

This is the pre-peer reviewed version of the following article, which has been published in final form at <http://onlinelibrary.wiley.com>

Understanding what controls the spawning distribution of North Sea whiting (*Merlangius merlangus*) using a multi-model approach

Christophe Loots¹, Sandrine Vaz^{1,*}, Benjamin Planque², Philippe Koubbi^{3,4}

¹ Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Laboratoire Ressources Halieutiques, 150 quai Gambetta, BP699, F-62321, Boulogne/mer, France

² Institute of Marine Research, Postboks 6404. 9294 Tromsø, Norway

³ UPMC Univ Paris 06, UMR 7093, Laboratoire d'Océanographie de Villefranche, 06230 Villefranche-sur-Mer, France

⁴ CNRS, UMR 7093, LOV, 06230 Villefranche-sur-Mer, France

*: Corresponding author : S. Vaz, email address : Sandrine.Vaz@ifremer.fr

Abstract:

The processes that control the spatial distribution of North Sea whiting (*Merlangius merlangus*) spawning adults are investigated using a statistical multi-model approach. Models of external and internal controls on the population, such as environmental conditions, spatial constraints, present or past spatial distribution, and demographic state of the population, are evaluated, compared and ranked to select those that are the best able to predict the observed distribution of spawning adults. Model selection is greatly influenced by the selection method, either based on data fitting or prediction, as well as by the threshold value used to stop the selection. Model selection based on prediction tends to select simpler models than selection based on data fitting. The hypotheses underlying the selected models are inferred to play a significant role in controlling North Sea whiting spatial distribution. The multi-model inference approach developed in this study enables comparison of several theoretical concepts and hypotheses and the results provide important clues on the processes involved in the control of the spatial distribution of whiting. We conclude that whiting has a high spatial fidelity to spawning site which can be linked to either geographical attachment or year-to-year persistence of the spatial distribution of the population. Environmental factors – temperature and salinity – appear to influence the geographical extent of spawning whiting distribution, whereas local abundance levels are primarily controlled by internal factors, i.e., population size and spatial segregation between ages.

Keywords: Akaike information criterion; hypotheses inference; model selection; North Sea; spawning distribution; whiting

INTRODUCTION

1
2
3
4
5
6 45 North Sea whiting (*M. merlangus*) is one of the key species of the North Sea
7 demersal fish assemblages (Greenstreet and Hall, 1996) as well as the third most
8 exploited gadoid fish in Scotland (Stratoudakis *et al.*, 1999). Not studied as much as cod
9 and haddock, there are nevertheless several studies on population structure (Hislop and
10 MacKenzie, 1976; Rico *et al.*, 1997; Charrier *et al.*, 2007), stock assessment (Pope and
11 Macer, 1996), diet (Hislop *et al.*, 1991; Pedersen, 1999) and spatial distribution (Zheng
12 *et al.*, 2001; Zheng *et al.*, 2002; Atkinson *et al.*, 2004). Whiting has one of the longest
13 spawning periods among North Sea species, lasting from February to June, with a peak
14 in April (Gibb *et al.*, 2004). Spawning progressively occurs later in the season at
15 increasing latitudes. Whiting becomes mature at 2 years old. Adults are predominantly
16 found in shallow waters, at temperatures between 6 and 9°C, and have a protracted
17 spawning period of between 6 weeks and 2 months. Eggs are pelagic and released in the
18 water in several batches (ICES, 2005). As for many fish species in the North Sea,
19 whiting stock abundance has substantially declined since the 1970s (Hislop, 1996;
20 Serchuk *et al.*, 1996) because of increasing fishing mortality since the 1940s
21 (Greenstreet and Hall, 1996), a high discard rate (Stratoudakis *et al.*, 1999) and poor
22 recruitment in the 1980s and 1990s (ICES, 2008). Gibb *et al.* (2004) indicate that this
23 decline may be associated with a contraction in the spatial distribution and a loss of
24 distinct reproductive populations. However, Daan *et al.* (1990) highlight that little
25 information on spawning distribution and the factors that determine spawning ground
26 selection is available (Coull *et al.*, 1998; Gibb *et al.*, 2004; ICES, 2005).
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53 65
54
55
56
57
58
59
60

Many processes related to both external and internal controls of fish population can explain spatio-temporal variations in spawning fish distribution (Planque *et al.*, 2010). External controls, such as environmental constraints, are independent from

1
2
3 population condition. Internal controls are directly linked to the status of the population
4
5
6 70 under study and include density-dependent processes, the effect of demographic
7
8 structure, trophic interactions and past conditions of the population. Environment
9
10 constraint is by far the most commonly considered process, as it is based on the well
11
12 developed ecological niche theory of Hutchinson (1957). A common spatial structure
13
14 may be shared by species and environment conditions, as the environment is highly
15
16
17 75 structured in space and many species have ecological preferences for specific
18
19 environmental ranges. However, individual interactions, especially during reproduction,
20
21 can also lead to spatial dependency in fish spatial distribution. It may be difficult to
22
23 separate pure environmental control from combined space-environment control, but
24
25 recent studies have developed ways to evaluate this shared component (Borcard *et al.*,
26
27
28 80 1992; Borcard and Legendre, 1994; Borcard and Legendre, 2002; Borcard *et al.*, 2004).
29
30 Other processes related to the internal condition of the population, mainly its density
31
32 and demographic structure, may also influence fish spatial distribution. According to the
33
34 ideal free distribution (Fretwell and Lucas, 1970; Sutherland, 1983) and the density-
35
36 dependent habitat selection (McCall, 1990; Rosenzweig, 1991), fish populations may
37
38
39 85 have a contracted spatial distribution at low level of population density whilst
40
41 expanding at high abundance, leading to an abundance-area relationship (Swain and
42
43 Morin, 1996; Fisher and Frank, 2004; Blanchard *et al.*, 2005). Old and young
44
45 individuals may also be found in distinct geographical areas in variable proportions,
46
47 thus the relative abundance of different age classes in a given year may partly control
48
49
50
51
52 90 the population's spatial distribution (Zheng *et al.*, 2001). Finally, recurrent spawning
53
54 areas are commonly found for many fish species, indicating a strong geographical
55
56 attachment. This ability to return to a specific location from year to year may result
57
58 from geographical attachment and/or from a number of mechanisms which promote
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

persistence, or conservatism, in spatial distribution patterns (Corten, 2002; Petitgas *et al.*, 2006). We group the latter processes under the term of ‘population memory’. A more complete description of the various ecological hypotheses on the control of fish population spatial distributions is provided in Planque *et al.* (2010).

To elucidate the relative importance of hypotheses that control spawning fish spatial distribution, a multi-inference model and information theoretic approach may be used (Hilborn and Mangel, 1997; Burnham and Anderson, 2002; Johnson and Omland, 2004; Stephens *et al.*, 2007; Diniz-Filho *et al.*, 2008). Using such an approach, Loots *et al.*, (2010) have successfully compared several models reflecting various hypotheses of control based on a metric distance between model predictions and observations. Hypotheses underlying the selected set of reasonably good models can thus be inferred to represent the dominant controls of the spatio-temporal variability of a population’s distribution pattern. Based on the approach developed in Planque *et al.* (2010) and applied in Loots *et al.* (2010), the present study aims to determine what are the dominant controlling factors of the spatial distribution of North Sea whiting at the beginning of its spawning season. We investigate several processes of control using statistical models that are evaluated, compared and selected. In addition, the sensitivity of the model selection and inference to datasets and selection thresholds are investigated.

MATERIAL AND METHODS

The biological response

1
2
3
4
5
6
7
8
9
10 115 The spatial distribution of whiting at the beginning of its spawning season was
11 investigated using the 1980 to 2007 abundance data of spawning adult whiting from the
12 first quarter of the International Bottom Trawl Surveys (January-March) (Fig. 1). These
13 abundances have been computed for the seven standard roundfish areas of ICES
14 (International Council for the Exploration of the Sea, Fig. 1), from the available
15 proportion of stage 3 adults (corresponding to spawning adults with fluent gonads,
16 ICES, 2004) within any given length class (Fig. 2). Males and females data were
17 merged together as the data were insufficient to calculate proportions of spawning
18 adults for each sex. The total number of trawling stations available was 9435. For each
19 of these stations, total abundance of spawning adults was calculated as the product of
20 the total abundance and the corresponding proportion of spawning adults within each
21 size class.
22
23
24
25
26
27
28
29
30
31
32
33 125
34
35
36
37
38

Modelling the control hypotheses

39
40
41
42 Ten possible factors of control have been modelled: persistent and non-persistent
43 environment, three scales of spatial dependency, population size, annual and spatial age
44 structure, geographical attachment and population memory. Persistent environment
45 consists of environmental conditions that are spatially structured but do not display
46 interannual variations during the study period (e.g. bottom depth). Non-persistent
47 environment are environmental conditions that are also spatially structured but have
48 changed during the study period (e.g. temperature). Spatial dependency (resulting in
49 spatial autocorrelation) describes the spatial structure present in the distribution of
50 whiting. Spatial dependency may potentially be shared by other controls like persistent
51
52
53
54
55
56
57
58 135
59
60

1
2
3 and non-persistent environment. Population size is the size of the population in biomass
4
5 and number of individuals by year. Annual age structure (referred to as annual
6
7 demography) corresponds to the proportions of each year-class in the whole area,
8
9
10 140 whereas spatial age structure (referred to as spatial demography) is the spatial
11
12 distribution of each year-class at each trawling station. Geographical attachment
13
14 corresponds to absolute geographical coordinates and population memory depicts the
15
16 degree of persistence of whiting distribution from one year to the next. Geographical
17
18 attachment supposes that whiting returns to the same spawning site every year whereas
19
20
21 145 population memory reflects that current spatial distribution depends upon recent past
22
23 distribution(s). Geographical attachment is equivalent to local random processes with a
24
25 fixed mean, with no memory in the process, but if the variance is low, the different
26
27 realizations will be close to the mean and therefore local abundance is similar between
28
29 years. Population memory is equivalent to an autoregressive process. There is memory
30
31
32 150 in the process and the mean is not fixed, therefore local abundance is expected be more
33
34 similar between consecutive years than between years that are separated in time.
35
36
37

38
39 Five environmental variables were used: depth (m), bottom temperature ($^{\circ}\text{C}$),
40
41 salinity, seabed stress ($\text{N}\cdot\text{m}^{-2}$) and sediment type. Depth, bottom temperature and
42
43 salinity were measured at each trawl location from 1980 to 2007. Seabed stress is a
44
45 155 measure of water shear friction on the seabed due to the tidal current and is provided by
46
47 the Proudman Oceanographic Laboratory (POL, Aldridge and Davies, 1993). Sediment
48
49 type is taken from the MARGIS project seafloor sediment of the North Sea (Schlüter
50
51 and Jerosch, 2008). Sediment classification was aggregated into five classes: fine sand,
52
53 coarse sand, mud, pebbles and gravel. A value of seabed stress and a class of sediment
54
55
56 160 were allocated to each trawl by resampling corresponding maps at trawl location using
57
58
59
60 ArcMap's Hawth's Analysis Tools extension (Beyer, 2004). Depth, seabed stress and

1
2
3 sediment type were grouped to represent persistent environment, whereas bottom
4
5 temperature and salinity were grouped to represent non-persistent environment.
6
7

8 Control through spatial dependency between individuals was modelled using
9
10 165 principal coordinates of neighbour matrices (PCNM). PCNM variables result from the
11 diagonalized distance matrix and represent a spectral decomposition of the spatial
12 relationships among the study sites. They produce orthogonal maps that maximise
13 spatial auto-correlation patterns that may arise at different spatial scales within the
14 sampling network. They produce orthogonal maps that maximise
15 spatial auto-correlation patterns that may arise at different spatial scales within the
16 sampling network. The PCNM variables resulting from a given sampling design are
17 independent from any response variables observed by the mean of this sampling
18 strategy. As such, they can be directly linked to the spatial patterns of the response
19 distribution to evaluate which spatial scales and auto-correlation patterns are most
20 relevant to the studied response. Extraction of the PCNM was achieved following the
21 method of Borcard and Legendre (2002). A Euclidean distance matrix was built using
22 170 latitude and corrected longitude ($longitude' \cos((latitude' p)/180)$) of the 9435
23 trawling stations. This distance matrix was then truncated by recoding all the distances
24 above a particular threshold up to four times this threshold. The threshold was taken as
25 the maximum distance between two neighboring stations in order to keep all the stations
26 connected together. A principal coordinate analysis was then performed on this
27 truncated distance matrix, leading to 9435 potential PCNM, from which only positive
28 PCNM (eigenvectors with positive eigenvalues) were kept. Then, PCNM that described
29 all the scales relevant to whiting distribution were selected, based on the method
30 described by Blanchet *et al.* (2008). In order to remove the significant spatial trend that
31 cannot be modelled using PCNM, log-transformed abundances were detrended prior to
32 selection using a quadratic trend based on latitudinal and corrected longitudinal
33 185 coordinates of the trawling stations, as prescribed by Borcard and Legendre (2002).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 PCNM were then selected from these detrended log-transformed abundances and a
4
5 global test including all the positive PCNM was performed to model the biological
6
7 response. As this test was significant, the analysis was carried on to the next step which
8
9
10 190 consisted in the forward selection. The significance of each PCNM was assessed using a
11
12 999 Monte Carlo permutations test and, according to a p -value and the adjusted- r^2
13
14 (explained variance) of the model calculated with all the PCNM. PCNM were tested
15
16 successively and the selection procedure stopped if the newly added PCNM was not
17
18 significant at the chosen p -value threshold or, if it made the adjusted- r^2 of the whole
19
20
21 195 model increase more than the value calculated for the model with all the PCNM. Spatial
22
23 scale of the selected PCNM was determined following the method outlined in Bellier *et*
24
25 *al.* (2007). A Gaussian model, used as a theoretical variogram, was fitted to each spatial
26
27 predictor to determine their range, which represents the spatial scale described by each
28
29 PCNM. PCNM describing similar spatial scale were grouped together by plotting their
30
31
32 200 eigenvalue against their range. Finally, three scales corresponding to three spatial sub-
33
34
35 models were determined and used as distinct hypotheses of control (see Loots *et al.*,
36
37 2010 for more details).
38
39
40

41 Age structure control was expressed as the relative proportions of each cohort (1
42
43 to 5 year olds) for the whole area (annual demography) and for each trawling station
44
45 (spatial demography). Estimations of annual abundances for the five cohorts are
46 205
47 available since 1980 (ICES, 2008). Proportions of each cohort at each trawling station
48
49 were calculated from the 1980 to 2007 biometric data of the International Bottom Trawl
50
51 Surveys.
52
53

54
55 Population size was expressed using the spawning stock biomass (SSB) and the
56
57 210 total size of the population provided since 1980 (ICES, 2008, Fig. 3). Although some
58
59 discrepancies between stock assessments using Extended Survivors Analysis (XSA) and
60

1
2
3 those using survey indices have been identified before 1990 (ICES, 2009), the XSA-
4
5 based SSB and total population size are considered to be the best estimates, as they are
6
7 based on all available sources of data during the time-period considered and over the
8
9
10
11 215 whole distribution area of this species.

12
13 The ICES statistical rectangles were used as a categorical variable to represent
14
15 the geographical location. Memory of the population was expressed using the observed
16
17 abundance of spawning adults in the previous year. For trawling positions with no
18
19 observation in the previous year, the observed abundance of the closest location was
20
21
22 220 used. This reduced the total available number of trawling stations from 9435 to 8098 for
23
24 the 1981 to 2007 period.
25

26 27 28 *Model structures*

29
30 Generalised additive models (GAM, Hastie and Tibshirani, 1990; Guisan *et al.*,
31
32 2002; Wood and Augustin, 2002) were used to construct the models (Loots *et al.*,
33
34
35 225 2010). The modelling procedure was split into two steps: a binomial model with a logit
36
37 link was built on presence-absence (where positive abundances are recoded as 1) and a
38
39 Gaussian model with an identity link was constructed for non-null log-transformed
40
41 abundances (Stefánsson, 1996; Le Pape *et al.*, 2003). This modelling design allows for
42
43 the appropriate handling of zero-inflated log-normal distribution, and also permits the
44
45
46
47 230 investigation and distinction between factors that control the geographical extent of the
48
49 distribution from those that control the geographical patterns of abundance (Koubbi *et*
50
51 *al.*, 2006; Loots *et al.*, 2010). For the binomial model, the following hypotheses of
52
53 control were tested: geography, persistent and non-persistent environment, the three
54
55 spatial sub-models, annual demography, population size and memory. For the Gaussian
56
57
58
59 235 model, the spatial demography was also added.
60

Model adjustment and prediction

In order to study the impact of the size of the prediction dataset on model complexity, years of the calibration and prediction datasets were chosen to retain a larger prediction dataset as advised by Loots *et al.* (2010). Also, the prediction dataset was constrained to be within the environmental and biological range covered by observations in the calibration data (Table 1). Nine years (1986-1989 and 2003-2007) were used for the calibration dataset, which corresponds to 2943 and 2753 stations used for fitting the binomial and Gaussian models. The remaining 18 years (1981-1985 and 1990-2002) were used for the prediction dataset, which represents 4150 stations used for the binomial model and 3894 for the Gaussian model. Degrees of smoothing were adjusted simultaneously on the corresponding fitting period for all the variables of the same hypothesis (Table 2). Finally, for each class of model (either binomial or Gaussian), all the combinations of the hypotheses were calculated resulting in 2^n-1 candidate models (with n , the number of hypotheses).

Model selection procedures

Models were selected for their ability to fit and predict the spatial distribution of whiting. Models were ranked based on their goodness of fit using Akaike Information Criterion (AIC, Akaike, 1974) and AIC difference (Δ_i , Burnham and Anderson, 2002; Johnson and Omland, 2004). The best model among all candidates is the one with the lowest AIC value (i.e. the model with the best compromise between adjustment to observed data and parsimony in the number of explanatory variables). Other models were compared to this best model by calculating the Δ_i (i.e. the difference of AIC between the best model and each model). Loots *et al.* (2010) indicated that the

1
2
3 260 conventional threshold value of 10 for Δ_i , advised by Burnham and Anderson (2002),
4
5
6 was too restrictive. Here, we use a range of threshold values from 0 to 60 as the criteria
7
8 for stopping the selection and the comparison of results was used to study the influence
9
10 of the selection threshold on model complexity.
11

12
13 Models were also compared for their predictive abilities, using the log-likelihood
14
15 265 between observations and predictions (McCullagh and Nelder, 1989; Burnham and
16
17 Anderson, 2002). The best model is the one with the highest log-likelihood value. Other
18
19 models were compared to it using a likelihood ratio test (LRT, Burnham and Anderson,
20
21 2002). Models were ranked according to their LRT from smallest to largest. The same
22
23 selection threshold of 60 (as for Δ_i) was used in order to keep only a reduced set of
24
25
26
27 270 selected models. Details on calculation of the AIC, Δ_i , log-likelihood and likelihood
28
29 ratio test can be found in the references above and in Loots *et al.* (2010).
30
31
32
33
34

35 *Model evaluation*

36
37 Binomial and Gaussian models selected based on the likelihood ratio test were
38
39 275 re-applied on the 3894 stations of the prediction dataset that were used to fit Gaussian
40
41 models. Predicted probabilities of presence and log-transformed abundances were then
42
43 combined in a delta approach. The delta approach consists of multiplying predicted
44
45 probabilities of presence by predicted abundances (conditional on presence) at each
46
47 location in order to derive the predicted spatial distribution (Stefánsson, 1996; Le Pape
48
49
50
51 280 *et al.*, 2003). This resulted in several predicted distributions that were compared to the
52
53 observed distribution using a Taylor diagram (Taylor, 2001).
54
55
56
57
58
59
60

RESULTS

PCNM extraction and spatial sub-models

Principal Coordinates of Neighbour Matrices were used to depict spatial dependency in the spatial distribution of North Sea whiting spawning adults. A threshold of 1.13 (decimal degrees of latitude) was used to truncate the distance matrix between stations. Overall, 3832 positive PCNM were extracted from the principal coordinate analysis and 22 were determined as significant by the forward selection (p -value ≤ 0.001 , adjusted $R^2 \geq 0.003$). Three sub-models representing three spatial scales were determined (Fig. 4): broad scale (166-355 km), medium scale (55-166 km) and fine scale (< 55 km). Three PCNM accounted for the broad scale, ten for the medium scale and nine for the fine scale.

Complexity of models selected

On the basis of the hypotheses tested, 512 binomial and 1024 Gaussian models were constructed, evaluated and ranked. Using a limit threshold of selection of 60 for Δ_i and LRT, a total number of 320 models were finally retained. Selected models varied according to the selection procedure and retained hypotheses varied with the type of model (binomial or Gaussian). Selected models based on data fitting performances were on average more complex, in terms of number of hypotheses and explanatory variables they contained, than models selected on predictive capabilities (Fig. 5). Selected Gaussian models were also more complex than selected binomial models (Fig. 5).

Complexity of the selected models varied with the selection threshold (Fig. 5). In binomial models selected on the basis of fitting performances ($\Delta_i < 60$), the number of hypotheses and explanatory variables contained decrease as the selection threshold increased. This was not the case for binomial models selected based on prediction

1
2
3 performances (LRT < 60). Selected models with lower LRT value (0-20) contained
4 fewer hypotheses and explanatory variables than models with higher LRT values (20-
5
6
7
8
9
10
11
12
13 310 was not the case for models selected on the basis of prediction. Models with low LRT
14 values (0-30) contained fewer hypotheses and explanatory variables than models with
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Frequency of the hypotheses

Frequencies of each control hypothesis were calculated from the 24 binomial
315 and 17 Gaussian models retained in the selection procedure based on predictive
performances (Table 3). Frequencies of the various hypotheses differ between binomial
and Gaussian models and vary with the value of the selection threshold (Table 4).

For binomial models, three hypotheses dominate regardless of the threshold
value: non-persistent environment, medium spatial scale and population memory.
320 Control by population size is seen in models with LRT values between 10 and 20. The
number of hypotheses shifts from four to six in models with a LRT value between 20
and 40, with persistent environment and broad scale sub-model as supplementary
hypotheses. Control by fine spatial scale is not frequent and only present in models with
a LRT value between 50 and 60. Geography and annual demography are not present in
325 selected models.

For Gaussian models, six hypotheses dominate in models regardless of the
selection threshold value: geographical attachment, permanent and non-persistent
environment, population size, spatial demography and population memory. Persistent
and non-persistent environment are less frequent in models with LRT values between
330 20 and 60. The three scale sub-models are not present in models with LRT values

1
2
3 between 0-10, whereas they become increasingly frequent in models with higher LRT
4
5 values. Annual demography is never retained in the selected models.
6
7

8 9 *Evaluation of delta models*

10
11 The 17 binomial and 24 Gaussian selected models (Table 3) were combined in a
12
13 335 delta approach, resulting in 408 predictions. These predictions were evaluated using the
14
15 Taylor diagram (Fig. 6). The Taylor diagram indicates that all the models have
16
17 comparable values of Spearman correlation (close to 0.8), Root Mean Square Error
18
19 (close to 0.6) and standard deviation (between 0.7 and 0.8).
20
21
22

23 This is confirmed by the maps of predicted distribution of spawning whiting,
24
25 340 illustrated for the years 1992 and 1998 (Fig. 7). These two years show similar pattern of
26
27 distribution with higher abundances in the northern and southern parts of the North Sea.
28
29 Abundances are lower for year 1998, especially in the southern part. For each year,
30
31 predictions of the simplest and the most complex models among those selected (see
32
33 Table 3 for formulae) have been mapped. Whereas these models do not have the same
34
35 LRT value or complexity, they are both able to depict the observed distribution for both
36
37 345 years.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

Based on a set of 28 years of biological and environmental data collected in the North Sea, the statistical modelling approach used in this study revealed the key mechanisms which are involved in the control of the geographical extent of spawning whiting and the spatial distribution of its abundance during the spawning season. We discuss below new and important insights on the technical aspects of the analysis as well as on the ecological implications of the results obtained.

Selection based on prediction versus selection based on fitting

In this study, the prediction dataset was much larger than the calibration dataset and therefore potentially supports more variability and a higher number of autocorrelated data. Thus, we could have supposed that more complex models might have been selected from prediction than from fitting. However, this was clearly not the case as models selected on predictive performances were less complex on average than models selected from fitting. These results support those obtained by Loots *et al.* (2010). In their first attempt to perform multi-model selection, they built a calibration dataset larger than the prediction dataset and shown that selected models from prediction were less complex than those selected from fitting. The present study confirms that this was not due to the size of the datasets because in the present study the prediction dataset was larger and the models selected from prediction were still less complex compared to those selected from fitting. This is an important insight as it supports the principle that once processes that control the spatial distribution have been correctly identified through hypotheses selection, they will be able to depict the observed distribution whatever the complexity of the spatial pattern. The present study has shown that this feature could be better achieved using prediction-based selection

1
2
3 rather than fitting-based selection (see also discussion on this aspect in Planque *et al.*,
4
5
6 2010).

7
8 Consequently, we argue that the difference in complexity between models
9
10 selected from fitting compared to those selected from prediction is not due to the size of
11
12 the datasets, but to the principles underlying the calculation of the AIC, as argued in
13 375 Planque *et al.* (2010). The AIC is designed to favour models with high data fitting
14
15 performance whilst penalizing models of high complexity. In the presence of
16
17 autocorrelated data, the amount of data is much greater than the number of true
18
19 independent observations, and the AIC is heavily skewed towards favouring fitting
20
21 performance rather than model parsimony. In such situation the use of the AIC can lead
22
23
24 380 to the selection of overfitted and more complex models, as is observed from the current
25
26 data.
27
28
29
30

31
32 This problem is absent when testing performances using the LRT on predictions
33
34 because the selection criteria is independent from the true number of independent
35
36 385 observations in the fitting dataset. Based on the results of Loots *et al.* (2010) and the
37
38 present study, it appears that spatial autocorrelation acts as a statistical nuisance when
39
40 model selection is based on data fitting (using AIC) and it should be recommended that
41
42 model selection be based on prediction as a more robust method to infer the role of the
43
44 various hypotheses.
45
46
47
48

49 390 *Impact of a the selection threshold*

50
51 We chose a wide range of LRT threshold values to evaluate how these can affect
52
53 model selection and hypotheses inference. Unexpectedly, evaluation of delta models has
54
55 revealed that all the predictions have similar correlation with the observations and
56
57 depict the same associated error (Taylor diagram in Fig. 6). This highlights that all
58
59 selected models can predict the main patterns of observed spatial distributions with
60 395

1
2
3 relatively similar skills, even though they may have different quantitative performances,
4
5 as revealed by LRT values.
6
7

8 Although the selected models display similar predictive skills, they can have
9
10 different levels of complexity and include different sets of hypotheses. In the selection
11
12 process based on prediction, simple models were selected first and hypotheses that were
13 400 determined as the most important (i.e. the most frequent) in these first selected models
14
15 remained the most important in following models. Therefore, considering a large
16
17 number of models does not lead to undermining of the most important hypotheses, but it
18
19 may lead to attributing too much importance to secondary hypotheses.
20
21
22
23

24 405 One may wonder which value of selection threshold should be chosen. For the
25
26 binomial models of this study, the simplest model was in the set of models that have a
27
28 LRT lower than 10. In this case, restricting the selection threshold to 10 allows for the
29
30 selection of a model with a small number of underlying hypotheses which is also the
31
32 best model. For Gaussian models, the simplest model is not within models with an LRT
33
34 value of 10 or less and restricting the selection threshold to 10 led to the exclusion of a
35
36 410 simple model which has a predictive power similar to more complex ones. This reveals
37
38 that choosing a too restrictive selection threshold, even with prediction based selection,
39
40 can restrict the choice to models that are too complex, whereas simple models, that are
41
42 not automatically within the threshold of 10 (as advised by Burnham and Anderson,
43
44 2002), may have similar predictive performances. Therefore, the present study
45
46 415 demonstrates that model ranking should not be considered as the only criterion for
47
48 model selection when using biological data. Rather, it can be used as a first step to
49
50 evaluate and compare a selection of models. Once the various models have been ranked,
51
52 models of equivalent predictive power should be examined.
53
54
55
56
57
58
59
60

420 *Control of the spatial distribution of whiting*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In our study, a selection threshold of 10 for binomial models and 20 for Gaussian models appears to be appropriate to evaluate the relevance of the different hypotheses. Selected binomial models were clearly different from selected Gaussian models. The results reveal that the geographical extent of the spawning population and geographical distribution of abundance are not controlled by the same processes. The geographical extent of the population (i.e. distribution of areas of presence-absence) is controlled by non-persistent environment and population memory. Within this, the distribution of high abundances of spawning whiting is influenced by geographical attachment, population size, age structure and population memory. Ecologically, this reveals that the level of abundance of whiting in an area is determined by a more complex association of factors than is the geographical spread of the population. This was also observed for plaice (*Pleuronectes platessa*) in the North Sea by Loots *et al.* (2010).

The importance of the memory hypothesis shows that the areas of presence-absence tend to be maintained in time as the presence-absence of whiting for one year is highly related to that of the previous year. Whereas the importance of the memory hypothesis reveals a high conservatism in spawning migrations, Zheng *et al.* (2001, 2002) argued that the spatial pattern of sea surface temperature has an important influence on the spatial distribution of whiting, in particular during winter and spring. Although our study can not rule out the effects of temperature of whiting distribution at the time of spawning, it appears that such effects are likely minor because whiting has a wide area of presence and few areas of absence (only 6% of zeros) in the North Sea.

The relevance of the memory hypothesis as well as the geographical attachment highlights a strong persistence of whiting distribution and abundance from year to year.

1
2
3 445 Such persistence may reflect population strategy which can ensure life cycle closure,
4
5
6 despite unpredictable variations in the environmental conditions from year-to-year.
7
8 Such mechanism creates a time lag between environmental changes and changes in the
9
10 spatial distribution of a species which allows endurance of exceptional environmental
11
12 conditions (Corten, 2002). In the short term, this can result in poor recruitment in a
13
14
15 450 specific year because eggs and larvae may not be released in optimal conditions for their
16
17 survival. On the other hand, such strategy may ensure good average recruitment over
18
19 decadal time scales.
20
21

22 Our results indicate that temporal variations in abundance may be influenced by
23
24 the size and age structure of the population. The overall declining trend in population
25
26
27 455 size seemed to better match the apparent contraction of whiting distribution in recent
28
29 years, which probably contributed to the importance of this control factor in explaining
30
31 whiting spatial distribution in the present study. Control through population size
32
33 suggests a density dependent effect where the spatial extent of areas of high
34
35 concentrations of spawning adults may vary in relation to population abundance. In
36
37
38 460 addition, our results indicate segregation between young individuals, that are mainly
39
40 distributed in shallow areas, and old individuals distributed in deeper areas (Zheng *et*
41
42 *al.*, 2001). This age-dependent segregation pattern is important to explain and predict
43
44 the spatial distribution of abundance of spawning whiting, in particular in the face of
45
46 changes in the population age-structure.
47
48
49

50
51 465 In summary, although environmental factors such as temperature and salinity
52
53 may play a role in the location and extent of whiting distribution, year-to-year
54
55 persistence in the spatial distribution allows North Sea whiting to buffer for short term
56
57 environmental variations. Therefore, as long as changes in present environmental
58
59 conditions remain within the fundamental niche of whiting, it should be expected that
60

1
2
3 470 the geographical configuration of spawning areas will vary little. This may not be the
4
5 case if future ocean climate changes beyond the niche range of whiting. Numerical
6
7 simulations of future climate (coupled with the selected models) will be necessary to
8
9 assess the impact of such changes on whiting spawning distribution. However, this may
10
11 remain a difficult challenge, if the environmental and biological ranges encompassed by
12
13 475 scenarios are not within those of the calibration dataset, the model will extrapolate so
14
15 that the predictions will be poorly reliable.
16
17
18
19
20

21 Inside the areas of presence, the areas of high abundances are also persistent in
22
23 time. Variations in their extent are no longer determined by the environment, but rather
24
25 by internal population controls, mainly population size and demography. A high level of
26
27 480 abundance as well as a higher diversity in the age structure of the population result in
28
29 wider spatial coverage of the spawning population. When such a situation occurs, eggs
30
31 and larvae are released over a wide range of environmental conditions, which can then
32
33 limit the risk of recruitment failure. Within this context, the loss of a part of the
34
35 population abundance as well as the truncation of the age pyramid due to fishing
36
37 485 mortality can strongly reduce the spatial coverage of spawning, thereby increasing the
38
39 risk of recruitment failure caused by poor environmental conditions.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CONCLUSION

1
2
3
4
5
6
7
8
9
10
11
12 490
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47 505
48
49
50
51
52
53
54
55
56
57
58
59
60

The present study concludes that model selection based on data fitting performances - as still commonly practised in fisheries and marine ecology studies - leads to selection of overly complex and overfitted models. Instead, prediction-based selection provides a valuable way for model selection and hypothesis inference and should be recommended over fitting or cross validation based selection.

The multi-model approach developed in this study reveals that a large amount of model uncertainty exists, due to the difficulty of selecting an appropriate final set of models that may differ in their complexity and hypotheses involved, but that share similar predictive performance. We contend that a range of models, supported by different hypotheses, can be selected as good candidates to explain and predict the spatial distribution of spawning whiting. These results support the idea that accounting for only one or two hypotheses and considering only one (best) model, as it is commonly performed, may lead to erroneous conclusions in understanding the distribution of fish population and predicting it under different climatic or biological change scenarios.

Despite such uncertainty, the selected models revealed that whiting distribution is primarily controlled by population memory and geographical attachment, leading to the conservation of spatial distribution over years, whilst variations are mainly due to population size and demography. Recent fluctuations in environmental conditions seem to play a minor role.

ACKNOWLEDGMENTS

1
2
3
4
5
6
7 This work was part-funded by the EU as part of the RECLAIM project (STREP-
8
9 510 FP6, contract n044133). The authors also would like to acknowledge financial supports
10
11 by the Region Nord-Pas de Calais. They are also grateful to all the persons involved
12
13 each year in the International Bottom Trawl Surveys for their precious work which
14
15 allow making this study feasible.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

REFERENCES

- 1
2
3
4
5 515 Akaike, H. (1974) A new look at the statistical model identification. *IEEE T. Automat.*
6
7
8 *Contr.* **19**: 716-723.
- 9
10 Aldridge, J.N. and Davies, A.M. (1993) A high resolution three-dimensional
11
12 hydrodynamic tidal model of the Eastern Irish Sea. *J. Phys. Oceanogr.* **23**: 207-224.
- 13
14 Atkinson, C.J.L., Bergmann, M. and Kaiser, M.J. (2004) Habitat selection in whiting. *J.*
15
16
17 520 *Fish Biol.* **64**: 788-793.
- 18
19
20 Bellier, E., Monestiez, P., Durbec, J. and Candau, J. (2007) Identifying spatial
21
22 relationships at multiple scales: principal coordinates of neighbour matrices (PCNM)
23
24 and geostatistical approaches. *Ecography* **30**: 385-399.
- 25
26
27 Beyer, H.L. (2004) Hawth's Analysis Tools for ArcGIS. Available at
28
29 525 <http://www.spatial ecology.com/htools>.
- 30
31
32 Blanchard, J.L., Mills, C., Jennings, S., Fox, C.J., Rackham, B.D., Eastwood, P.D. and
33
34 O'Brien, C.M. (2005) Distribution-abundance relationships for North Atlantic cod
35
36 (Gadus morhua): observation versus theory. *Can. J. Fish Aquat. Sci.* **62**: 2001-2009.
- 37
38
39 Blanchet, F.G., Legendre, P. and Borcard, D. (2008) Forward selection of explanatory
40
41 530 variables. *Ecology* **89**: 2623-2632.
- 42
43
44 Borcard, D., Legendre, P. and Drapeau, P. (1992) Partialling out the spatial component
45
46 of ecological variation. *Ecology* **73**: 1045-1055.
- 47
48
49 Borcard, D. and Legendre, P. (1994) Environmental control and spatial structure in
50
51 ecological communities: an example using oribatid mites (Acari, Oribatei). *Environ.*
52
53 535 *Ecol. Stat.* **1**: 37-61.
- 54
55
56 Borcard, D. and Legendre, P. (2002) All-scale spatial analysis of ecological data by
57
58 means of principal coordinates of neighbour matrices. *Ecol. Model.* **153**: 51-68.
- 59
60

- 1
2
3 Borcard, D., Legendre, P., Avois-Jacquet, C. and Tuomisto, H. (2004) Dissecting the
4 spatial structure of ecological data at multiple scales. *Ecology* **85**: 1826-1832.
5
6
7
8 540 Burnham, K.P. and Anderson, D.R. (2002) *Model selection and multimodel inference. A*
9 *practical information-theoretic approach*. USA: Springer-Verlag, 488pp.
10
11
12 Charrier, G., Coombs, S.H., McQuinn I H and Laroche, J. (2007) Genetic structure of
13 whiting *Merlangius merlangus* in the northeast Atlantic and adjacent waters. *Mar.*
14 *Ecol.-Prog. Ser.* **330**: 201-211.
15
16
17
18
19 545 Corten, A. (2002) The role of "conservatism" in herring migrations. *Rev. fish biol.*
20 *fisher.* **11**: 339-361.
21
22
23
24 Coull, K.A., Johnstone, R. and Rogers, S.I. (1998) Fisheries sensitivity maps in British
25 waters., 63 pp.
26
27
28
29
30
31 550 Daan, N., Bromley, P.J., Hislop, J.R.G. and Nielsen, N.A. (1990) Ecology of North Sea
32 *Fish. Neth. J. Sea Res.* **26**: 343-386.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55 560 Diniz-Filho, J.A.F., Rangel, T.F.L.V.B. and Bini, L.M. (2008) Model selection and
56 information theory in geographical ecology. *Global Ecol. Biogeogr.* **17**: 479-488.
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
- Borcard, D., Legendre, P., Avois-Jacquet, C. and Tuomisto, H. (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**: 1826-1832.
- 540 Burnham, K.P. and Anderson, D.R. (2002) *Model selection and multimodel inference. A practical information-theoretic approach*. USA: Springer-Verlag, 488pp.
- Charrier, G., Coombs, S.H., McQuinn I H and Laroche, J. (2007) Genetic structure of whiting *Merlangius merlangus* in the northeast Atlantic and adjacent waters. *Mar. Ecol.-Prog. Ser.* **330**: 201-211.
- 545 Corten, A. (2002) The role of "conservatism" in herring migrations. *Rev. fish biol. fisher.* **11**: 339-361.
- Coull, K.A., Johnstone, R. and Rogers, S.I. (1998) Fisheries sensitivity maps in British waters., 63 pp.
- Daan, N., Bromley, P.J., Hislop, J.R.G. and Nielsen, N.A. (1990) Ecology of North Sea *Fish. Neth. J. Sea Res.* **26**: 343-386.
- Diniz-Filho, J.A.F., Rangel, T.F.L.V.B. and Bini, L.M. (2008) Model selection and information theory in geographical ecology. *Global Ecol. Biogeogr.* **17**: 479-488.
- Fisher, J.A.D. and Frank, K.T. (2004) Abundance–distribution relationships and conservation of exploited marine fishes. *Mar. Ecol.-Prog. Ser.* **279**: 201-213.
- 555 Fretwell, S.D. and Lucas, H.L. (1970) On territorial behavior and other factors influencing habitat distribution in birds. *Acta Biotheor.* **19**: 16-36.
- Gibb, F.M., Wright, P.J., Gibb, I.M. and O'Sullivan, M. (2004) Haddock and whiting spawning areas in the North Sea and Scottish west coast, 26 pp.
- Greenstreet, S.P.R. and Hall, S.J. (1996) Fishing and the ground-fish assemblage structure in the north-western North Sea: an analysis of long-term and spatial trends. *J. Anim. Ecol.* **65**: 577-598.

- 1
2
3 Guisan, A., Edwards, T.C. and Hastie, T. (2002) Generalized linear and generalized
4 additive models in studies of species distributions: setting the scene. *Ecol. Model.* **157:**
5
6 89-100.
7
8
9
10 565 Hastie, T.J. and Tibshirani, R.J. (1990) *Generalized additive models*. London:
11 Chapman and Hall, 335pp.
12
13
14
15 Hilborn, R. and Mangel, M. (1997) *The ecological detective. Confronting models with*
16
17 *data. Monographs in population biology, 28*. Princeton: Princeton University Press,
18
19 315pp.
20
21
22 570 Hislop, J.R.G. and MacKenzie, K. (1976) Population studies of the whiting *Merlangius*
23
24 *merlangus* (L.) of the northern North Sea. *ICES J. Mar. Sci.* **37:** 98-110.
25
26
27 Hislop, J.R.G., Robb, A.P., Bell, M.A. and Armstrong, D.W. (1991) The diet and food
28
29 consumption of whiting (*Merlangius merlangus*) in the North Sea. *ICES J. Mar. Sci.* **48:**
30
31 139-156.
32
33
34 575 Hislop, J.R.G. (1996) Changes in North Sea gadoid stocks. *ICES J. Mar. Sci.* **53:** 1146-
35
36 1156.
37
38
39 Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harb. Symp. Quant. Biol.*
40
41 **22:** 415-427.
42
43
44 ICES (2004) Manual for the international bottom trawl surveys, rev. VII. International
45
46 580 Bottom Trawl Survey Working Group, ICES, Copenhagen. 51 pp.
47
48 <http://www.ices.dk/datacentre/datras/NSIBTSmanualRevVIIdraft.pdf>.
49
50
51 ICES (2005) Report of the Planning Group on North Sea Cod and Plaice Egg Surveys in
52
53 the North Sea (PGEGGS), 10–12 May 2005, Lowestoft, UK. ICES Document CM
54
55 2005/G:11, 85 pp.
56
57
58
59
60

- 1
2
3 585 ICES (2008) Report of the working group on the assessment of demersal stocks in the
4 North Sea and Skagerrak - combined spring and autumn (WGNSSK), 7-13 May 2008,
5 Copenhagen, Denmark. ICES CM 2008/ACOM:09, 921 pp.
6
7
8
9
10 ICES (2009) Report of the Benchmark and Data Compilation Workshop for Roundfish
11 (WKROUND), January 16–23 2009, Copenhagen, Denmark. ICES CM
12
13
14
15 590 2009/ACOM:32, 259 pp.
16
17 Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution.
18 *Trends Ecol. Evol.* **19**: 101-108.
19
20
21
22 Koubbi, P., Loots, C., Cotonnec, G., Harlay, W., Grioche, A., Vaz, S., Martin, C.,
23
24
25
26
27 595 *Solea solea*, *Pleuronectes flesus* and *Limanda limanda* fish larvae in the eastern English
28 Channel during the spring. **70S2**: 147-157.
29
30
31
32 Le Pape, O., Chauvet, F., Mahévas, S., Lazure, P., Guérault, D. and Désaunay, Y.
33
34 (2003) Quantitative description of habitat suitability for the juvenile common sole
35 (*Solea solea*, L.) in the Bay of Biscay (France) and the contribution of different habitats
36
37
38
39 600 to the adult population. *J. Sea. Res.* **50**: 139-149.
40
41
42
43
44
45
46
47
48
49
50
51 605 University of Washington Press, 153pp.
52
53
54
55
56
57
58
59
60
McCullagh, P. and Nelder, J.A. (1989) *Generalized linear models, Monographs on
statistics and applied probability 37*. London: Chapman and Hall, 511pp.
Pedersen, J. (1999) Diet comparison between pelagic and demersal whiting in the North
Sea. *J. Fish Biol.* **55**: 1096-1113.

- 1
2
3 610 Petitgas, P., Reid, D., Planque, B., Nogueira, E., O'Hea, B. and Cotano, U. (2006) The
4 entrainment hypothesis: an explanation for the persistence and innovation in spawning
5 migration and life cycle patterns. *ICES CM 2006/B:07*, 9 pp.
6
7
8
9
10 Planque, B., Loots, C., Petitgas, P., Lindstrom, U. and Vaz, S. (2010) Understanding
11 what controls the spatial distribution of fish populations using a multi-model approach.
12
13
14
15 615 *Fish. Oceanogr.* In press.
16
17
18
19
20
21
22
23
24
25
26
27 620 Rico, C., Ibrahim, K.M., Rico, I. and Hewitt, G.M. (1997) Stock composition in North
28 Atlantic populations of whiting using microsatellite markers. *J. Fish Biol.* **51**: 462-475.
29
30
31
32
33
34
35
36
37
38
39 625
40 information regarding geology, geochemistry, oceanography and biology. Available at
41 [http://www.awi.de/en/research/research_divisions/geosciences/marine_geochemistry/m](http://www.awi.de/en/research/research_divisions/geosciences/marine_geochemistry/marine_gis/digital_atlas_of_the_north_sea/)
42 [arine_gis/digital_atlas_of_the_north_sea/](http://www.awi.de/en/research/research_divisions/geosciences/marine_geochemistry/marine_gis/digital_atlas_of_the_north_sea/).
43
44
45
46
47
48
49
50
51 630
52
53
54
55
56
57
58
59
60
- Pope, J.G. and Macer, C.T. (1996) An evaluation of the stock structure of North Sea cod, haddock, and whiting since 1920, together with a consideration of the impacts of fisheries and predation effects on their biomass and recruitment. *ICES J. Mar. Sci.* **53**: 1157-1169.
- Rosenzweig, M.L. (1991) Habitat Selection and Population Interactions: The Search for Mechanism. *Am. nat.* **137**: S5-S28.
- Schlüter, M. and Jerosch, K. (2008) Digital atlas of the North Sea (DANS). Geo-information regarding geology, geochemistry, oceanography and biology. Available at http://www.awi.de/en/research/research_divisions/geosciences/marine_geochemistry/marine_gis/digital_atlas_of_the_north_sea/.
- Serchuk, F.M., Kirkegaard, E. and Daan, N. (1996) Status and trends of the major roundfish, flatfish, and pelagic fish stocks in the North Sea: thirty-year overview. *ICES J. Mar. Sci.* **53**: 1130-1145.
- Stefánsson, G. (1996) Analysis of groundfish survey abundance data: combining the GLM and delta approaches. *ICES J. Mar. Sci.* **53**: 577-588.
- Stephens, P.A., Buskirk, S.W. and del Rio, C.M. (2007) Inference in ecology and evolution. *Trends Ecol. Evol.* **22**: 192-197.

- 1
2
3 635 Stratoudakis, Y., Fryer, R.J., Cook, R.M. and Pierce, G.J. (1999) Fish discarded from
4
5
6
7
8
9
10
11
12
13
14
15 640 Swain, D.P. and Morin, R. (1996) Relationships between geographic distribution and
16
17
18
19
20
21
22
23
24
25
26
27 645 Wood, S.N. and Augustin, N.H. (2002) GAMs with integrated model selection using
28
29
30
31
32
33
34
35
36
37
38
39 650 Zheng, X., Pierce, G.J., Reif, D.G. and Jolliffe, I.T. (2002) Does the North Atlantic
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Scottish demersal vessels: estimators of total discards and annual estimates for targeted gadoids. *ICES J. Mar. Sci.* **56**: 592-605.
- Sutherland, W.J. (1983) Aggregation and the ideal free distribution. *J. Anim. Ecol.* **52**: 821-828.
- abundance of American plaice (*Hippoglossoides platessoides*) in the southern Gulf of St. Lawrence. *Can. J. Fish Aquat. Sci.* **53**: 106-119.
- Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.* **106**: 7183-7192.
- penalized regression splines and applications to environmental modelling. *Ecol. Model.* **157**: 157-177.
- Zheng, X., Pierce, G.J. and Reid, D.G. (2001) Spatial patterns of whiting abundance in Scottish waters and relationships with environmental variables. *Fish Res.* **50**: 259-270.
- current affect spatial distribution of whiting? Testing environmental hypotheses using statistical and GIS techniques. *ICES J. Mar. Sci.* **59**: 239-253.

FIGURES

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

655 Figure 1: The North Sea International Bottom Trawl Surveys. Black dots indicate trawl stations of the first quarter (January-March) from 1980 to 2007, for which the abundance of spawning adult North Sea whiting is given. Countries involved in the surveys, ICES statistical rectangles, standard roundfish areas and main locations are also indicated.

660 Figure 2: Proportions of adult stage 3 (spawning adults) North Sea whiting per size class, for the seven standard roundfish areas. The datapoints represent the proportions observed, and the lines the adjusted proportions using a smoothing spline for missing size classes.

Figure 3: ICES estimation (ICES Working Group) of North Sea whiting population size (thousands of fish) and spawning stock biomass ('000 t) from 1980 to 2007.

665 Figure 4: PCNM spatial sub-models. Left, plot of the range values of the 22 selected PCNMs against their eigenvalue. Broad scale, eigenvectors with a range between 166 km and 355 (1.5-3.2 decimal degrees of latitude); medium scale, eigenvectors with a range between 55 km and 166 km (0.5-1.5 decimal degrees of latitude); fine scale, eigenvectors with a range between 0 and 55 km (<0.5 decimal degrees of latitude).
670 Right, maps of the spatial sub-models (arbitrary unit) giving the scale of aggregation of spawning North Sea whiting. Top, broad scale sub-models; center, medium scale sub-models; right, fine scale sub-models. Black lines denote zero values for PCNM.

Figure 5: Complexity of selected binomial and Gaussian models. Complexity is represented by the mean number of hypotheses and explanatory variables within selected models. They are given according to the selection procedure and for different value of selection threshold for Δ_i and LRT between 0 and 60.

1
2
3 Figure 6: Plot of the Taylor diagram. Only the part representing a positive correlation is
4 shown. Predictions of delta models (triangle) are compared to the observations using the
5 standard deviation, the root mean square error (RMSE) and the Spearman correlation.
6
7

8
9
10 680 Observations are symbolized as a point referred to as the reference point and are
11 normalized so that the standard deviation is equal to 1, the correlation equal to 1 (the
12 correlation between the observations and themselves) and the RMSE equal to 0 (the
13 difference between the observations and themselves). Values of the indices for the
14 predictions are plotted with the standard deviation on the y-axis, the correlation on the
15 radii of the circle and the RMSE on concentric circles around the reference point.
16
17
18
19
20
21
22 685
23

24 Figure 7: Maps of spawning adult whiting. Maps for the most simple and complex
25 models among those selected (see Table 3 for formulae of the models) were computed
26 for years 1992 and 1998. The corresponding maps of the observed distribution of
27 spawning adults for the same two years are also shown.
28
29
30
31
32

33
34 690
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

TABLES

Table 1. Environmental and biological ranges encompassed by the calibration and the prediction datasets for building the binomial and the Gaussian models.

Table 2. The formulae for the hypotheses. The explanatory variables for each hypothesis for the binomial and the Gaussian models were fitted on the calibration dataset. An *s* means that a smoothing function has been used and the corresponding degree of smoothing is indicated in parentheses. PCNM are denoted by the term A_x . Hypotheses: Geo. Att., geographical attachment; Pers. Eenvt., persistent environment; N. Pers. Eenvt., non-persistent environment; Br. Sc., broad scale sub-model, Med. Sc., medium scale sub-model; Fine Sc., fine scale sub-model; Pop. Size, population size; Ann. Dem., annual demography; Spa. Dem., spatial demography; Pop. Mem., population memory.

Table 3. The formulae for the models selected from the prediction. The best model is the one with a value of LRT of 0. Values of Δ_i are also given. Formulae of the simplest (1) and the most complex (2) delta models used for figure 7 are indicated. Hypotheses: Geo. Att., geographical attachment; Pers. Eenvt., persistent environment; N. Pers. Eenvt., non-persistent environment; Br. Sc., broad scale sub-model; Med. Sc., medium scale sub-model; Fine Sc., fine scale sub-model; Pop. Size, population size; Spa. Dem., spatial demography; Pop. Mem., population memory.

Table 4. Frequencies of the different hypotheses, giving the percentage of the model selected from prediction in which each hypothesis is present, for different threshold of selection. Recoding was according to five levels: 0 (absence), + (<25%), ++ (25-50%), +++ (50-75%), +++++ (75-100%). Hypotheses: Geo. Att., geographical attachment; Pers. Eenvt., persistent environment; N. Pers. Eenvt., non-persistent environment; Br. Sc., broad scale sub-model; Med. Sc., medium scale sub-model; Fine Sc., fine scale sub-model;

1
2
3 Pop. Size, population size; Ann. Dem., annual demography; Spa. Dem., spatial
4
5
6 demography; Pop. Mem., population memory.
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

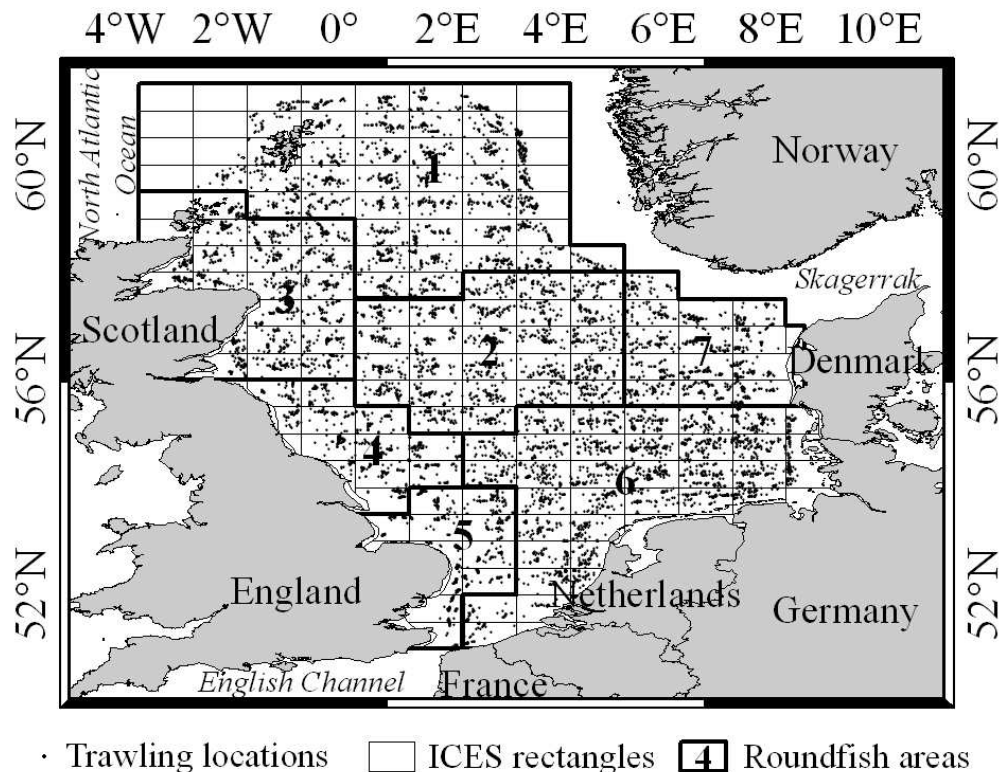


Figure 1: The North Sea International Bottom Trawl Surveys. Black dots indicate trawl stations of the first quarter (January-March) from 1980 to 2007, for which the abundance of spawning adult North Sea whiting is given. Countries involved in the surveys, ICES statistical rectangles, standard roundfish areas and main locations are also indicated.
79x62mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

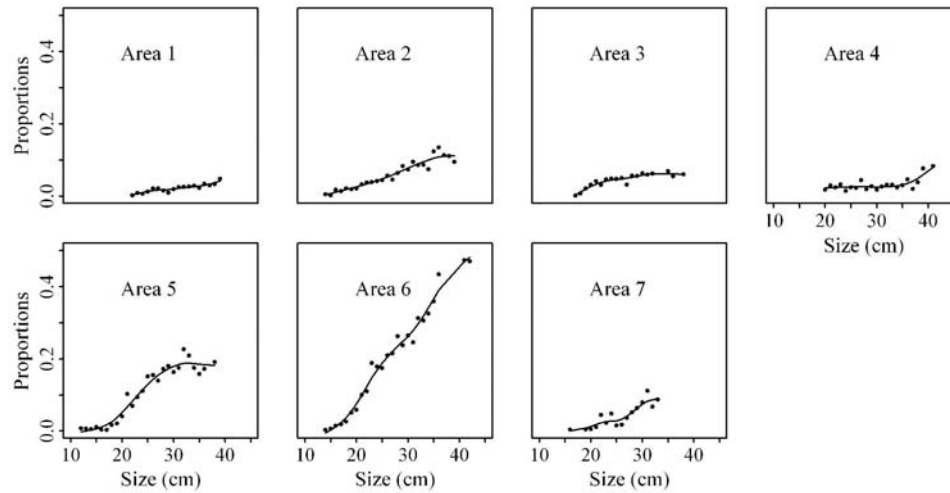


Figure 2: Proportions of adult stage 3 (spawning adults) North Sea whiting per size class, for the seven standard roundfish areas. The datapoints represent the proportions observed, and the lines the adjusted proportions using a smoothing spline for missing size classes.
160x80mm (600 x 600 DPI)

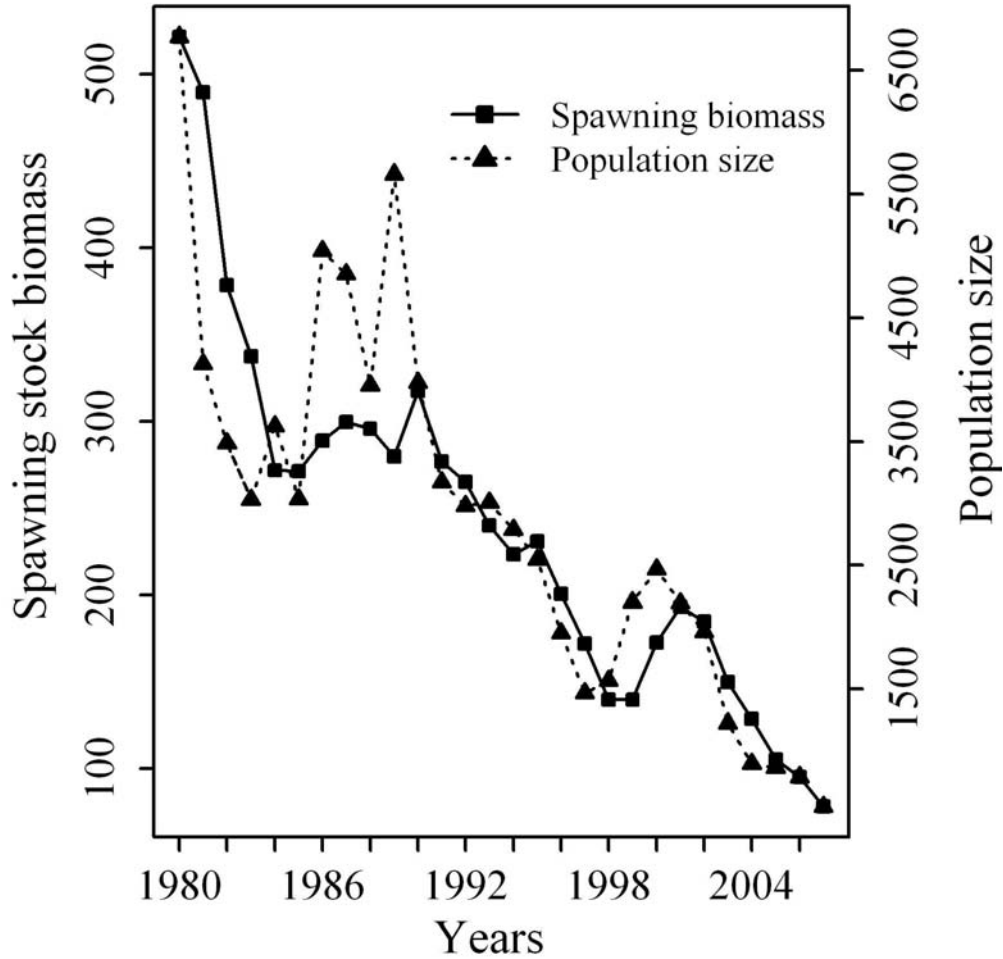


Figure 3: ICES estimation (ICES Working Group) of North Sea whiting population size (thousands of fish) and spawning stock biomass ('000 t) from 1980 to 2007.
74x74mm (600 x 600 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

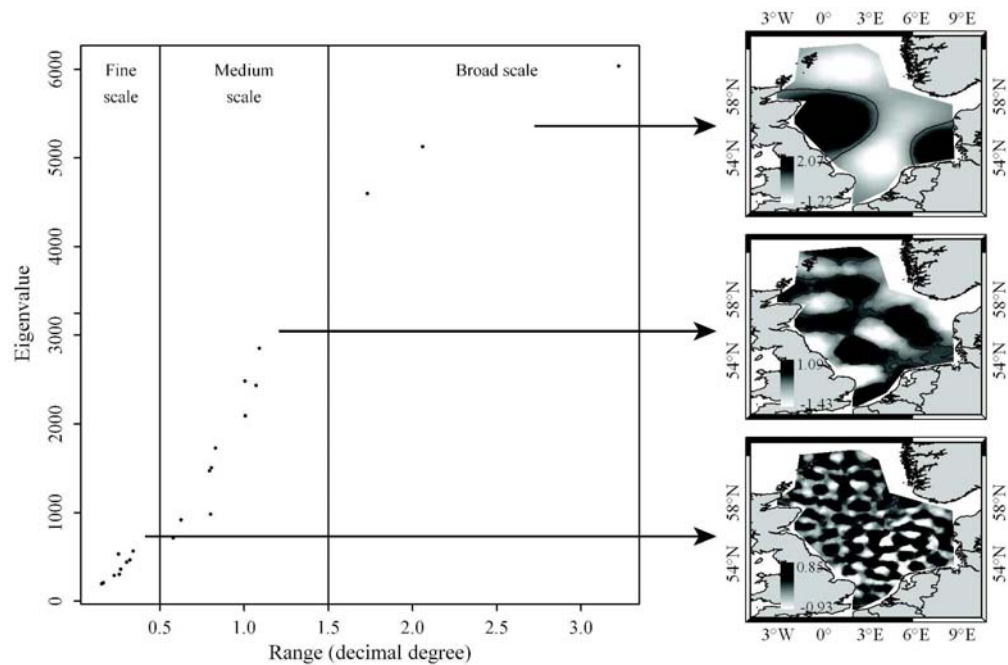


Figure 4: PCNM spatial sub-models. Left, plot of the range values of the 22 selected PCNMs against their eigenvalue. Broad scale, eigenvectors with a range between 166 km and 355 (1.5-3.2 decimal degrees of latitude); medium scale, eigenvectors with a range between 55 km and 166 km (0.5-1.5 decimal degrees of latitude); fine scale, eigenvectors with a range between 0 and 55 km (<0.5 decimal degrees of latitude). Right, maps of the spatial sub-models (arbitrary unit) giving the scale of aggregation of spawning North Sea whiting. Top, broad scale sub-models; center, medium scale sub-models; right, fine scale sub-models. Black lines denote zero values for PCNM.

156x103mm (600 x 600 DPI)

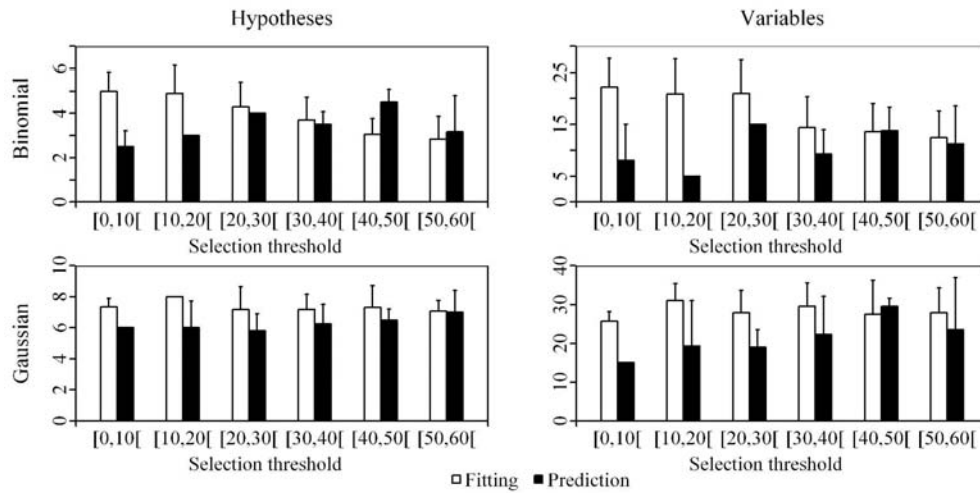


Figure 5: Complexity of selected binomial and Gaussian models. Complexity is represented by the mean number of hypotheses and explanatory variables within selected models. They are given according to the selection procedure and for different value of selection threshold for and LRT between 0 and 60.
160x80mm (600 x 600 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

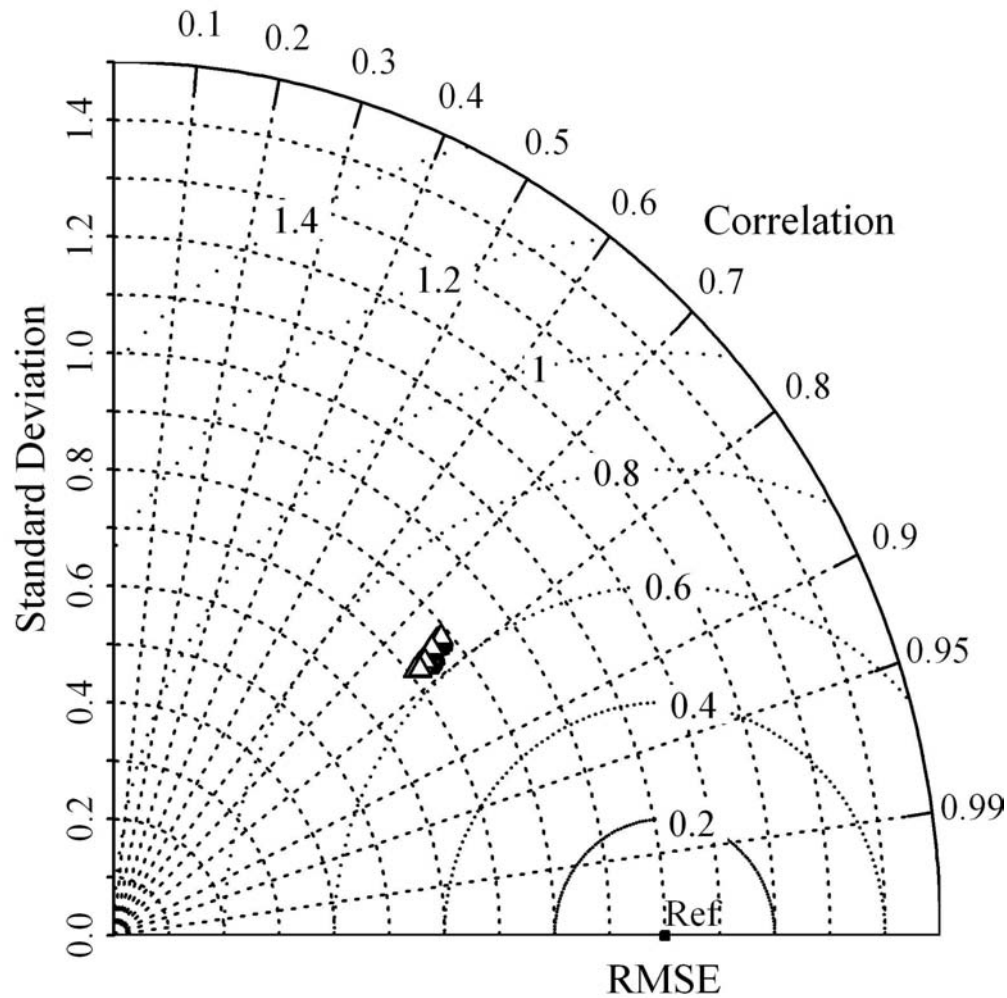


Figure 6: Plot of the Taylor diagram. Only the part representing a positive correlation is shown.

Predictions of delta models (triangle) are compared to the observations using the standard deviation, the root mean square error (RMSE) and the Spearman correlation. Observations are symbolized as a point referred to as the reference point and are normalized so that the standard deviation is equal to 1, the correlation equal to 1 (the correlation between the observations and themselves) and the RMSE equal to 0 (the difference between the observations and themselves). Values of the indices for the predictions are plotted with the standard deviation on the y-axis, the correlation on the radii of the circle and the RMSE on concentric circles around the reference point.

80x80mm (600 x 600 DPI)

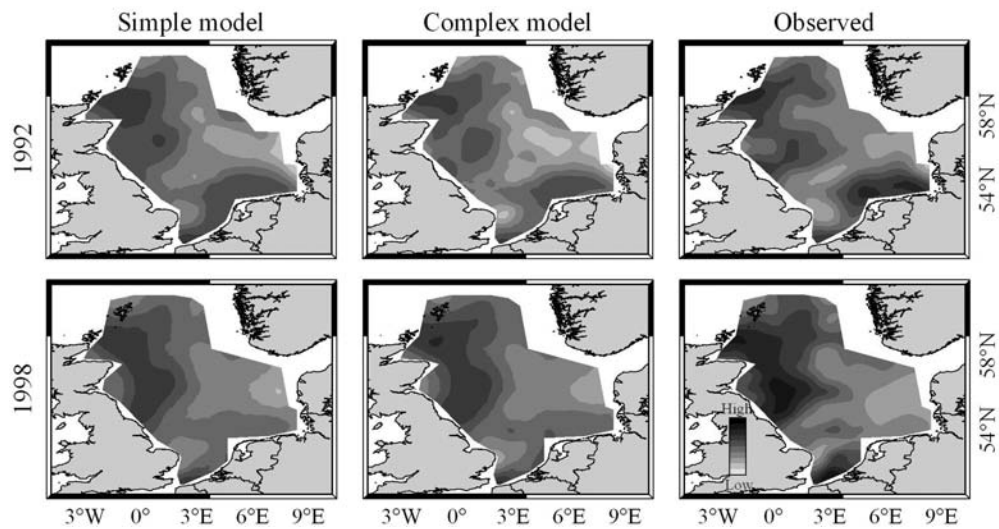


Figure 7: Maps of spawning adult whiting. Maps for the most simple and complex models among those selected (see Table 3 for formulae of the models) were computed for years 1992 and 1998. The corresponding maps of the observed distribution of spawning adults for the same two years are also shown.

160x84mm (300 x 300 DPI)

Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

	Calibration dataset				Prediction dataset			
	Binomial		Gaussian		Binomial		Gaussian	
	Min	Max	Min	Max	Min	Max	Min	Max
Depth	10	195	10	195	10	194	10	194
Temperature	-1.4	9.8	-1.4	9.8	-1.2	9.7	-0.8	9.7
Salinity	26	35.4	26	35.4	26.8	35.4	26.8	35.4
Bedstress	0	2.8	0	2.8	0	2.8	0	2.8
Spawning stock biomass	78,045	299,684	78,045	599,684	139,540	276,842	139,540	276,842
Population Size	549,997	5,661,687	549,997	5,661,687	1,465,866	3,624,792	1,465,866	3,624,792
Age 1	32.04	77.8	32	77.8	49.4	72.3	49.4	72.3
Age 2	10.97	31	10.9	31	15	27.5	15	27.5
Age 3	5.18	19.1	5.1	19.1	5.5	17.2	5.5	17.2
Age 4	1.53	13.3	1.5	13.3	1.9	7.6	1.9	7.6
Age 5	0.35	8.6	0.3	8.6	0.7	2.9	0.7	2.9

Hypotheses	Binomial models	Gaussian models
Geo. Att..	as.factor(StatisticalRectangle)	as.factor(StatisticalRectangle)
N. Pers. Envt.	s(Depth, 4) + s(Bedstress, 4) + as.factor(Sediments)	s(Depth, 4) + s(Bedstress, 4) + as.factor(Sediments)
Pers. Envt.	Temperature + s(Salinity, 3)	s(Temperature, 4) + s(Salinity, 4)
Br. Sc.	s(Ax1, 3) + s(Ax2, 4) + s(Ax4, 4)	s(Ax1, 4) + s(Ax2, 4) + s(Ax4, 4),
Med. Sc.	Ax11 + s(Ax13, 2) + Ax14 + s(Ax16, 4) + s(Ax18, 3) + Ax20 + Ax21 + Ax25 + s(Ax26, 3) + s(Ax28, 1)	s(Ax11, 4) + s(Ax13, 4) + s(Ax14, 4) + s(Ax16, 4) + s(Ax18, 4) + Ax20 + s(Ax21, 4) + s(Ax25, 4) + Ax26 + s(Ax28, 4)
Fine Sc.	s(Ax31, 4) + Ax33 + s(Ax36, 4) + s(Ax39, 1) + s(Ax54, 2) + Ax70 + s(Ax73, 1) + Ax104 + s(Ax110, 2)	s(Ax31, 4) + s(Ax33, 4) + s(Ax36, 3) + s(Ax39, 4) + s(Ax54, 4) + s(Ax70, 3) + s(Ax73, 4) + s(Ax104, 3) + s(Ax110, 4)
Pop. Size	PopulationSize + s(SpawningStockBiomass,1)	PopulationSize + s(SpawningStockBiomass, 4)
Ann. Dem.	s(Age1a, 4) + Age2a + Age3a + Age4a + Age5a	Age1a + Age2a + s(Age3a, 3) + s(Age4a,4) + Age5a
Spa. Dem.	-	s(Age1s, 2) + s(Age2s, 4) + s(Age3s, 1) + s(Age4s, 3) + s(Age5s, 4)
Pop. Mem.	s(PreviousYearAbundance, 2)	s(PreviousYearAbundance, 4)

Models	Formula	
Binomial	N. Pers. Eenvt. + Pop. Mem. (1)	
	N. Pers. Eenvt. + Med. Sc. + Pop. Mem.	
	N. Pers. Eenvt. + Pop. Size + Pop. Mem.	
	N. Pers. Eenvt. + Med. Sc. + Pop. Size + Pop. Mem.	
	N. Pers. Eenvt. + Br. Sc. + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Med. Sc. + Pop. Mem.	
	N. Pers. Eenvt. + Br. Sc. + Pop. Size + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Pop. Size + Pop. Mem.	
	N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Pop. Mem.	
	Pop. Mem.	
	N. Pers. Eenvt. + Med. Sc.	
	Pers. Eenvt. + N. Pers. Eenvt. + Pop. Size + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Med. Sc. + Pop. Size + Pop. Mem.	
	N. Pers. Eenvt. + Fine Sc. + Pop. Mem.	
	Pop. Size + Pop. Mem.	
	Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Pop. Size + Pop. Mem. (2)	
	N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Pop. Size + Pop. Mem.	
	Med. Sc. + Pop. Mem.	
	Pers. Eenvt. + Br. Sc. + Pop. Mem.	
	N. Pers. Eenvt. + Med. Sc. + Fine Sc. + Pop. Mem.	
	N. Pers. Eenvt.	
	Gaussian	Geo. Att. + Pers. Eenvt. + N. Pers. Eenvt. + Pop. Size + Spa. Dem. + Pop. Mem.
		Geo. Att. + Pop. Size + Spa. Dem. + Pop. Mem. (1)
		Pers. Eenvt. + N. Pers. Eenvt. + Med. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.
Geo. Att. + Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Pers. Eenvt. + N. Pers. Eenvt. + Med. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Br. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Med. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Pop. Size + Pop. Mem.		
Geo. Att. + Br. Sc. + Med. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Br. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Pers. Eenvt. + N. Pers. Eenvt. + Br. Sc. + Med. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem. (2)		
Geo. Att. + N. Pers. Eenvt. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Med. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Br. Sc. + Med. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + Pers. Eenvt. + N. Pers. Eenvt. + Med. Sc. + Fine Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		
Geo. Att. + N. Pers. Eenvt. + Br. Sc. + Pop. Size + Spa. Dem. + Pop. Mem.		

LRT	Δi
0	54.34
9.21	10.10
17.54	46.96
27.54	5.71
31.27	12.41
33.82	17.46
38.51	36.80
38.90	10.02
42.65	9.87
44.40	11.39
45.59	15.11
45.70	0.98
50.36	73.41
50.47	21.24
51.43	32.80
52.43	7.59
53.47	38.21
53.80	61.15
53.80	10.49
54.59	0
55.52	27.97
57.93	36.53
58.89	8.68
59.60	83.98
0	42.65
11.21	160.69
16.50	244.76
16.95	55.97
21.69	106.13
21.74	50.90
23.47	166.84
23.69	162.04
26.93	61.75
30.22	166.72
34.04	116.02
37.27	245.29
39.13	90.28
40.37	118.85
47.12	129.28
53.37	0.10
58.74	97.84

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

	Binomial models						Gaussian models					
	[0-10[[0-20[[0-30[[0-40[[0-50[[0-60[[0-10[[0-20[[0-30[[0-40[[0-50[[0-60[
Geo. Att..	0	0	0	0	0	0	++++	++++	++++	++++	++++	++++
N. Pers. Eenvt.	++++	++++	++++	++++	++++	++++	++++	++++	+++	+++	++	+++
Pers. Eenvt.	0	0	0	++	++	++	++++	++++	+++	++	++	++
Br. Sc.	0	0	0	+	+++	++	0	+	++	++	++	++
Med. Sc.	+++	++	+++	++	++	++	0	+	++	++	++	++
Fine Sc.	0	0	0	0	0	+	0	+	+	++	++	++
Pop. Size	0	++	+++	+	++	++	++++	++++	++++	++++	++++	++++
Ann. Dem.	0	0	0	0	0	0	0	0	0	0	0	0
Spa. Dem.	-	-	-	-	-	-	++++	++++	++++	++++	++++	++++
Pop. Mem.	++++	++++	++++	++++	++++	++++	++++	++++	++++	++++	++++	++++

For Peer Review