

## Use of Remotely-Derived Bathymetry for Modelling Biomass in Marine Environments

MAIGORZATA M. WIECZOREK,<sup>1</sup> WALDEMAR A. SPALLEK,<sup>1</sup> TOMASZ NIEDZIELSKI,<sup>1,2</sup> JASMIN A. GODBOLD,<sup>3,4</sup> and IMANTS G. PRIEDE<sup>3</sup>

**Abstract**—The paper presents results on the influence of geometric attributes of satellite-derived raster bathymetric data, namely the General Bathymetric Charts of the Oceans, on spatial statistical modelling of marine biomass. In the initial experiment, both the resolution and projection of the raster dataset are taken into account. It was found that, independently of the equal-area projection chosen for the analysis, the calculated areas are very similar, and the differences between them are insignificant. Likewise, any variation in the raster resolution did not change the computed area. Although the differences were shown to be insignificant, for the subsequent analysis we selected the cylindrical equal area projection, as it implies rectangular spatial extent, along with the automatically derived resolution. Then, in the second experiment, we focused on demersal fish biomass data acquired from trawl samples taken from the western parts of ICES Sub-area VII, near the sea floor. The aforementioned investigation into processing bathymetric data allowed us to build various statistical models that account for a relationship between biomass, sea floor topography and geographic location. We fitted a set of generalised additive models and generalised additive mixed models to combinations of trawl data of the roundnose grenadier (*Coryphaenoides rupestris*) and bathymetry. Using standard statistical techniques—such as analysis of variance, Akaike information criterion, root mean squared error, mean absolute error and cross-validation—we compared the performance of the models and found that depth and latitude may serve as statistically significant explanatory variables for biomass of roundnose grenadier in the study area. However, the results should be interpreted with caution as sampling locations may have an impact on the biomass–depth relationship.

**Key words:** GEBCO, bathymetry, biomass modelling, projection, resolution, statistical model.

### 1. Introduction

Biogeoscience, which combines biology with broadly understood Earth sciences, is steadily confirming its significance as a discipline and goes far beyond traditional biogeography. Biological processes are influenced by distance in accordance with the first law of geography as postulated by TOBLER (1970), but are also driven by a variety of geospatial variables. These variables are observable and are an expression of both the spatial and temporal dynamics of the physical Earth. The corresponding data may be obtained using in situ measurements and, increasingly, through remote sensing. Not uncommonly, the observations are records of various geophysical processes. The integration of up-to-date geophysical data with modern biological analyses provides us with new tools that may support and enhance classical studies, such as those stated in The Theory of Island Biogeography (MACARTHUR and WILSON 1967).

Geophysics itself is now defined in several ways, one of which states that it is the use of physical methods and data to interpret the Earth as a system. For decades, scientists have been trying to combine geophysical and biological concepts. For instance, there have been numerous studies looking at how animals respond to the dynamics of various geophysical fields. GRIFFIN (1969) considered a set of signals from both atmosphere and the solid Earth to understand how birds navigate, especially in nocturnal conditions. Later, and in a slightly different context, PRIEDE (1984) pioneered satellite tracking of fish, a technique that enabled the study of fish motion in light of numerous physical features of the ocean such as along thermal fronts (PRIEDE and MILLER 2009). Another example of how pure geophysics of the solid Earth interacts with biology is a recent study

<sup>1</sup> Institute of Geography and Regional Development, University of Wrocław, pl. Uniwersytecki 1, 50-137 Wrocław, Poland. E-mail: malgorzata.wieczorek@uni.wroc.pl

<sup>2</sup> Space Research Centre, Polish Academy of Sciences, ul. Bartycka 18A, 00-716 Warsaw, Poland.

<sup>3</sup> Oceanlab, University of Aberdeen, Newburgh, Aberdeenshire AB41 6AA, UK.

<sup>4</sup> National Oceanography Centre, Southampton, University of Southampton, Waterfront Campus, European Way, Southampton SO14 3ZH, UK.

by CHARZEWSKA *et al.* (2010). The authors found that sunflower autonomic movements are coherent with semidiurnal and diurnal tides as well as with plumb line variations. The list of such associations might be easily extended. However, a clear message is that geophysical data and methods are very useful for biological investigations into animal behaviour and distribution, and it is within this context that our present study is placed.

There are numerous studies on how marine habitats, considering their spatial and temporal variability, are a reflection of sea floor topography (KENNY *et al.* 2003). One of the major challenges facing management of resources in the oceans is the continued depletion of fish stocks around the world. There is strong evidence that as fisheries in coastal waters have declined (JACKSON *et al.* 2001; CHRISTENSEN *et al.* 2003) the industry has moved to progressively deeper waters increasing the mean depth of catch by 32 m for each decade in the North Atlantic Ocean since 1950 (MORATO *et al.* 2006). However, it is very difficult to assess the size and status of deep-sea fish resources. Most deep-sea species are confined to characteristic depth zones (REX and ETTER 2010). In bottom-living fishes, this kind of zonation is manifested as a succession of different species with increasing depth (PRIEDE *et al.* 2010). As depth has been found to be a strong predictor of probability of species occurrence and relative abundance, using information on patterns of species abundance with depth it is possible to estimate stock biomass in a given area by extrapolation using known bathymetry.

Associated with this is the search for the most suitable statistical models that may be used to predict biomass as a function of depth and other explanatory variables. The development of such models is now feasible using global bathymetric data, which have become available through remote sensing, both from ships towards the sea floor and from satellites towards the Earth. Hence, global bathymetry grids (MARKS and SMITH 2006) include the inverse information about the altimeter-derived Earth's gravity field. Accurate bathymetry is critical for appropriateness of the aforementioned models and thus recent progress towards the 30-arc seconds solution of the General Bathymetric Charts of the Oceans (GEBCO), with

properly combined data from ship-mounted sonars and sensors integrated on satellites, offers new perspectives for modelling marine environments at a range of spatial scales.

This paper presents results of research on the influence of geometric attributes of raster bathymetric dataset (GEBCO) (HALL 2006) on spatial statistical modelling of fish biomass. In particular, the aim of the research was to identify the changes of biomass of bottom-living fishes in the Northeastern Atlantic Ocean from pre-commercial trawling levels (1977–1989) to the post-commercial trawling period (2000–2002). Data were modelled from scientific trawl samples taken from the western parts of ICES Sub-area VII (Fig. 1). In order to calculate the biomass of demersal fish (tonnes), we needed to determine the relationship for biomass ( $\text{kg km}^{-2}$ ) as a function of depth. Then knowing depth across the ICES Sub-area VII total biomass can be estimated by integrating values of biomass multiplied by the area ( $\text{km}^2$ ). The result can then be compared with biomass estimates obtained using standard methods used in fishery management. Typically, fishery scientists use cohort or virtual population analysis (HILBORN and WALTERS 1992) derived from commercial landings statistics and population age structure data to reconstruct the stock biomass from which those landings were derived. The International Council for Exploration of the Sea (ICES 2012) reports landings data and stock biomass estimates each year for deep water stocks in the Northeastern Atlantic Ocean by sub-area. Our approach also enables estimation of biomass of species not landed by commercial fishermen.

In the first part of the paper, we examine impact of cartographic projection (we tested nine equal area projections) and resolution (three different resolutions) on the prediction of total fish biomass in ICES Sub-area VII using a linear model with generalised least squares estimation (GODBOLD *et al.* 2013). The aim of these analyses is to verify whether choice of cartographic projection and raster resolution influences biomass estimates. The objective of the second part of the paper is twofold. Firstly, using in situ trawling data obtained in two distinct periods in the Porcupine Seabight (the Northeastern Atlantic), we fit various generalised additive models (GAM) and generalised additive mixed models (GAMM) that aim

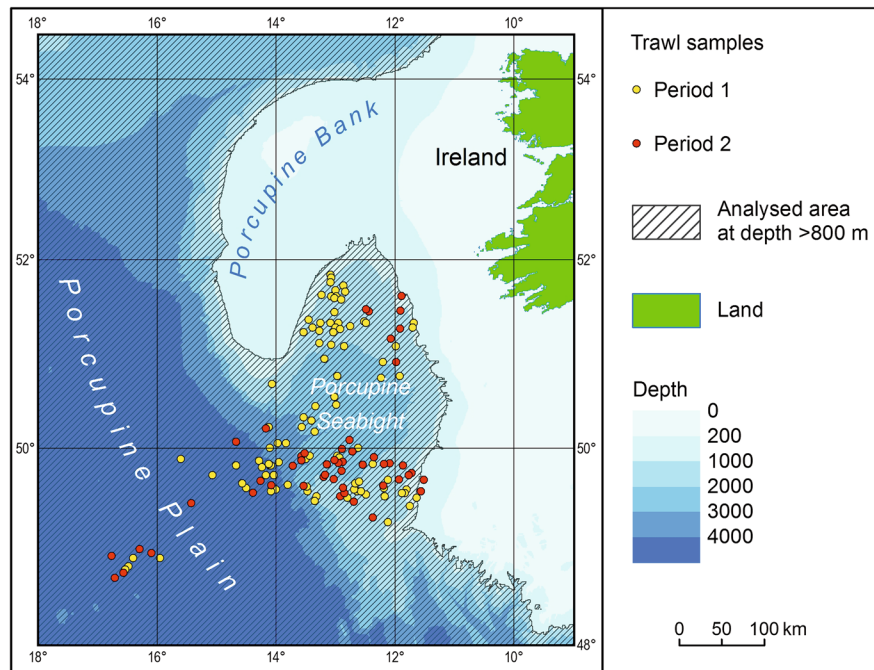


Figure 1

Locations of the trawl sample points in ICES Sub-area VII, divisions b, c, j and k. *Period 1*, before the advent of a commercial deep water fishery, corresponds to years 1977–1989; and *period 2*, during the commercial fishery, corresponds to years 2000–2002. The isobath 800 m indicates the minimum depth sampled in this study

to predict *Coryphaenoides rupestris* (roundnose grenadier) biomass as a function of explanatory variables which are chosen from 1-, 2- and 3-dimensional spaces. GODBOLD *et al.* (2013) proposed a straightforward one-dimensional biomass–depth relationship, and by fitting a spatial autocorrelation structure within GAMM considered variation in the spatial distribution of *C. rupestris* across the Porcupine Seabight. Herein, we extend that approach and statistically examine more complicated models, with latitude and/or longitude as explanatory variables, which may improve prediction accuracy. Secondly, the most suitable models are then run using a set of values (of depth, latitude or/and longitude), which were derived from a randomised GEBCO bathymetry extracted for the Porcupine Seabight. As a result, we are able to migrate to geospatial modelling with geographic information system (GIS) techniques and use actual sea floor topography to estimate total biomass in a given marine environment. The entire study aims to provide a better understanding of environmental conditions, mostly

based on the topography of the sea floor, that control distribution of fish.

## 2. Materials

### 2.1. Global Bathymetry

Mapping of deep waters offshore using conventional sounding methods has been hampered by sparse data sets with errors in navigation, transcription and digitisation. Major progress has been made through the use of satellite location provided by the Global Positioning System (GPS), which has reduced navigational errors in conventional ship-borne surveys. A further step has been the derivation of bathymetry from satellite microwave altimetry. Since the early 1970s, altimetric satellites have been providing a wealth of geophysical data on the spatial and temporal variability of the sea level and, as a consequence, altimeter-derived gravity field of the oceans. Early altimetric missions—Skylab, Geos-3, Seasat and Geosat—did not guarantee the accuracy

required by the detailed oceanographic studies. However, the launch of TOPEX/Poseidon in 1992 commenced the era of accurate altimetry (FU *et al.* 1994). Its superior performance was later continued by its successor, Jason-1 (PERBOS 2003). At present, a continuous time series of sea surface topography is being maintained by Jason-2 (LAMBIN *et al.* 2010) and, additionally, by Jason-1, Cryosat-2 and HY-2. Altimetry can be utilised to produce sea floor bathymetry data through application of inversion algorithms (DIXON *et al.* 1983; BAUDRY and CALMANT 1991; CALMANT and BAUDRY 1996) based on the fact that depth influences local gravity. The description of the problem may be summarised as follows: “depth variations of the seafloor can be considered as height variations of mass elements the density  $\Delta\rho$  of which is given by the contrast between rock and sea water densities” (CALMANT and BAUDRY 1996). This concept was used by HAXBY *et al.* (1983) and HAXBY (1985) who first produced a global map of the marine gravity field. There are a few procedures that allow us to invert satellite altimetric data to sea floor topography data, and the development of these methods was initiated by DIXON *et al.* (1983). A review of more advanced approaches was presented by CALMANT and BAUDRY (1996).

SMITH and SANDWELL (1997) produced a global digital bathymetric map of the oceans with a horizontal resolution of 2 arc-minutes by combining high-resolution marine gravimetry from Geosat and ERS-1 with carefully validated ship-borne depth soundings. Depth soundings were used to correct inversion algorithms for sediment thickness and substrate density variations. The high resolution achieved revealed important previously unknown features such as seamounts.

Satellite-derived bathymetry is now incorporated into the General Bathymetric Chart of the Oceans (GEBCO) produced under the auspices of the International Hydrographic Organization (IHO) and the Intergovernmental Oceanographic Commission (IOC) of UNESCO (MARKS and SMITH 2006). Satellite data are used to interpolate between sparse depth soundings. Inversion algorithms are in turn calibrated by available soundings. Since 2010, GEBCO bathymetry has become available for public download at a global resolution of 30 arc-seconds

either the GEBCO official webpage or the website of the British Oceanographic Data Centre (BODC) (The GEBCO\_08 Grid). The same resolution of bathymetric data is available from yet another global dataset, namely SRTM30\_PLUS derived from the Shuttle Radar Topography Mission (SRTM) that flew onboard the Space Shuttle Endeavour in 2000 (BECKER *et al.* 2009).

In the context of the present study, the resolution in question plays a key role in the performance of our statistical models. Indeed, sea floor topography is assumed as a geophysical field that, along with latitude and longitude, creates a framework for multivariate modelling.

## 2.2. The Study Area

The study area covers the Sub-area VII delineated by the International Council for the Exploration of the Sea (ICES) in the North-East Atlantic Ocean, south west of Ireland. For the present analyses, only the area with depth greater than 800 m was taken into account, as these are the depths from which the deep-sea fisheries operate. Hence ICES Sub-area VII divisions a, d, e, f, g, h, which are all too shallow, were excluded and the model was confined to divisions b, c, j and k, which are located in the western part of ICES Sub-area VII. These divisions form an ellipsoidal trapezoid, limited by the following parallels:  $B_1 = 48^{\circ}00'N$ ,  $B_2 = 54^{\circ}30'N$ , and meridians:  $L_1 = 9^{\circ}00'W$ ,  $L_2 = 18^{\circ}00'W$  (Fig. 1).

## 2.3. Trawling Data

The deep-sea demersal fishes of the Porcupine Seabight and Abyssal Plain areas of the Northeast Atlantic Ocean (approx.  $50^{\circ}N$ ,  $13^{\circ}W$ ) were surveyed by scientific bottom trawl from 1977 to 2002. For the present analysis we used trawls at depths ranging from 800 to 4,865 m [=146 trawls, GODBOLD *et al.* (2013)], which is a subset of the full data described by BAILEY *et al.* (2009) and PRIEDE *et al.* (2010). Trawls from depths  $<800$  m were excluded from the analysis as no shallow water trawling was carried out after 1989. In addition, eight trawls from 1997 were omitted from the analysis, because these trawls lacked ‘time on bottom’ data, which are required to

calculate fish biomass ( $\text{kg km}^{-2}$ ) and abundance (individuals  $\text{km}^{-2}$ ) from trawl swept-area (calculated from time on bottom, vessel speed and door spread). The survey data were split into two distinct time periods, ‘early’ period from 1977 to 1989 (95 trawls) and ‘late’ period from 2000 to 2002 (51 trawls). The ‘early’ period represents the state of the deep-sea fish assemblage before and during the initial development of the commercial fishery in this area, whilst the ‘late’ period represents the time when the fishery was well established (BAILEY *et al.* 2009; PRIEDE *et al.* 2011).

### 3. GIS Methods

Biomass of all bottom-living fish species data were modelled by GODBOLD *et al.* (2013) and they found a linear relationship between total biomass and depth. As noted above, the biomass of fish ( $\text{kg km}^{-2}$ ) was multiplied by area ( $\text{km}^2$ ) to calculate the total biomass of demersal fish (in tonnes). All pixels in GEBCO raster dataset have the same size in degree (30 arc-seconds), but different areas. Hence, we needed to resample the raster dataset of ICES Sub-area VII onto an equal area projection, which has no area distortion (KENNEDY and KOPP 2000). Another problem was the choice of raster size for modelling spatial differentiation of biomass, as modifications of the resolution implied dissimilar estimates.

#### 3.1. Area Measurement

We tested a few equal-area projections, available in ArcView 9.3 by ESRI (Table 1), by measuring the area of ellipsoidal trapezoid extracted from GEBCO bathymetric dataset. The trapezoid was resampled to three different resolutions and projected into nine different equal area projections (Fig. 2). Within the first group of raster datasets, a variety of similar resolutions were tested (an average of 748 m). The latter number was automatically generated by ArcView software as optimized for a particular projection. The remaining resolutions, i.e., 500 and 1,000 m, were chosen to test raster datasets in smaller and bigger pixel size.

The obtained results (Table 2) were compared to the reference area, which was calculated using Eq. (1) for the area of the ellipsoidal trapezoid ( $P$ ) limited by parallels  $B_1$   $B_2$  and meridians  $L_1$   $L_2$  stored in the Geographical Coordinate System (GCS) based on the World Geodetic System (WGS) 1984 ellipsoid:

$$P = \frac{1}{2}a^2(1 - e^2)(L_2 - L_1) \left( \frac{\sin B}{1 - e^2 \sin^2 B} + \frac{1}{2e} \ln \frac{1 + e \sin B}{1 - e \sin B} \right) \Big|_{B_1}^{B_2} \quad (1)$$

where  $a$  is semi-major axis and  $e$  is eccentricity (BALCERZAK and PĘDZICH 2006).

The most promising results, which are highly similar to the reference area in WGS 1984, are obtained for the pseudocylindrical projection and the cylindrical equal area (CEA) projection (Table 2). For all of them, a difference between area of the ellipsoidal trapezoid on raster dataset and ellipsoid WGS 1984 is  $<0.2\%$  of the reference area. However, the difference for other projections was not significantly greater, i.e., the values do not exceed  $0.3\%$ . Hence, area measurement error was found to be negligible. Nevertheless, the cylindrical projection fits better to rectangular raster cells than the others because of the rectilinear and right-angled shape of cartographic network.

#### 3.2. Changes in Total Demersal Fish Biomass as Function of Bathymetry-Derived Depth

Application of three resolutions and nine equal area projections results in 27 different datasets of bathymetry for the study area. Since deep-sea fisheries do not operate at depths shallower than 800 m, only area at depths equal or deeper than 800 m below sea level were extracted from each of 27 raster datasets for ICES Sub-area VII. To predict the total biomass for both periods we used the following linear functions (GODBOLD *et al.* 2013):

$$\text{Period 1: } y = -0.2253x + 1,351, \quad (2)$$

$$\text{Period 2: } y = -0.1261x + 853.27, \quad (3)$$

where  $y$  is total biomass ( $\text{kg km}^{-2}$ ), and  $x$  is depth (m).

Table 1

*Attributes of equal area projections used in the experiment*

Projection	Type by projection surface	Aspect	Standard parallels*
Lambert azimuthal	Azimuthal	Tangent oblique	51°N**
Albers	Conical	Secant normal	49.5°N and 53°N
Bonne	Pseudoconic	Tangent normal	51°N
Lambert cylindrical	Cylindrical	Tangent normal	0°N
Behrmann	Cylindrical	Secant normal	30°S and 30°N
Cylindrical equal area	Cylindrical	Secant normal	51°S and 51°N
Eckert IV	Pseudocylindrical	Tangent normal	0°N
Eckert VI	Pseudocylindrical	Tangent normal	0°N
Mollweide	Pseudocylindrical	Tangent normal	0°N

\* Longitude of central meridian for all projections is 13.5°W

\*\* Latitude of projection centre

In order to model the spatial distribution of all demersal fish biomass in the study area between the two periods (Fig. 3), the equations were applied to GEBCO bathymetry using Map Algebra tool in ArcView 9.3 in combination with nine different projections and three different resolutions. As indicated by many authors (e.g., NYERGES and JANKOWSKI 1989; KHALID 2006), the problem of selecting an appropriate projection remains an ongoing challenge. In particular, it has been confirmed that dissimilar equal-area projections may lead to results that are unlike each other (USERY and SEONG 2001), and the differences are driven by numerous factors. Hence, we decided to focus on that problem in the context of bathymetry data. In practise, it is often unclear which projection should be used and how our choice influences the estimates of the area. Therefore, we performed the exercise, the aim of which was to show the real impact of projection on calculation of area and, subsequently, on computation of total fish biomass. The best solution to calculate the total biomass of demersal fish (tonnes) was to multiply the mean value of biomass ( $\text{kg km}^{-2}$ ) by the area ( $\text{km}^2$ ) at depths >800 m. The mean values were extracted using Raster Dataset Properties in ArcView 9.3. As a result, we obtained 54 values of demersal fish total biomass in tonnes. A descriptive overview of the results on the statistics that were computed is shown in Table 3. Regardless of the choice of projection, the biomass values have been found to be very similar. The differences, regardless of resolutions and periods, do not exceed 0.4 % of the minimum and

maximum values. However, the lowest differentiation of the obtained biomass estimates is typical for models based on the automatically generated resolution. As expected, highly dissimilar, and probably the worst results were obtained at 1,000 m resolution, but results based on the automatically generated resolution and the resolution bigger than 500 m were highly similar (Table 3). This suggests that increasing spatial resolution does not significantly modify modelling results. Furthermore, higher resolution may increase the size of the raster files, and thus data processing may become more time-consuming. Hence, the best solution is to choose a raster resolution generated by aforementioned software.

#### 4. Statistical Models and Spatial Modelling

Lack of complex/adequate/representative data is typical for investigations into changing components of environment, especially for marine environments. If they exist, they are often constrained because they are:

- initially collected for other reasons,
- collected at different times,
- collected when the conditions (weather, ship route as well as trawl route) allowed, and therefore samples cover the study area in a spatially random fashion.

Hence, there are considerable difficulties in assessing the state of phenomena across different time periods, and consequently, their changes are tough to measure.

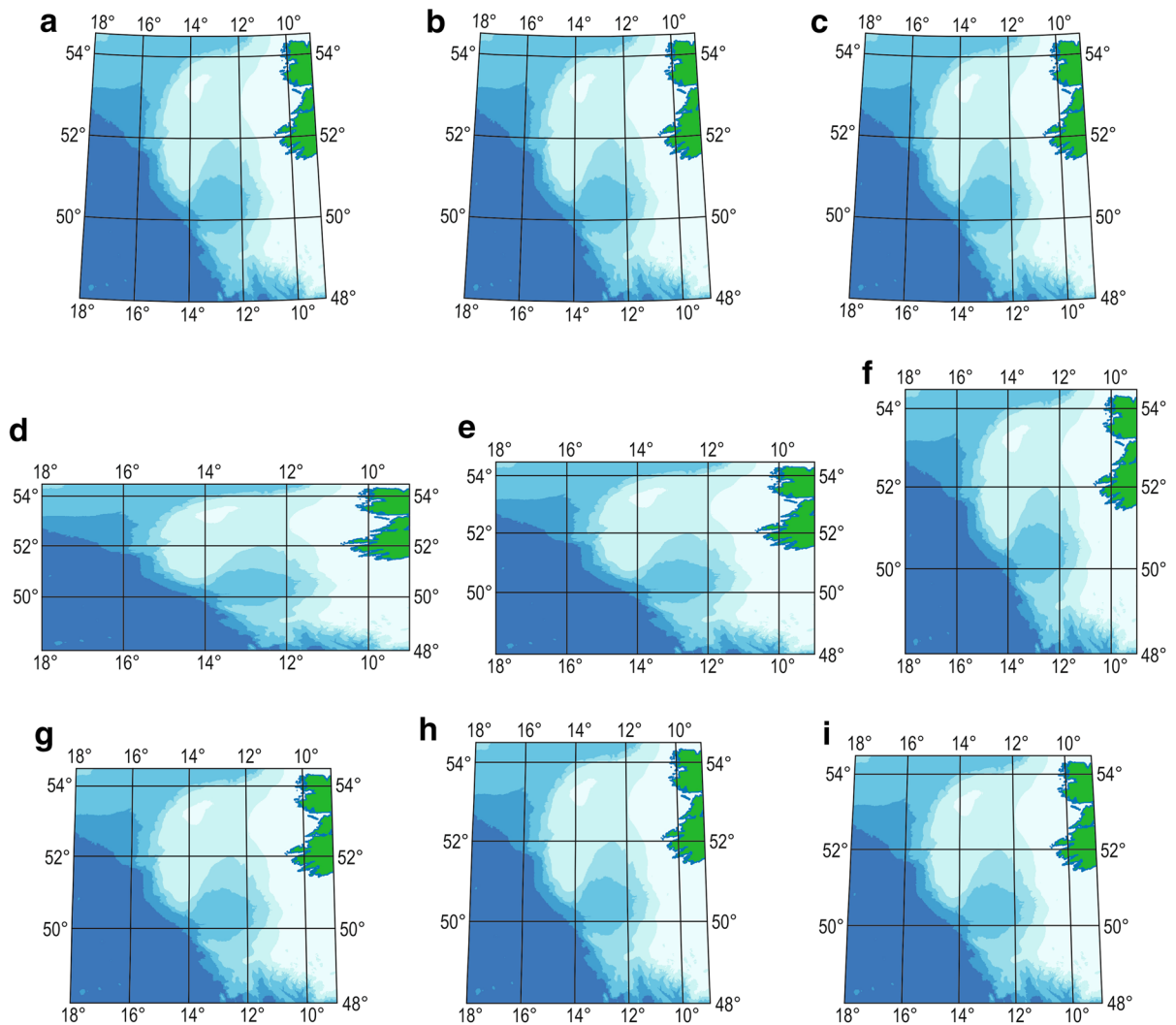


Figure 2

Maps of the ICES Sub-area VII (divisions b, c, j and k) projected in nine various equal area projections: **a** Lambert azimuthal, **b** Albers conical, **c** Bonne pseudoconic, **d** Lambert cylindrical, **e** Behrmann cylindrical, **f** cylindrical equal area, **g** Eckert IV pseudocylindrical, **h** Eckert VI pseudocylindrical, **i** Mollweide pseudocylindrical. Scale is equal to 1:25,000,000

#### 4.1. Descriptive Statistics of Trawling Data

In the experiment, we used roundnose grenadier biomass data ( $\text{kg km}^{-2}$ ), which were collected in the same way as the total biomass of demersal fish data described above. Figure 4 shows the spatial distribution of the sample points. Maximum differences between latitude and longitude of sample points are the same, but in metric units they are different. Distance between the most southern and the most northern sample points is ca. 300 km and between the most eastern and the most western sample points is

ca. 180 km. Therefore, we tested not only depth impact on biomass but also focused on how geographical and metrical coordinates influence biomass estimates.

Roundnose grenadier is the main species targeted by the commercial fishery. Figure 5 shows biomass values for the species collected during the sampling procedure. Depth range of the sample points is from 804 to 1,932 m. The maximum biomass of roundnose grenadier was collected at depth 1,360 m in period 1 ( $1,172.7 \text{ kg km}^{-2}$ ) and 1,541 m in period 2 ( $1,383.7 \text{ kg km}^{-2}$ ).

Table 2  
 Comparison of area of the ellipsoidal trapezoid for different projections and WGS 1984

Projection	Pixel size (m)	Area of ellipsoidal trapezoid (km <sup>2</sup> )	Difference between area on raster dataset and reference area (km <sup>2</sup> )	Difference as percent of reference area
Reference area	–	454,155.327	0	0
Lambert azimuthal	763.808	455,112.372	–957.045	–0.211
Albers	764.166	455,108.737	–953.410	–0.210
Bonne	763.727	455,122.007	–966.679	–0.213
Lambert cylindrical	734.215	455,077.818	–922.491	–0.203
Behrmann	734.215	455,079.974	–924.647	–0.204
Cylindrical equal area	734.215	454,857.877	–702.550	–0.155
Eckert IV	742.175	454,348.348	–193.021	–0.043
Eckert VI	747.747	454,652.009	–496.682	–0.109
Mollweide	749.388	454,336.206	–180.878	–0.040
Lambert azimuthal	500	455,121.500	–966.173	–0.213
Albers	500	455,119.000	–963.673	–0.212
Bonne	500	455,115.500	–960.173	–0.211
Lambert cylindrical	500	455,362.000	–1,206.673	–0.266
Behrmann	500	454,921.500	–766.173	–0.169
Cylindrical equal area	500	455,083.750	–928.423	–0.204
Eckert IV	500	454,517.000	–361.673	–0.080
Eckert VI	500	454,579.750	–424.423	–0.093
Mollweide	500	454,367.750	–212.423	–0.047
Lambert azimuthal	1,000	455,121.000	–965.673	–0.213
Albers	1,000	455,128.000	–972.673	–0.214
Bonne	1,000	455,122.000	–966.673	–0.213
Lambert cylindrical	1,000	455,362.000	–1,206.673	–0.266
Behrmann	1,000	455,356.000	–1,200.673	–0.264
Cylindrical equal area	1,000	455,040.000	–884.673	–0.195
Eckert IV	1,000	454,150.000	5.327	0.001
Eckert VI	1,000	454,582.000	–426.673	–0.094
Mollweide	1,000	454,720.000	–564.673	–0.124

#### 4.2. Statistical Inference

We fitted various GAM and GMM models to predict biomass of roundnose grenadier as a function of explanatory variables, which were chosen from one-, two- and three-dimensional spaces (ZUUR *et al.* 2007, 2009). Finally, we obtained 14 specific models, both based on GAM and GMM approaches (Table 4). Firstly, a straightforward one-dimensional biomass-depth relationship was prepared by GODBOLD *et al.* (2013) who used latitude and longitude for defining a spatial correlation kernel. In the latter approach, however, information on location was not explicitly included in explanatory variables. Fluctuations of the biomass (kg km<sup>–2</sup>) of roundnose grenadier with depth, between the early and the late periods, were analysed using GAM and GMM models, applied to log<sub>10</sub>-transformed data, with a Gaussian distribution. We used log<sub>10</sub> transformation to stabilise the variance and

reduce the effect of large values. Spatial autocorrelation was detected and the biomass data were analysed using GMM models with the Gaussian spatial correlation structure (corGauss) (PINHEIRO *et al.* 2008). Both GAM and GMM approaches included factor called ‘period’ (early or late) and a smoother over depth. To identify a presence of significant correlations between explanatory variables we used the analysis of variance (ANOVA).

As we were specifically interested in spatial distribution of biomass, the GAM and GMM models were extended by incorporating coordinates as potential explanatory variables. Hence, we obtained eight 2-dimensional models based on depth and one of the geographical coordinates measured in degrees (latitude or longitude) or rectangular coordinates measured in metres (*X* or *Y* according to geodetic notation). Furthermore, we prepared four three-dimensional



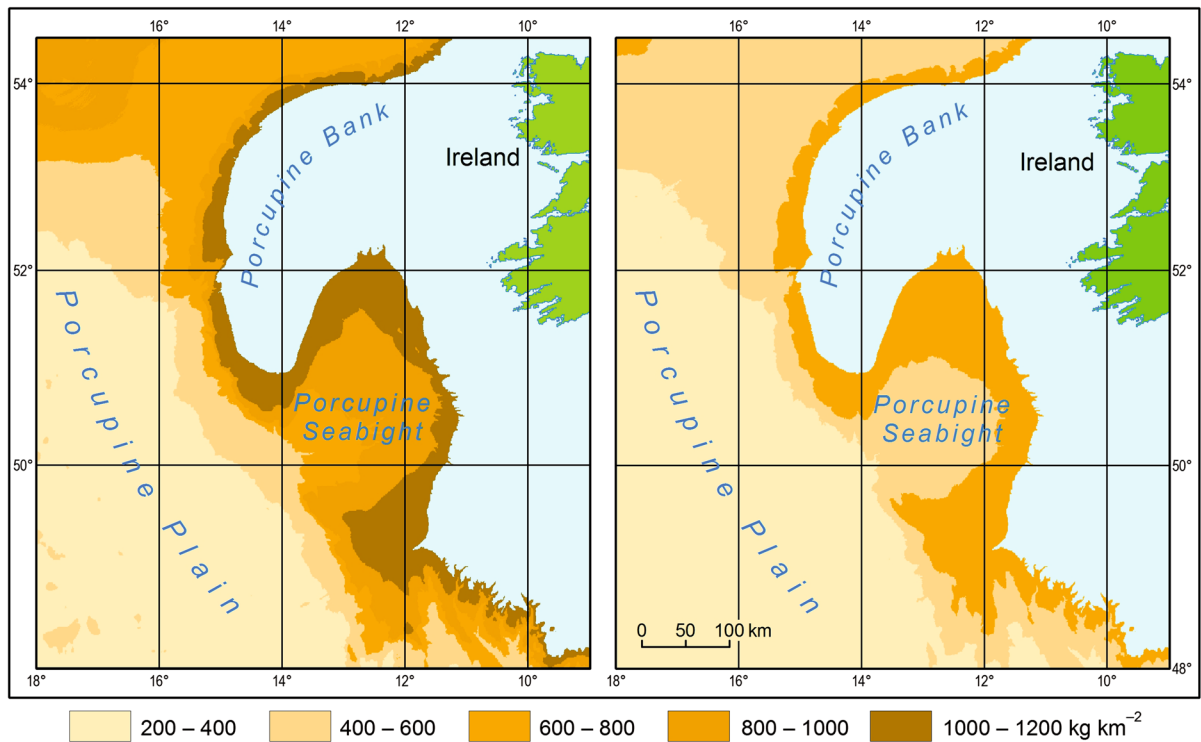


Figure 3

Example of biomass predictions based on the GIS model for spatial distribution of total fish biomass ( $\text{kg km}^{-2}$ ) in ICES Sub-area VII (divisions b, c, j, k) computed using Eqs. (2) and (3). Spatial data are projected onto the cylindrical equal area (CEA) projection. *Left figure* corresponds to period 1 (1997–1989), and *right figure* shows the results for period 2 (2000–2002) (the same as in GODBOLD *et al.* (2013) but in different projection)

Table 3

*Descriptive statistics of the modelled biomass of fish for dissimilar resolution and equal area projections*

Period	Resolution	Range (t)	Mean (t)	Standard deviation (t)	Range as a percent of minimum value	Range as a percent of maximum value
1	Automatic	312.26	169,720	121.00	0.1842	0.1838
	500 m	382.59	169,717	116.33	0.1938	0.1934
	1,000 m	517.90	169,775	146.87	0.3057	0.3047
2	Automatic	226.52	122,445	86.67	0.1852	0.1849
	500 m	243.73	122,446	83.00	0.1993	0.1989
	1,000 m	390.09	122,475	113.11	0.3191	0.3181

models, in which we took into account depth and full location of the trawling points, expressed in geographical or rectangular coordinates. All models were compared using the Akaike information criterion (AIC), root mean square error (RMSE) and mean absolute error (MAE). Analyses were conducted in R version 2.13.2 (The R Foundation for Statistical Computing 2011) using the ‘mgcv’ library for GAM and GAMM models (WOOD 2007; ZUUR *et al.* 2009).

### 5. Results and Discussion on Statistical Modelling

Firstly, having assumed the GIS-based prerequisites identified in Sect. 3, we aimed to check whether explanatory variables are significantly correlated with biomass. Therefore, we used ANOVA to check all models under study. As expected, depth is a significant explanatory variable for biomass, and this holds for all models. The analysis of variance showed that

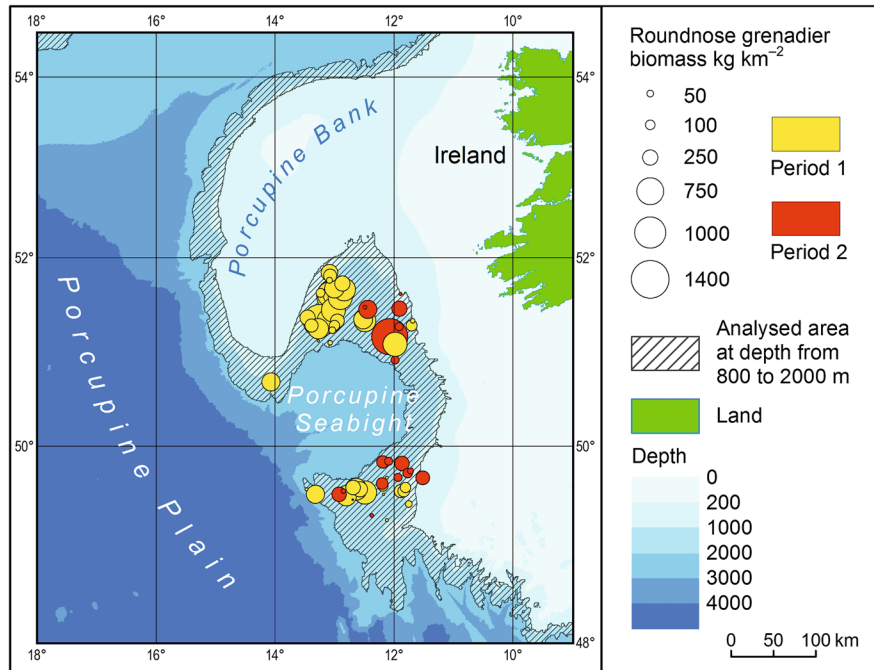


Figure 4

Locations of sample points at depth from 800 to 2,000 m with biomass of roundnose grenadier collected in the ICES Sub-area VII in early and late period

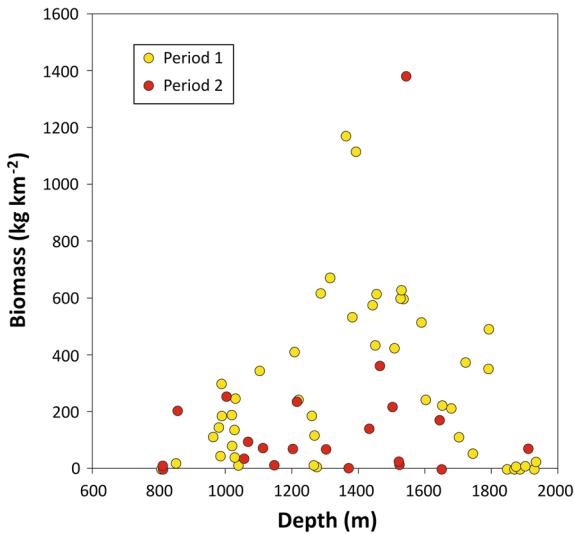


Figure 5

Relations of roundnose grenadier biomass ( $\text{kg km}^{-2}$ ) to depth in the Porcupine Seabight area of the Northeast Atlantic Ocean between the early (1977–1989, period 1) and late (2000–2002, period 2) trawling periods

longitude and the corresponding  $Y$  coordinates are insignificant, because both are correlated with depth (Table 4). However, as also inferred from Table 4,

longitude and  $Y$  coordinates were found to reveal a significant impact on biomass. According to AIC, we argue that the most appropriate approach is the 2-dimensional GAM based on depth and latitude (referred to as GAM\_depth\_lat) or  $Y$  coordinate (named also as GAM\_depth\_Y). It is worth noting that AIC criterion attained minimum for the three-dimensional GAM models, which additionally included longitude (known hereafter as GAM\_depth\_lat\_lon) or coordinate  $X$  (named as GAM\_depth\_Y\_X), but they had to be rejected due to high  $p$  value of these additional variables obtained in the ANOVA investigation. The one-dimensional model (known as GAMM\_depth) and the two-dimensional model based on depth and latitude (referred to as GAMM\_depth\_lat) may be identified as the most appropriate amongst GAMM models used in the study. To obtain two types of RMSE and MAE, we first calculated model residuals (RM) and subsequently utilised cross-validation (CV) based on leave-one-out procedure. According to RM, eight models (all three-dimensional and four two-dimensional based on the depth and latitude or coordinate  $Y$ ) have

Table 4  
ANOVA and errors for models of a relationship between biomass of roundnose grenadier and depth

Model	Variables	ANOVA		AIC	RM		CV	
		<i>F</i>	<i>p</i> value		RMSE	MAE	RMSE	MAE
GAM_depth	Depth	8.51	0.000	151.1	0.654	0.490	0.744	0.552
GAM_depth_lat	Depth	8.39	0.000	141.5	0.600	0.455	0.690	0.521
	Lat	11.56	0.001					
GAM_depth_lon	Depth	8.23	0.000	153.0	0.654	0.490	0.755	0.559
	Lon	0.13	0.715					
GAM_depth_lat_lon	Depth	8.46	0.000	142.5	0.595	0.450	0.699	0.525
	Lat	12.07	0.001					
	Lon	0.70	0.408					
GAM_depth_Y	Depth	8.38	0.000	141.5	0.600	0.455	0.690	0.521
	Y km	11.56	0.001					
GAM_depth_X	Depth	7.97	0.000	152.9	0.654	0.489	0.754	0.559
	X km	0.18	0.671					
GAM_depth_Y_X	Depth	8.46	0.000	142.5	0.596	0.450	0.699	0.525
	X km	0.69	0.409					
	Y km	12.01	0.001					
GAMM_depth	Depth	7.06	0.000	155.4	0.657	0.493	0.722	0.538
GAMM_depth_lat	Depth	7.79	0.000	154.0	0.604	0.461	0.680	0.515
	Lat	6.68	0.012					
GAMM_depth_lon	Depth	6.84	0.000	158.6	0.657	0.492	0.730	0.544
	Lon	0.03	0.874					
GAMM_depth_lat_lon	Depth	7.69	0.000	157.0	0.601	0.457	0.688	0.520
	Lat	6.60	0.013					
	Lon	0.30	0.589					
GAMM_depth_Y	Depth	6.42	0.000	160.2	0.606	0.465	0.680	0.518
	Y km	6.10	0.016					
GAMM_depth_X	Depth	5.82	0.000	163.2	0.657	0.496	0.730	0.548
	X km	0.00	0.960					
GAMM_depth_Y_X	Depth	6.07	0.000	171.4	0.600	0.460	0.688	0.522
	X km	0.46	0.502					
	Y km	6.12	0.016					

lower RMSE (by approximately 0.05) than the others. The corresponding MAE are smaller by about 0.035. The similar findings hold for CV-based errors. Indeed, the same eight models have smaller errors than the others (by about 0.04 for RMSE and about 0.02 for MAE).

Considering all the above-mentioned criteria we may conclude that in each group, GAM and GAMM, the best models are those based on depth and latitude (GAM\_depth\_lat and GAMM\_depth\_lat). In addition, AIC, RMSE and MAE calculated for model residuals are smaller for GAM model, however, GAMM models reveal smaller RMSE and MAE values for the cross-validation experiment. It is thus difficult to unequivocally recommend any of the two classes of additive models as tools for biomass-depth-location modelling.

For further investigation we selected models that include location as explanatory variable. These are: GAM\_depth\_lat and GAMM\_depth\_lat. For comparison, we used the one-dimensional models, namely GAM\_depth and GAMM\_depth. The models were used to compute the values of biomass (tonnes) in 100-m depth bins and incorporated into the GIS modelling following the same modelling procedure described in the first part of this paper for total demersal fish biomass.

In the one-dimensional models, calculation of biomass in 100-m depth bins was straightforward. In the case of the two-dimensional models sample points contained information about depth (100-m depth bins) and latitude, and the latter values were selected randomly from the GEBCO bathymetry. Figures 6 and 7 show the distribution of biomass depending on

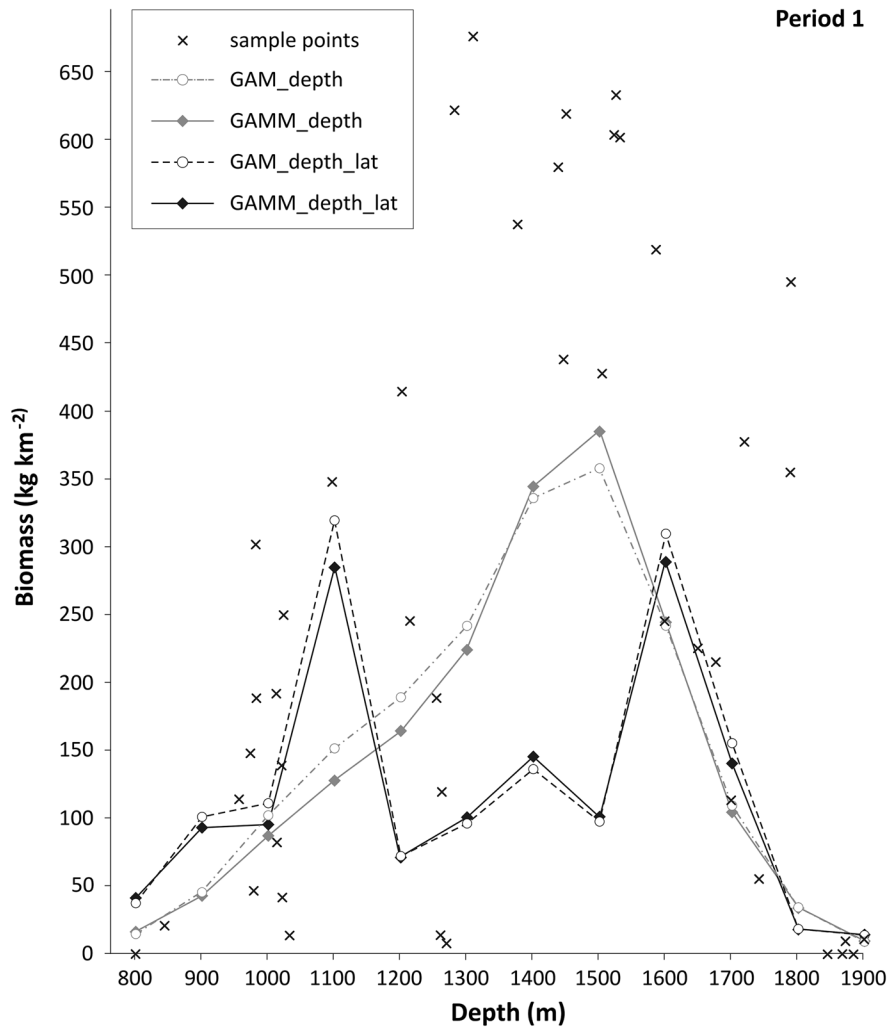


Figure 6

Biomass as a function of depth or depth and latitude obtained from models for period 1, a few extreme data have been excluded from the figure for the purpose of presentation—see Fig. 5 for comparison

depth or depth and latitude obtained from the models in two periods. The one-dimensional models predict a unimodal distribution of biomass as a function of depth, in contrast to the two-dimensional models that have a bimodal distribution. The unimodal distribution seems to better describe a relation between the biomass of roundnose grenadier and depth, because like most deep-sea species, grenadier are confined to a characteristic depth zone. The best models in terms of statistics are characterised by a bimodal distribution, but they seem to be incorrect in biological sense. Bimodal distribution probably results from the spatial distribution of sample points, which occur in two

groups-clusters, in the North and in the South of Porcupine Seabight (Fig. 4).

Biomass of the roundnose grenadier as a function of depth was found to be non-linear. Therefore, for estimating total biomass for both periods we used segmented regression models. Biomass values were calculated for each cell of the raster dataset, with depth  $\geq 800$  m (Fig. 8) and subsequently were converted to total roundnose grenadier biomass (in tonnes) by multiplying the mean biomass ( $\text{kg km}^{-2}$ ) by area at depths  $\geq 800$  m ( $\text{km}^2$ ). The integrated values of the biomass are presented in Table 5.

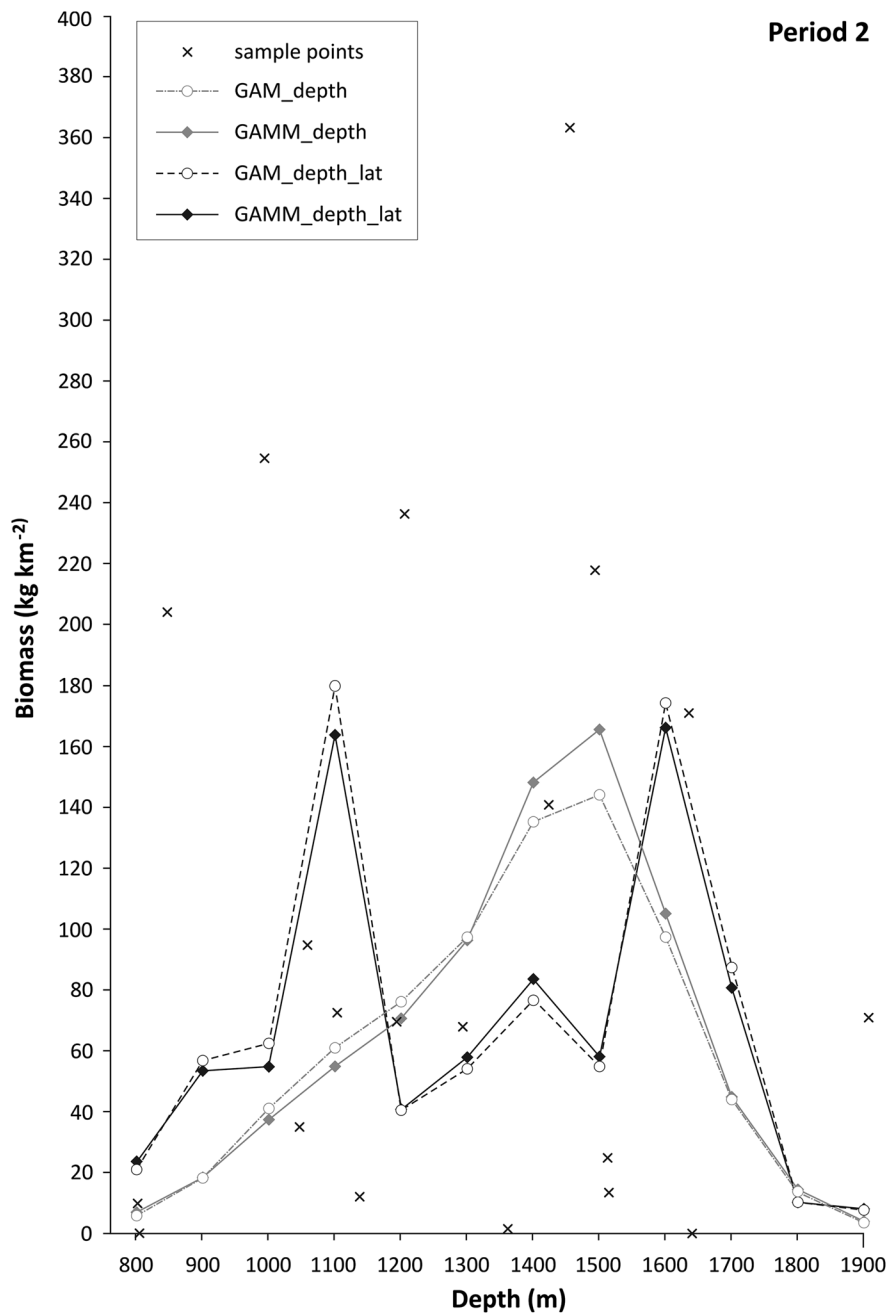


Figure 7

Biomass as a function of depth or depth and latitude obtained from models for period 2; for the purpose of presentation, a few extreme data have been excluded from the figure—see Fig. 5 for comparison

In period 1, the two-dimensional models give lower values of biomass than the 1-dimensional models, and they differ by about 17 % for GAM and 11.5 % for GAMM. In period 2, the two-dimensional models produce biomass value higher by about 16 %

for GAM and about 23 % for GAMM. Therefore, the differences of the estimated biomass of roundnose grenadier between the pre-and post-commercial fishing period in the particular models, vary up to 20 %. GAM\_depth model, which gives the extreme

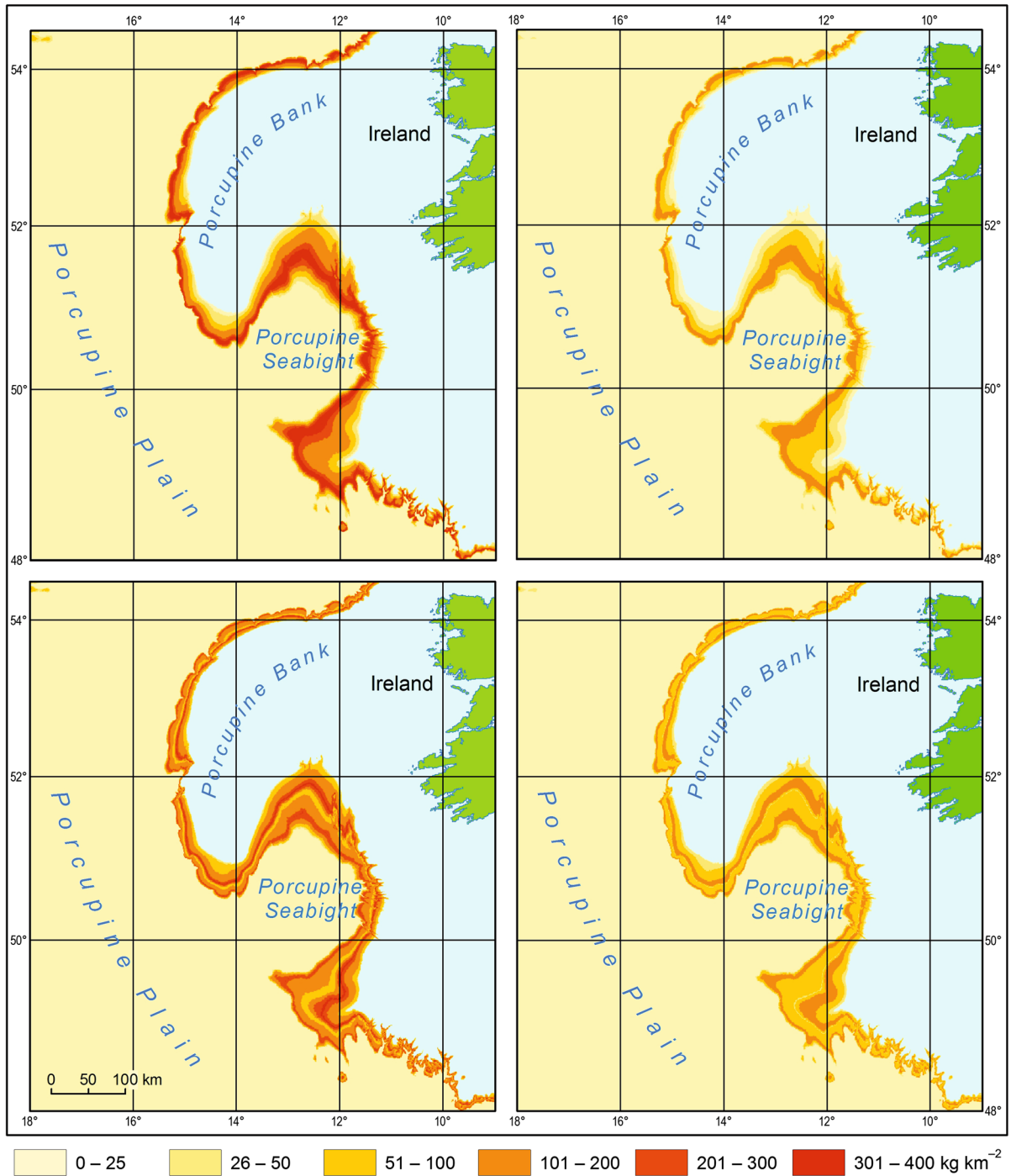


Figure 8

Examples of biomass predictions based on the GIS model for spatial distribution of total roundnose grenadier biomass ( $\text{kg km}^{-2}$ ) in ICES Sub-area VII (divisions b, c, j, k). *Left column* corresponds to period 1 (1997–1989), whereas *right column* corresponds to period 2 (2000–2002). *Top row* presents the distribution based on the GAMM\_depth model (Table 5) after Godbold et al. (2013), whereas the *bottom row* shows the results based on the GAMM\_depth\_lat model (Table 5)

Table 5

*Total biomass of roundnose grenadier in ICES Sub-area VII as results of GIS modelling*

Model	Biomass (t)		Difference between period 1 and 2	Decrease as % of the biomass in period 1
	Period 1	Period 2		
GAM_depth	7,815	3,143	4,672	60
GAM_depth_lat	6,480	3,641	2,838	44
GAMM_depth	7,525	3,232	4,294	57
GAMM_depth_lat	6,666	3,990	2,676	40

values of biomass in both periods and also the biggest difference, is the least reliable due to results of statistical inference. An important advantage of GAMM models is the possibility of taking into account spatial autocorrelation of the data, which is impossible in GAM models (ZUUR 2009).

## 6. Conclusions

In statistical modelling of satellite-derived bathymetric data the choice of equal area projection has no impact on the final estimate of area. The choice of resolution of the modelled bathymetric raster data seems to be more important if cell size is bigger than default. Otherwise resolution should not affect the calculation of areas. In spatial analyses, choice of statistical models cannot be based solely on statistical criteria, as well-fitted models do not always give better predictions. Indeed, ROBERTS and PASHLER (2000) discussed how crucial goodness-of-fit is and found that, in fact, it is more important to look at what the theory predicts. Not uncommonly conceptually and phenomenologically incorrect models produce acceptable predictions, as shown by LEFEBVRE *et al.* (1996) who focused on kriging-based predictions. It is therefore necessary to take into account the nature of a given phenomenon and its geophysical setting, which—in this study—is depth. In the deep-sea environment, besides specifying spatial extent in two dimensions, it is necessary to enter a range of depths. Outcomes of spatial models also depend on the abundance and distribution of the sample.

Finally, we obtained two models. The first one, GAMM\_depth, illustrates variation of grenadier biomass with depth and does it rather acceptably, but

does not take geographical differentiation into account. The second model, GAMM\_depth\_lat, is the best one from the statistical point of view, but—as it depends on latitude and hence is partially explained by zonal variation—is more sensitive to the spatial distribution of samples. Although the fit and model predictions computed for GAMM\_depth\_lat—also those based on a cross-validation procedure—are superior over the other models under study, the resulting distribution of roundnose grenadier as a function of depth is bimodal, with the first (lower) modal value around 1,000 m and the second (upper) mode at 1,600 m of depth. The bimodal distribution of roundnose grenadier is biologically improbable (PRIEDE *et al.* 2010; REX and ETTER 2010), and a potential explanation should be sought in the relatively limited number of data points available in each period. Indeed, the sparse data points are spatially clustered and thus it is difficult to unequivocally confirm the bimodal pattern of the distribution curve. However, given the predictive performance of the GAMM\_depth\_lat approach it is worth stating a working hypothesis for future investigations with the main message that it would be valuable to incorporate additional explanatory variables in order to improve the model predictions based on bathymetry data.

## Acknowledgments

This is a contribution to the EU-funded HERMIONE Project. We thank the ships' companies of the cruises that collected the raw data between 1977 and 2002. The research was partially supported from statutory funds of the Institute of Geography and Regional Development of the University of Wrocław, Poland. We wish to express our thanks to Dr. Alain Zuur who

offered us numerous valuable remarks on statistical modelling. Last but not least, we are grateful to Prof. Paweł Peździch for his comments on projections and their implications for area calculations.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## REFERENCES

- BAILEY, D.M., COLLINS, M.A., GORDON, J.D.M., ZUUR, A.F., and PRIEDE, I.G. (2009), *Long-term changes in deep-water fish populations in the northeast Atlantic: a deeper reaching effect of fisheries*, Proc. R. Soc. B, 276, 1965–1969.
- BALCERZAK, J., and PEDZICH, P., (2006), *Some methods of calculation of ellipsoidal polygons areas*, Annals of Geomatics, 4(3), 41–49 (in Polish).
- BAUDRY, N., and CALMANT, S., (1991), *3D Modelling of seamount topography from satellite altimetry*, Geophysical Research Letters 18, 1143–1146.
- 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., LADNER, R., MARKS, K., NELSON, S., PHARAOH, A., TRIMMER, R., VON ROSENBERG, J., WALLACE, G., and WEATHERALL, P., (2009), *Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30\_PLUS*, Marine Geodesy 32, 355–371.
- CALMANT, S., and BAUDRY, N., (1996), *Modelling bathymetry by inverting satellite altimetry data: a review*, Marine Geophysical Researches 18, 123–134.
- CHARZEWSKA, A., KOSEK, W., and ZAWADZKI, T., (2010), *Does circumnutation follow tidal plumb line variations?* Biological Rhythm Research 41, 449–455.
- CHRISTENSEN, V., GUÉNETTE, S., HEYMANS, J.J., WALTERS, C.J., WATSON, R., ZELLER, D., and PAULY, D., (2003), *Hundred year decline of North Atlantic predatory fishes*, Fish and Fish 4, 1–124.
- DIXON, T.H., NARAGHI, M., McNUTT, M.K., and SMITH, S.M., (1983), *Bathymetric predictions from SEASAT altimeter data*, Journal of Geophysical Research 88, 1563–1571.
- FU, L.-L., CHRISTENSEN, E.J., YAMARONE, C.A. JR., LEFEBVRE, M., MÉNARD, Y., DORRER, M., and ESCUDIER, P., (1994), *TOPEX/POSEIDON mission overview*, Journal of Geophysical Research 99(C12), 24369–24381.
- GODBOLD, J.A., BAILEY, D.M., COLLINS, M.A., GORDON, J.D.M., SPALLEK W.A., and PRIEDE, I.G., (2013) *Putative fishery-induced changes in biomass and population size structures of demersal deep-sea fishes in ICES Sub-area VII, North East Atlantic Ocean*. Biogeosciences 10, 529–539. doi:10.5194/bg-10-529-2013.
- GRIFFIN, D.R., (1969), *The physiology and geophysics of bird navigation*, The Quarterly Review of Biology 44, 255–276.
- HALL, J.K., (2006), *GEBCO Centennial Special Issue—Charting the secret world of the ocean floor: the GEBCO project 1903–2003*, Marine Geophysical Researches 27, 1–5.
- HAXBY, W.F., KARNER, G.D., LABRECQUE, J.L., and WEISSEL, J.K., (1983), *Digital images of combined oceanic and continental data sets and their use in tectonic studies*, EOS Trans. AGU 64, 995–1004.
- HAXBY, W.F., (1985), *Gravity Field of the World's Oceans*. US Navy Naval Office of Research (chart, scale 1:51,400,000).
- HILBORN, R., WALTERS, C., (1992), *Quantitative fisheries stock assessment: choice, dynamics and uncertainty*. New York, Chapman and Hall.
- ICES: Report of the Working Group on the Biology and Assessment of Deep-sea Fisheries Resources (WGDEEP), Denmark ICES document CM 2012/ACOM: 17, 2012.
- JACKSON, J.B.C., M.X. KIRBY, W.H. BERGER, K.A. BJORNDAAL, L.W. BOTSFORD, B.J. BOURQUE, R.H. BRADBURY, R. COOKE, J. ERLANDSON, J.A. ESTES, and Others (2001), *Historical overfishing and the recent collapse of coastal ecosystems*, Science 293, 629–638.
- KENNEDY, M., KOPP, S., *Understanding map projections*. GIS by ESRI (ESRI Inc. Redlands, CA, 2000).
- KENNY, A.J., CATO, I., DESPREZ, M., FADER, G., SCHÜTTEHELM, R.T.E., and SIDE, J., (2003), *An overview of seabed-mapping technologies in the context of marine habitat classification*. ICES Journal of Marine Science 60, 411–418.
- KHALID A.E., (2006), *A COM-based expert system for selecting the suitable map projection in ArcGIS*, Expert Systems with Applications 31, 94–100.
- LAMBIN, J., MORROW, R., FU, L.-L., WILLIS, J.K., BONEKAMP, H., LILLIBRIDGE, J., PERBOS, J., ZAUCHE, G., VAZE, P., BANNOURA, W., PARISOT, F., THOUVENOT, E., COUTIN-FAYE, S., LINDSTROM, E., and MIGNOGNO, M., (2010), *The OSTM/Jason-2 Mission*, Marine Geodesy 33, Supplement 1, 4–25.
- LEFEBVRE, J., ROUSSEL, H., WALTER, E., LECOINTE, D., and TABBARA, W., (1996), *Prediction from wrong models: the Kriging approach*, Antennas and Propagation Magazine IEEE 38(4), 35–45.
- MACARTHUR, R.H., and WILSON, E.O., *The Theory of Island Biogeography*. (Princeton University Press, Princeton, N.J., 1967).
- MARKS, K.M., and SMITH, H.F., (2006), *An evaluation of publicly available global bathymetry grids*, Marine Geophysical Researches 27, 19–34.
- MORATO, T., WATSON, R., PITCHER, T.J., and PAULY, D., (2006), *Fishing down the deep*. Fish and Fisheries 7, 24–34.
- NYERGES T.L., JANKOWSKI P., (1989). *A Knowledge Base for Map Projection Selection*. Cartography and Geographic Information Science 16, 29–38.
- PERBOS, J., (2003), *J-1: Satellite and system performance, one year after launch*, AVISO Newsletter 9, 4–5.
- PINHEIRO, J., BATES, D., DEBROY, S., and SARKAR, D. (2008), *Package 'nlme'. Linear and Nonlinear Mixed Effects Models*. <http://cran.r-project.org/web/packages/nlme/nlme.pdf>.
- PRIEDE, I.G., (1984), *A basking shark (Cetorhinus maximus) tracked by satellite together with simultaneous remote sensing*, Fisheries Research 2, 201–216.
- PRIEDE, I.G., and MILLER, P.I., (2009), *A basking shark (Cetorhinus maximus) tracked by satellite together with simultaneous remote sensing II: new analysis reveals orientation to a thermal front*, Fisheries Research 5, 370–372.
- PRIEDE, I.G., GODBOLD, J.A., KING, N.J., COLLINS, M.A., BAILEY, D.M., and GORDON, J.D.M., (2010), *Deep-sea demersal fish species richness in the Porcupine Seabight, NE Atlantic Ocean: global and regional patterns*, Marine Ecology 31, 247–260.
- PRIEDE, I.G., GODBOLD, J.A., NIEDZIELSKI, T., COLLINS, M.A., BAILEY, D.M., GORDON, J.D.M., and ZUUR A.F., (2011), *A review of the spatial extent of fishery effects and species vulnerability of the*



- deep-sea demersal fish assemblage of the Porcupine Seabight, Northeast Atlantic Ocean (ICES Sub-area VII)*, ICES Journal of Marine Science 68, 281–289.
- REX, M., and ETTER, R., Deep-sea biodiversity: pattern and scale. (Harvard University Press, Cambridge, MA 2010).
- ROBERTS, S., and PASHLER, H., (2000), *How persuasive is a good fit? A comment on theory testing*, Psychological Review 107, 358–367.
- SMITH, W., and SANDWELL, D., (1997), Measured and Estimated Seafloor Topography, World Data Service for Geophysics, Boulder Research Publication RP-1, poster, 34" X 53".
- TOBLER, W., (1970), *A computer movie simulating urban growth in the Detroit region*, Economic Geography 46, 234–240.
- The GEBCO\_08 Grid, version 20090202, <http://www.gebco.net>.
- USERY, L.E., SEONG J.C., (2001), *All Equal-Area Map Projections Are Created Equal, But Some Are More Equal Than Others*. Cartography and Geographic Information Science 28, 183–194.
- WOOD, S., (2007), The mgcv Package v. 1.3-28. GAMs with GCV smoothness estimation and GAMM by REM L/PQL, <http://cran.r-project.org/web/packages/mgcv/mgcv.pdf>.
- ZUUR, A.F., IENO, E.N., and SMITH, G.M., Analysing ecological data (Springer Science + Business Media, LLC, New York, 2007).
- ZUUR, A.F., IENO, E.N., WALKER, N.J., SAVELIEV, A.A., SMITH, G.M., Mixed effects models and extensions in ecology with R. (Springer Science + Business Media, LLC, New York, 2009).

(Received July 22, 2013, accepted July 29, 2013, Published online August 22, 2013)