population size and demography could eventually be added in our delta GAM for some functional groups provided the data are available (Loots et al., 2010). Finally, the selection of binomial and quasi-Poisson GAMs could be automated using selection procedures, such as stepwise regression using the Akaike Information Criterion (AIC) (using, e.g., the 'step.gam' function in the R package 'gam'; Hastie, 2006). However, stepwise regressions should be utilized with extreme care (Guisan et al., 2002; Loots et al., 2010). Stepwise regressions are high-variance operations in that small perturbations in the response variable can sometimes result in vastly different subsets of the predictors (Guisan et al., 2002). Thus, model selection using stepwise regression can lead to overfitted GAMs, which display a good adjustment to data but have a low predictive power (Loots et al., 2010).

As emphasized in Guisan et al. (2002), interactions between two or more predictors of GAMs should be evaluated. However, this is rarely done in practice. We evaluated whether environmental covariates were correlated to one another using Pearsons' correlation coefficient. None of the environmental covariates were found to be correlated with the exception of depth and bottom temperature (results not presented here). While we were primarily concerned with data availability, the environmental predictors of our GAMs represent the only long term-environmental predictors available across the entire GOM. The main concern with correlations between predictors is that: (1) collinear variables may appear non-significant if modeled together, but (2), if modeled separately, they may explain a significant proportion of the deviance. The second issue becomes critically important in our case, given the penalization terms which can drop variable influence from the GAMs, and the potential elimination of predictors from Eq. (1) to prevent GAM overfitting or runaway predictions. Relatively complex procedures exist to deal with interactions between two or more predictors of GAMs, which could be used in future studies. These procedures include stepwise model selection for correlated variables and defining the structure of the covariance using generalized additive mixed models (GAMMs) (Guisan et al., 2002; Wood, 2006).

Due to a scarcity or lack of data, it is not possible to use our delta GAM approach to estimate the spatial distribution of the early life stages of most functional groups for the OSMOSE-WFS model. Information from the literature, experts' opinion or simulations of larval dispersal using biophysical models can be used to produce distribution maps for these life-stage groups. Further modeling will be done using a biophysical model (the Connectivity Modeling System; Paris et al., 2013) to be able to generate density maps for younger juvenile red and gag groupers for OSMOSE-WFS.

The delta GAM approach introduced in the present study was developed essentially to produce density maps for the OSMOSE-WFS model, but it can be utilized to obtain distribution maps for other spatially explicit ecosystem models developed for the entire or part of the GOM shelf. This framework could also be adapted for initializing biomass distributions in spatially explicit ecosystem models designed for any geographical region, where wide spatial and taxonomic coverage is desired while benefits of high precision estimates are lost at run-time.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fishres.2014.05.005.

References

Ainsworth, C.H., Kaplan, I.C., Levin, P.S., Cudney-Bueno, R., Fulton, E.A., Mangel, M., Turk-Boyer, P., Torre, J., Pares-Sierra, A., Luna, H.N.M., 2011. Atlantis Model Development for the Northern Gulf of California. U.S. Dept. Commer, NOAA Tech. Memo. NMFS-NWFSC-110, 293 pp.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species–climate impact models under climate change. Glob. Change Biol. 11, 1504–1513.

Austin, M., 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. 200, 1–19.

Barry, S.C., Welsh, A.H., 2002. Generalized additive modelling and zero inflated count data. Ecol. Model. 157, 179–188.

Becker, J.J., Sandwell, D.T., Smith, W.H.F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.H., 2009. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. Mar. Geod. 32, 355–371.

Bielsa, L.M., Murdich, W.H., Labisky, R.F., 1983. Species Profiles: Life Histories and Environmental Requirements of Coastal Fishes and Invertebrates (South Florida): Pink Shrimp. U.S. Fish and Wildlife Service. FWS/OBS-82/11.17. U.S. Army Cops of Engineers, TR EL-82-4, 21 pp.

Boyer, T., Antonov, J., Baranova, O.K., Garcia, H.E., Johnson, D.R., Locarnini, R.A., Mishonov, A.V., O'Brien, T.D., Seidov, D., Smolyar, I.V., Zweng, M.M., 2009. World Ocean Database 2009. In: Levitus, S. (Ed.), NOAA Atlas NESDIS 66. U.S. Government Printing Office, Washington, DC, 216 pp.

Brochier, T., Ecoutin, J.M., de Morais, L.T., Kaplan, D., Lae, R., 2013. A multi-agent ecosystem model for studying changes in a tropical estuarine fish assemblage within a marine protected area. Aquat. Living Resour. 26, 147–158.

Buczkowski, B., Reid, J., Jenkins, C., Reid, J., Williams, S.J., Flocks, J., 2006. usSEABED: Gulf of Mexico and Caribbean (Puerto Rico and US Virgin Islands) Offshore Surficial Data Release: U.S. Geological Survey Data Series 146, version 1.0.

Campbell, M.D., Rademacher, K.R., Felts, P., Noble, B., Felts, M., Salisbury, J., 2013. SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Gag. SEDAR33-DW15. SEDAR, North Charleston, SC, 39 pp.

Christensen, V., Walters, C.J., 2004. Ecopath with Ecosim: methods, capabilities and limitations. Ecol. Model. 172, 109–139.

Coleman, F.C., Figueira, W.F., Ueland, J.S., Crowder, L.B., 2004. The impact of United States recreational fisheries on marine fish populations. Science 305, 1958–1960.

Costello, T.J., Allen, D.M., 1970. Synopsis of biological data on the pink shrimp *Penaeus Dourarum Duorarum* Burkenroad, 1939. FAO Fish. Rep. 57, 1499–1537.

Darcy, G.H., 1985. Synopsis of Biological Data on the Pinfish, *Lagodon rhomboides* (Pisces: Sparidae). NOAA Technical Report NMFS 23. U.S. Department of Commerce, NOAA, National Marine Fisheries Service, Seattle, Washington, 31 pp.

Drexler, M., Ainsworth, C.H., 2013. Generalized additive models used to predict species abundance in the Gulf of Mexico: an ecosystem modeling tool. PLoS ONE 8, e64458.

ESRI, 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.

Froese, R., Pauly, D., 2010. A Global Information System on Fishes. Fishbase http://www.fishbase.org

Fu, C., Ian Perry, R., Shin, Y.-J., Schweigert, J., Liu, H., 2013. An ecosystem modelling framework for incorporating climate regime shifts into fisheries management. Prog. Oceanogr. 115, 53–64.

Fulton, E.A., Parslow, J.S., Smith, A.D., Johnson, C.R., 2004. Biogeochemical marine ecosystem models II: the effect of physiological detail on model performance. Ecol. Model. 173, 371–406.

Fulton, E.A., Smith, A.D.M., Smith, D.C., 2007. Alternative Management Strategies for Southeast Australian Commonwealth Fisheries: Stage 2: Quantitative Management Strategy Evaluation. Australian Fisheries Management Authority. Fisheries Research and Development Corporation.

Garcia, S.M., Cochrane, K.L., 2005. Ecosystem approach to fisheries: a review of implementation guidelines. ICES J. Mar. Sci. 62, 311–318.

Gledhill, C.T., Ingram Jr., G.W., Rademacher, K.R., Felts, P., Trigg, B., 2005. NOAA Fisheries Reef Fish Video Surveys: Yearly Indices of Abundance for Gag (*Mycteroperca microlepis*). Report prepared for SouthEast Data Assessment and Review (SEDAR).

GMFMC, 2006. Options Paper Amendment 15 to the Shrimp Fishery Management Plan (including Environmental Impact Statement and Regulatory Impact Review). Gulf of Mexico Management Council, Tampa, FL.

GSMFC, 2011. SEAMAP Environmental and Biological Atlas of the Gulf of Mexico, 2009. Gulf States Marine Fisheries Commission.

Guisan, A., Edwards Jr., T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89–100.

Hansen, D.J., 1969. Food, growth, migration, reproduction, and abundance of pinfish, Lagodon rhomboides, and Atlantic Croaker, *Micropogon undulatus*, near Pensacola, Florida, 1963–65. Fish. Bull. 68, 135–146.

Hastie, T., 2006. GAM: Generalized Additive Models. R Package Version 0.98.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman and Hall, London, UK, 335 pp.

Henwood, T., Ingram, W., Grace, M., 2006. Shark/snapper/grouper longline surveys. In: Report Prepared by the NOAA/NMFS/SEFSC Pascagoula Lab, Mississipi, p. 22.

Horne, P.J., Kaplan, I.C., Marshall, K.N., Levin, P.S., Harvey, C.J., Hermann, A.J., Fulton, E.A., 2010. Design and Parameterization of a Spatially Explicit Ecosystem Model of the Central California Current. U.S. Dept. Commer., NOAA Tech. Memo. NMFS-NWFSC-104, 140 pp.

Ingram, W., Henwood, T., Grace, M., Jones, L., Driggers, W., Mitchell, K., 2005. Catch Rates, Distribution and Size Composition Of Large Coastal Sharks Collected During NOAA Fisheries Bottom Longline Surveys from the US Gulf of Mexico and US Atlantic Ocean, LCS05/06-DW-27.

Jenkins, C., 2011. Dominant Bottom Types and Habitats in Gulf of Mexico Data Atlas [Internet]. Stennis Space Center (MS): National Coastal Data Development Center.

Kaplan, I.C., Levin, P.S., Burden, M., Fulton, E.A., 2010. Fishing catch shares in the face of global change: a framework for integrating cumulative impacts and single species management. Can. J. Fish. Aquat. Sci. 67, 1968–1982.

Karnauskas, M., Schirripa, M.J., Kelble, C.R., Cook, G.S., Craig, J.K., 2013. Ecosystem Status Report for the Gulf of Mexico. NOAA Technical Memorandum NMFS-SEFSC-653, 52 pp.

Koubbi, P., Loots, C., Cotonnec, G., Harlay, X., Grioche, A., Vaz, S., Walkey, M., Carpentier, A., 2006. Spatial patterns and GIS habitat modelling of *Solea solea, Pleuronectes flesus* and *Limanda limanda* fish larvae in the eastern English Channel during the spring. Sci. Mar., 147–157.

Levin, P.S., Fogarty, M.J., Murawski, S.A., Fluharty, D., 2009. Integrated ecosystem assessments: developing the scientific basis for ecosystem-based management of the ocean. PLoS Biol. 7, e1000014.

Levin, P.S., Kelble, C.R., Shuford, R.L., Ainsworth, C., Dunsmore, R., Fogarty, M.J., Holsman, K., Howell, E.A., Monaco, M.E., Oakes, S.A., 2013. Guidance for implementation of integrated ecosystem assessments: a US perspective. ICES J. Mar. Sci., http://dx.doi.org/10.1093/icesjms/fst112.

Loots, C., Koubbi, P., Duhamel, G., 2007. Habitat modelling of *Electrona antarctica* (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. Polar Biol. 30, 951–959.

Loots, C., Vaz, S., Planque, B., Koubbi, P., 2010. What controls the spatial distribution of the North Sea plaice spawning population? Confronting ecological hypotheses through a model selection framework. ICES J. Mar. Sci. 67, 244–257.

Marzloff, M., Shin, Y.J., Tam, J., Travers, M., Bertrand, A., 2009. Trophic structure of the Peruvian marine ecosystem in 2000–2006: insights on the effects of management scenarios for the hake fishery using the IBM trophic model Osmose. J. Mar. Syst. 75, 290–304.

McLeod, K., Leslie, H., 2009. Ecosystem-based Management for the Oceans. Cambridge University Press, Cambridge, UK.

Mulholland, R., 1984. Habitat Suitability Index Models: Pink Shrimp. U.S. Fish and Wildlife Service FWS/OBS-82/10.76, 17 pp.

NOAA, 2010. Annual Commercial Landings Statistics. National Marine Fisheries Service. National Oceanic and Atmospheric Administration, Electronic Database. Available at http://www.st.nmfs.noaa.gov/st1/commercial/landings/annual_landings.html (accessed 27.10.12).

NOS, 2008. Gulf of Mexico at a Glance: U.S. National Ocean Service. Department of Commerce, National Oceanic and Atmospheric Administration, Washington, DC.

Okey, T.A., Vargo, G.A., Mackinson, S., Vasconcellos, M., Mahmoudi, B., Meyer, C.A., 2004. Simulating community effects of sea floor shading by plankton blooms over the West Florida Shelf. Ecol. Model. 172, 339–359.

Paris, C.B., Helgers, J., Van Sebille, E., Srinivasan, A., 2013. Connectivity modeling system: a probabilistic modeling tool for the multi-scale tracking of biotic and abiotic variability in the ocean. Environ. Model. Softw. 42, 47–54.

Pikitch, E., Santora, E.A., Babcock, A., Bakun, A., Bonfil, R., Conover, D.O., Dayton, P., Doukakis, P., Fluharty, D., Heheman, B., 2004. Ecosystem-based fishery management. Science 305, 346–347.

Planque, B., Bellier, E., Lazure, P., 2007. Modelling potential spawning habitat of sardine (*Sardina pilchardus*) and anchovy (*Engraulis encrasicolus*) in the Bay of Biscay. Fish. Oceanogr. 16, 16–30.

Rester, J.K., 2011. SEAMAP environmental and biological atlas of the Gulf of Mexico, 2009. Gulf States Marine Fisheries Commission, Ocean Springs, MS.

Samhouri, J.F., Haupt, A.J., Levin, P.S., Link, J.S., Shuford, R., 2013. Lessons learned from developing integrated ecosystem assessments to inform marine ecosystem-based management in the USA. ICES J. Mar. Sci., http://dx.doi.org/10.193/icesjms/fst141.

SEDAR, 2011. Fishery-independent Sampling: SEAMAP Trawl. SEDAR27-RD05. SEDAR, North Charleston, SC, 7 pp.

Shin, Y.J., Cury, P., 2001. Exploring fish community dynamics through size-dependent trophic interactions using a spatialized individual-based model. Aquat. Living Resour. 14, 65–80.

Shin, Y.J., Cury, P., 2004. Using an individual-based model of fish assemblages to study the response of size spectra to changes in fishing. Can. J. Fish. Aquat. Sci. 61, 414–431.

Shin, Y.J., Shannon, L.J., Cury, P.M., 2004. Simulations of fishing effects on the southern Benguela fish community using an individual-based model: learning from a comparison with Ecosim. Afr. J. Mar. Sci. 26, 95–114.

Steidinger, K.A., 2009. Historical perspective on *Karenia brevis* red tide research in the Gulf of Mexico. Harmful Algae 8, 549–561.

Travers, M., Shin, Y.J., Jennings, S., Machu, E., Huggett, J.A., Field, J.G., Cury, P.M., 2009. Two-way coupling versus one-way forcing of plankton and fish models to predict ecosystem changes in the Benguela. Ecol. Model. 220, 3089–3099.

Vaz, S., Pavoine, S., Koubbi, P., Loots, C., Coppin, F., 2006. Spatio-temporal Characteristics of Fish Populations in Relation to Environmental Forcing Functions as a Component of Ecosystem-based Assessment: Effects on Catchability, ICES CM 2006/O: 06.

Walters, C., Christensen, V., Walters, W., Rose, K., 2010. Representation of multi-stanza life histories in Ecospace models for spatial organization of ecosystem trophic interaction patterns. Bull. Mar. Sci. 86, 439–459.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F., Lindenmayer, D.B., 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecol. Model. 88, 297–308.

Wood, S.N., 2006. Generalized Additive Models: An Introduction With R. Chapman & Hall, London, UK.

Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression models for count data in R. J. Stat. Softw. 27, 1–25.