Stock Annex: Turbot (*Scophthalmus maximus*) in Subarea 4 (North Sea)

Stock specific documentation of standard assessment procedures used by ICES.

Stock:	Turbot
Working Group:	Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (WGNSSK)
Created:	
Authors:	
Last updated:	October 2018
Last updated by:	Jurgen Batsleer (IBPturbot2018)

A. General

A.1. Stock definition

Genetic studies within restricted geographical areas had illustrated the presence of distinct turbot populations in the Baltic and Irish Seas using neutral markers (e.g. Delbare and Declerck, 1999; Nielsen *et al*., 2004). Over the period 2006–20120, a genetic study of turbot population structure all over the species' distribution area has been conducted using both neutral and gene-associated genetic markers (Vandamme *et al.*, 2014). The neutral marker panel confirmed the break-up between the Baltic and Northeast Atlantic clusters. Within the latter, a more detailed pattern of genetic differentiation could be observed when gene-associated markers were also included in the analysis. This full analysis suggests a break between the southern and central parts of the North Sea, making turbot from the southern North Sea genetically more similar to those from the Western Waters. However, because it is unknown whether there are also differences in life history within the North Sea, and information on the number and location of spawning aggregations is missing, the break between the southern (4.c) and central parts (4.a and 4.b) is insufficiently supported to be recommended for management purposes Additionally, it is logistically difficult to split the North Sea into several management and assessment units. The proposed stock structure is represented in Figure A.1 below.

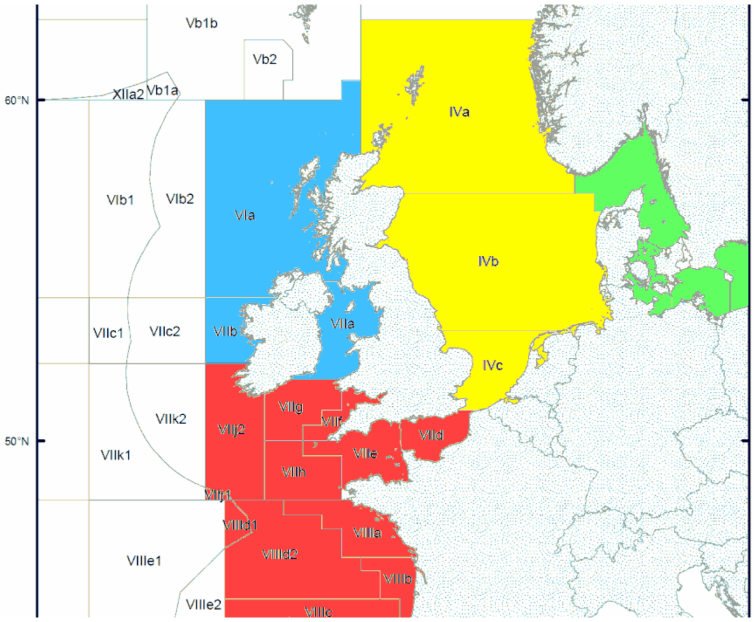

Figure A.1. Stock structure of turbot in the Northeast Atlantic as proposed by IBPNew 2012.

A.2. Fishery

In the 1950s the UK was the biggest contributor to the landings, with almost 50% of the landings coming from this country. In that early period, the landings fluctuated around 6000 tonnes per year. The last decade, the landings are around 3400 tonnes per year. Most of the landings stem from the Netherlands that contributes between 50 and 60%. Within the Netherlands most of the landings come from the 80 mm beam trawl fleet fishing for flatfish species sole and plaice. Also in most other countries turbot is caught in mixed fisheries trawls. The second largest contributor to the landings in the last decade is Denmark (between 15 and 20%). In Denmark there is a directed fishery for turbot using gillnets.

Within the Netherlands, most of the landings come from the Southern Bight and the German Bight. In Belgium, turbot is mainly caught in mid-class (301–900 Hp) and large (>900 Hp) beam trawlers. These vessels are mostly flatfish directed, particularly towards plaice and sole, together with the associated bycatch species such as turbot, brill, dab, lemon sole, anglerfish and some roundfish. In Danish fisheries, turbot is taken only as bycatch. In the North Sea, where most of the Danish landings of turbot are taken, the gillnet fishery accounts for almost half of the landings.

Before 1999, no information on discarding is available. In at least part of that period an EU-wide minimum landings size (MLS) of 30 cm was enforced. However, this minimum landings size was abandoned and member states have set their own MLS rules and regulations. For example, Belgium has a MLS of 30 cm, while the Netherlands implemented a minimum size of 25 cm. In recent years (2015–2017) Dutch Producer Organisations took measures, e.g. a gradually increase in MLS from 27 to 32 cm, to avoid an early exhaustion of the quota (Table A.1). Such measures may influence discard decisions in the fleet, causing an increase in discarding of the lower market size (younger age-classes) turbot in the Dutch flatfish fleet. Discard information for the Netherlands are available for the period 1999 to 2016. These data come from two Dutch monitoring programmes. An observer programme that has been carried out since 1999 mainly on the Dutch flatfish beam trawl fleet and a self-sampling programme in a reference fleet that is set up in 2010. From 2011 onwards, observer trips have taken place on board fishing vessels operating in the reference fleet. In 2017, most countries provided estimates of discards in 2015 to InterCatch.

Table A.1. Measures taken by the Dutch Producer Organisations from 2016 up to present.

Dutch PO-Measures			
Year	**Date**	**Max kg per week/trip**	**MLS**
2016	January	-	27 cm
2016	April	-	30 cm
2016	May	-	32 cm
2016	October	375 kg	32 cm
2016	November	600 kg	32 cm
2017	January	-	32 cm
2017	March	800 kg	32 cm
2017	November	2000 kg	30 cm

Conservation schemes and technical conservation measures

Fishing effort has been restricted for demersal fleets in a number of EC regulations (EC Council Regulation No. 2056/2001; EC Council Regulation No 51/2006; e.g. N°40/2008, annex IIa). For example, for 2007, Council Regulation (EC) No 41/2007 allocated different days at sea depending on gear, mesh size, and catch composition: Beam Trawls could fish between 123 and 143 days per year. Trawls or Danish seines could fish between 103 and 280 days per year. Gillnets could allowed to fish between 140 and 162 days per year. Trammelnets could fish between 140 and 205 days per year.

Several technical measures are applicable to the flatfish fishery in the North Sea: mesh size regulations, minimum landing size, gear restrictions and a closed area (the plaice box).

Mesh size regulations for towed trawl gears require that vessels fishing north of 55°N (or 56°N east of 5°E, since January 2000) should have a minimum mesh size of 100 mm, while to the south of this limit, where the majority the plaice fishery takes place, an 80 mm mesh is allowed. In the fishery with fixed gears a minimum mesh size of 100 mm is required. In addition to this, since 2002 a small part of North Sea plaice fishery is affected by the additional cod recovery plan (EU regulation 1342/2008) that prohibits trawl fisheries with a mesh size <120 mm in the area to the north of 56°N.

The maximum aggregated beam length of beam trawlers is 24 m. In the 12 nautical mile zone and in the plaice box the maximum aggregated beam-length is 9 m. A closed area has been in operation since 1989 (the plaice box). Since 1995 this area was closed in all quarters. The closed area applies to vessels using towed gears, but vessels smaller than 300 HP are exempted from the regulation.

B. Data

B.1. Commercial catch

Landings

The landings of turbot are available through the EuroStat database. This database holds the officially recorded landings for all countries landing turbot in the North Sea. There are no records for the Dutch landings in the EuroStat database between 1984 and 1987. However, for the North Sea these missing landings have been estimated in a Dutch/Belgian research project, and have been used to fill in the gaps (Boon and Delbare, 2000).

In the 1950s, the UK was the biggest contributor to the landings, with almost 50% of the landings coming from this country. In that early period, the landings fluctuated around 6000 tonnes per year. In the last decade, the landings are around 3400 tonnes per year. Most of the landings stem from the Netherlands that contributes between 50 and 60%. Within the Netherlands most of the landings come from the 80 mm beam trawl fleet fishing for flatfish species sole and plaice. Also in most other countries turbot is caught in mixed fisheries trawls. The second largest contributor to the landings in the last decade is Denmark, where there is a directed fishery for turbot using gillnets.

There is no long-term continuous programme for age sampling of landings in any of the countries. Therefore, the age structure of the landings is estimated using data from different sources in different time periods. What follows is a brief description of all data sources that are available.

From 1975, there is a four-year time period for which the age structure of the landings has been estimated by Weber (1979). This age structure is estimated from market samples taken in Cuxhaven and Hamburg, and from samples collected on board research vessel surveys. Most of the samples represent landings in the eastern part of the North Sea (27.4.b). The age structure is estimated from a total of 9360 length and 6389 weight measurements combined with 6788 age samples. The total numbers-at-age in the overall landings were estimated in the original publication by combining the age samples with the quarterly landings for England, the Netherlands and Germany. The German data were in the end not included in the assessment because of concerns about data quality and a potential bias because less than 20% of the landings stem from the German fleets at that time.

The second dataset from 1981 to 1990 is derived from landings in the Netherlands and reported in the "Datubras" project (Boon and Delbare, 2000). A stratified sampling scheme was used to collect the samples, using quarters, auctions, and market categories as stratification levels. Between 398 and 862 age samples were taken annually for age determination of fish. Most of the samples represent areas 4.b and 4.c. The Dutch data are subsequently raised to the total international landings.

The third dataset spans the period 2000–2002. It was supplied by Cefas and based on the UK landings of turbot. These were raised on an annual basis to the total landings. The UK data represented only 10–15% of the catches in this period and consisted of older ages compared to other fisheries.

The fourth and final dataset stems again from the Netherlands. It spans the years 1998 and 2004–2016. The age structure is estimated from stratified sampling accounting for auctions, quarters and market categories. These are raised to total Dutch landings by quarter. Between 494 and 1921 age samples were taken per year. The total Dutch landings are subsequently raised to the total international landings per year.

In preparation of the inter-benchmark, Danish age and length structure for landings were made available. The age structure was available for 2014–2016, and the length structure was available for 2002–2016. The age data were imported into InterCatch, and used to raise new age structures for landings in the period 2014–2016.

During the Interbenchmark in 2017, the different data-sources were trialled in a sensitivity analysis of the baserun:

- The German data showed not to be informative in the estimation of trends. To maintain an evenly consistent time-series, it was decided to drop the German catch-at-age data.
- The UK catch-at-age data (2000–2002) were excluded given the uncertainty around these data and selects older ages compared to other fisheries.
- The Dutch data are included and raised to the total international landings per year.
- The Danish age samples were found to be of good quality and not inconsistent with the Dutch data. Since this dataset provides valuable age composition information (other métiers than those from the Dutch sampling programme) and the data collection effort is ongoing, the data are used in the assessment.

By excluding the German and UK catch-at-age data, the final set of catch-at-age data, using Dutch and Danish age information, spans the time period from 1981 to present, coinciding with the start of the Dutch catch-at-age data.

Discards

Before 1999, no information on discarding is available. In at least part of that period an EU-wide minimum landings size (MLS) of 30 cm was enforced. However, this minimum landings size was abandoned and member states have set their own MLS rules and regulations. For example, Belgium has a MLS of 30 cm, while the Netherlands implemented a minimum size of 25 cm. In recent years (2015–2017), Dutch Producer Organisations took measures, e.g. a gradual increase in MLS from 27 to 32 cm, to avoid an early exhaustion of the quota (Table A.1). Such measures may influence discard decisions in the fleet, causing an increase in discarding of the lower market size (younger age-classes) turbot in the Dutch flatfish fleet. Discard information for the Netherlands are available for the period 1999 to 2017. These data come from two Dutch monitoring programmes. An observer programme that has been carried out since 1999 mainly on the Dutch flatfish beam trawl fleet and a self-sampling programme in a reference fleet that is set up in 2010. From 2011 onwards, observer trips have taken place on board fishing vessels operating in the reference fleet. The discard data are used for producing an index of discarding (dpue) over time (1999–2016) (Figure B.1). The dpues obtained from both programmes show an increase in most recent years, but absolute differences are great. The increase may be related to the measures taken by the Dutch POs, resulting in a decrease in the landings of age two turbot. It is important to note that these dpues are based on a small number of observations (i.e. a small number of trips were sampled). During the Interbenchmark, different configurations of the dpue were analysed (per gear, per quarter, per monitoring programme), but due to the small number of samples the discard indices were not used in the assessment.

In 2017, most countries provided estimates of discards in 2015 to InterCatch. However, there is very limited age sampling of the discards. Very few fish were sampled in the discards of some Danish métiers (<10 per métier) which is not enough to be used in the raising of international landings.

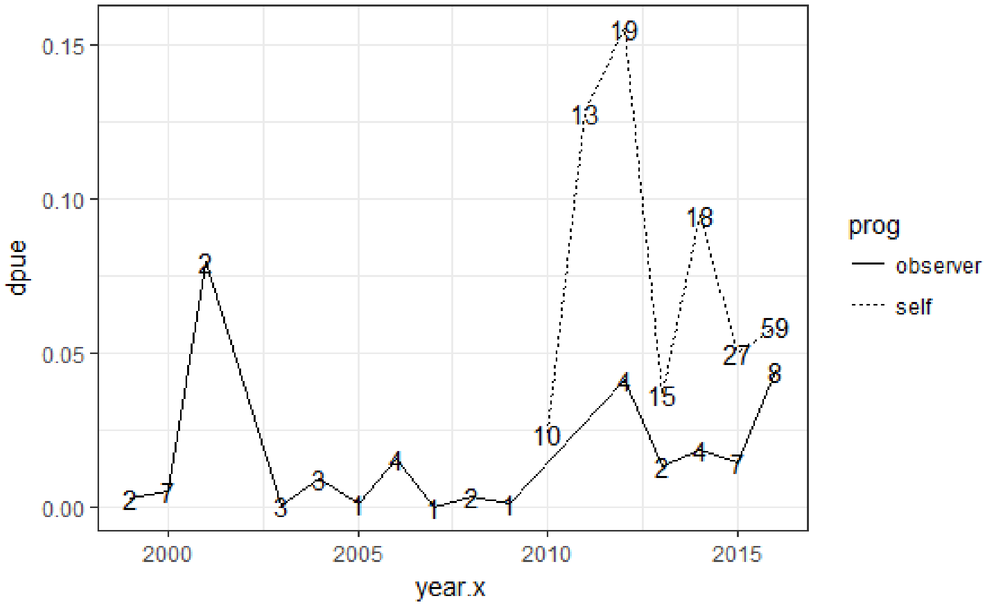

Figure B.1 Dpues of both discard monitoring programmes (self-sampling; dashed line, and observer: solid line). Number of observations (trip sampled) per year are shown.

In general, only an incomplete catch-at-age matrix is available that needs to be reconstructed by modelling for the years 1991 to 1997 and 1999 to 2001. Age information for recent years is mainly available only from the Dutch fishing fleets representing more than 50% of total landings. Data from Denmark were available for the three most recent years. Danish data demonstrated a shift towards older fish compared to the Dutch data. This may also be true for other countries but no information is available. Although the inclusion of Danish data impacted the overall catch-at-age matrix only to a minor extent, this highlights the need to get catch-at-age data not only from the Netherlands but also from other countries. Also, retrospective data would be highly beneficial if available.

The 2018 Inter-benchmark reviewed the plus-group settings. For the analysis, data from 1981 to present and age 1 to 10 were chosen as initial values. Sensitivity runs were performed using a step-wise reduction of the plus-group to 6. A comparison of the runs showed small differences between the runs with a plus-group of 10 to 8. Using a smaller plus-group of 6 or 7 results in a lower estimate of the SSB and higher estimate of Fbar, while there is no influence on the recruitement. The runs were critically reviewed based on Mohn's rho as well as model diagnostics. Solely based on the Mohn's rho a plus-group of 9 would be preferable. Model diagnostics, however, showed a more consistent selectivity pattern throughout the time period of the assessment for the plus-group of 8. In addition, the selectivity for the older ages seems more stable. Given the better selectivity pattern and small differences in model outcomes, it was agreed to use a plus-group of 8 for the catch-at-age.

B.2. Biological Data

B.2.1. Weight-at-age

Weight-at-age data in the catch for this stock are available for most but not all of the years during which there is age sampling of the landings (Figure B.2). Data are available for the period 1981–1990 from the DATUBRAS database (Boon and Delbare, 2000), and then again for the years 1998, and 2004 to present from Dutch market sampling.

Stock weights are estimated as the catch weights in Q2, coinciding with peak spawning of the stock. Hence stock weights estimates are available for the same time period, but excluding the years 2005 and 2006 where no samples were available in the second quarter. In addition to this average weights-at-age for the stock during the period 1976–1979 are available from Weber (1979). For both the catch and stock weights, estimated values for ages 6 and greater tend to show large interannual fluctuations, due to the limited number of fish sampled at these ages. The vast majority of landings are for ages 4 and younger and this is reflected in the number of samples for these ages.

With no data except a single year available in the 1990s (1998) modelling was required to infer the trend in weight-at-age over the period 1991 to 2003. The group decided that using a constant annual weight-at-age vector over the entire period as input to the stock assessment models would be inappropriate, especially since significant increases in weight-at-age have been observed for other flatfish species in the North Sea during this time. Hence a time-varying growth model was included in the assessment. This growth model determines the catch weights-at-age and stock weight-at age. Stock weights-at age are defined in a two-step process. First, time varying length-at-age is modelled using a von Bertalanffy growth model where length-at-age a (in m) in a given year t is calculated:

$$L_{a,t} = L_{\infty t}\,(1\text{-exp}(-K(a + a_0)))$$

where $L_{\infty t}$ is the asymptotic length in year t, K is a curvature parameter, and a_0 determines the point in time when the fish has zero length. Stock weights-at-age in a given year $W^S_{a,t}$ (in kg) are calculated using an allometric growth model:

$$W^S_{a,t} \ = \ \alpha \ L_{a,t}{}^{\beta}$$

With parameters α= 0.00001508 and β=3.090, as estimated by Bedford *et al.* (1986). Catch weights-at-age $W^C_{a,t}$are linked to stock weights-at-age by a simple age-independent scaling factor such that $W^C_{a,t} = \gamma W^S_{a,t}$.

Linking the weights–at-age model to data

The weights model has three scalar parameters: K, a_0 and γ. In addition, there is a vector $L_{\infty t}$ that needs to be estimated. This vector is a complex function of time and specifying an *a priori* shape may not fully address the multitude of processes that take place in shaping its functional form. Therefore, we used a smooth function of time, constructed using a number of b-spline basis functions (de Boor, 2001). These functions can be viewed as transformations of the explanatory variable t. For simplicity, the number of parameters (and the flexibility of the resulting $L_{\infty t}$) is taken to be equal to the number of parameters used for the spline describing the variation of fishing mortality over time.

The parameter fitting is done in the likelihood function of the assessment model. In short, the available observations of catch weights-at-age and landings weights-at-age are used in a likelihood component of the model that assumes a normal distribution of errors in the observations, with age-dependent standard deviations σ^S_a and σ^C_a for the stock and catch weights, estimated for each age separately.

Results for the fitted landings weights–at-age and stock weights-at age for the assessment model with nine ages are found in Figure B.2. Clearly L_∞ has changed over time, increasing in the period 1975–1990, and decreasing in the period 1995–2010. This pattern is also observed in other flatfish in the North Sea. Residual variance appears to increase with age. Most of the data appear to fit the model quite well, apart from the Weber data for which the older ages are overestimated by the model.

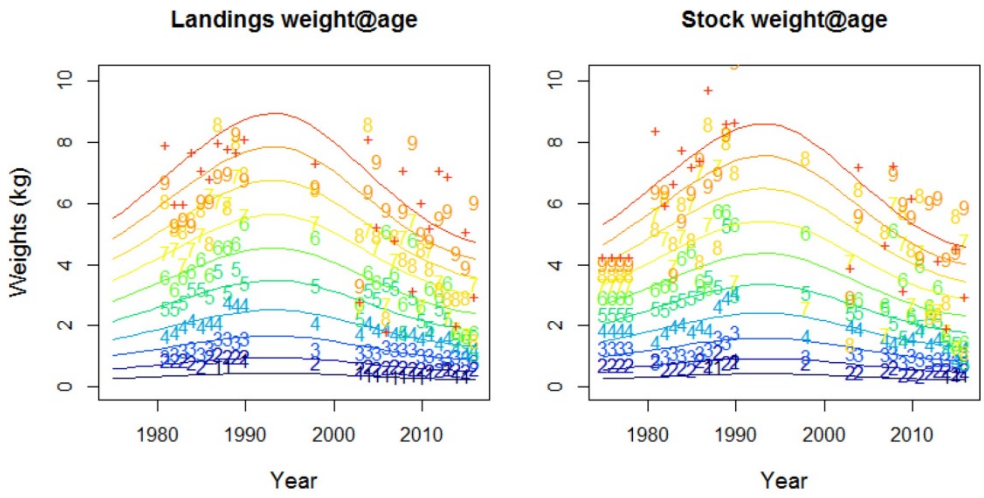

Figure B.2. (Top left) Landings weights assuming gradually changing weights-at-age, following a von Bertalanffy growth curve, (to right) stock weights assuming gradually changing weights-at-age, following a von Bertalanffy growth curve. L∞t is a 5 parameter spline in this example.

B.3. Surveys

Two scientific survey-series catching turbot are available. The Beam Trawl Survey (BTS-ISIS), and the Sole Net Survey (SNS). The BTS_ISIS is an offshore beam trawl survey designed to catch demersal species. The survey is performed in quarter 3. The index is based on the catch in one of the two nets and on catches between 52 and 239 individuals per year. The number of individuals used to generate an age–length key can be larger than the number of individuals used for the index, because the index is based on only the catch in one of the two nets, while age samples can be taken from both nets. The years included in the assessment run from 1991 to present (2017) and ages used are 1 to 7 (7 as a true age).

The procedure to create an age-structured index series from the BTS-ISIS was updated prior to the working group. Previously, each individual fish caught was linked to an age–length key based on its length. The age–length key was based on all age samples in the BTS survey since 1991. The updated procedure first links the individual fish from which otoliths are taken to the length sample. This allows direct ageing of the fish in the index. Those fish for which no direct age sample is available are then assigned to ages using the age–length key based on all fish in the period 1991–present.

The SNS is a nearshore beam trawl survey designed to monitor flatfish fauna and samples transects further offshore than the other inshore surveys. The SNS survey area overlaps with those of the Dutch DFS (inshore) and BTS-Isis (offshore) (ICES, 2016). It is also performed in quarter 3. The years included in the assessment run from 2004 to present (2017) and ages used are 1 to 6 (6 as a true age). Note that the age classes of the SNS survey used in the assessment are different compared to the previous assessment. This was decided after evaluating the results of both age-class configurations (from previous assessment vs. current age configuration) in the new assessment.

Within the Interbenchmark several other surveys were trialled. A BTS index using Delta-GAM standardisation including:

- The Dutch BTS-ISIS survey
- The Dutch BTS-Tridens survye

- The Belgian BTS survey
- The German BTS-Solea survey

The index spanned the period 2002–2016 and included ages 1–5. After testing the Delta-GAM BTS-index the group chose to keep using the BTS-ISIS. The Delta-GAM BTS-index did not improve the model diagnostics or retrospective patterns and resulted in a reduction in the length of the survey data time-series.

Also, the International Bottom Trawl Survey (IBTS) in quarter 1 and 3 was evaluated. The IBTS turbot data were trialled in several variations of the data: in cpue, numbers per hours and exploitable biomass. Including the IBTS turbot data did not improve the assessment since observation variances are high in all cases except for number-per-hour. The assessment model was however not taillored to fit this kind of data and is considered therefore invalid. In addition, catches of turbot in the IBTS are so minute, the group decide to exclude the IBTS from the assessment.

B.4. Commercial cpue

In addition to the survey based indices, there is also an index based on the Dutch 80 mm beam trawl fleet lpue. Since 2009, the fleet has transitioned to innovative gear types, replacing the beam with a wing design and replacing the tickler chains with electrical pulse stimuli. These changes may result in different catchabilities. Figure 3.8 shows the individual lpues for these fleets, clearly showing markedly different absolute lpue values, while trends in especially pulse and traditional beam trawl seem similar.

Not only their catchabilities are different, also their predominant fishing grounds differ as well. The pulse fleet now inhabits the more traditional sole grounds in the southern North Sea, while the remaining traditional beam trawlers fish in the northerly areas of the original distribution range, in areas where more turbot is caught. Therefore, we have to correct for gear and area when standarizing the lpue.

This lpue time-series is standardised by building a statistical model that includes interactions in space, time and gear. Raw lpues are calculated per trip and per ICES rectangle. The fishing effort per rectangle is then taken as a weighting factor in the analysis. Only those rectangles where fishing occurred in eleven or more years are then used. This dataset amounted to 99% of all turbot catches since 1995. The eleven years are considered an arbitrary number, but restricts the statistical model to predict effort in areas that are poorly sampled. Sensitivities were executed by reducing or expanding the eleven years without any effect on model results.

Several different model configurations were tested and analysed (Table B.1). In general, residuals did not show patterns and all configurations seem statistically appropriate. AIC and BIC criteria were calculated.

Although model D shows a clear drop in AIC, the increased use of parameters is substantial as shown with the BIC criteria. All standardized lpue models were trialled in the turbot assessment. After comparing the results, the inter-benchmark decided to use the model D in the turbot assessment.

Table B.1 Description of standardized lpue models and corresponding AIC and BIC criteria.

MODEL	DESCRIPTION	AIC	BIC
A	te(SI_LONG,SI_LATI, k = 5) + as.factor(year) + LE_GEAR_INNOV	2935	3180
B	te(SI_LONG,SI_LATI, year,k = 5)+LE_GEAR_INNOV	2656	3112
C	te(SI_LONG,SI_LATI, k = 5) + te(year, k = 10) + LE_GEAR_INNOV	3053	3230
D	te(SI_LONG,SI_LATI, by = as.factor(year), k = 5) + as.factor (year, k = 10) + LE_GEAR_INNOV	1618	3540

B.5 Internal consistency

The available scientific surveys have a low internal consistency especially for older ages leading to a low ability to track cohorts over time. Because of this, the assessment is strongly influenced by a Dutch lpue index. This index has been standardized for changes in fishing areas and gears used (i.e. traditional beam trawls vs. pulse trawls). It is also used as exploitable biomass index without age information to avoid that the catch-at-age matrix is used twice. However, a scientific survey with higher catch rates for turbot and a better internal consistency would be preferable.

C. Assessment: data and method

The state–space model SAM (Nielsen and Berg, 2014) offers a flexible way of describing the entire system, with relative few model parameters. It allows for objective estimation of important variance parameters, leaving out the need for subjective *ad-hoc* adjustment numbers, which is desirable when managing natural resources.

The total vector of model parameters for this model is:

$$\vartheta = (Q_{s=1,\,a=1-6+},\, Q_{s=2,\,a=1-6+},\, Q_{s=3},\, \sigma^2_R,\, \sigma^2_{S,a=1,2+},\, \sigma^2_{F,\,a=2,(1,3-10)},\, \sigma^2_{\circ,\,a=1,2,3+},\, \sigma^2_{s=1,\,a=1,2+},\, \sigma^2_{s=2,\,a=1,2+},\, \rho)$$

The Q parameters are catchabilities corresponding to the survey fleets (these parameters are survey- and age-specific for the SNS and BTS-ISIS, covering ages 1–6+, and a single value for the NL_BT2 exploitable biomass index). The variance parameters σ^2_R, $\sigma^2_{S,a=1,2+}$, and $\sigma^2_{F,\,a=2,(1,3-10)}$ are process variances for recruitment, survival, and development in fishing mortality respectively (the survival separately for age 1 and 2+ and the fishing mortality separately for age 2 and the remaining ages (1, 3–10+). The remaining σ^2 parameters are describing the variance of different observations divided into fleet and age classes. Finally ρ is the correlation parameter (among the ages) for the random walks on the fishing mortalities.

The WKNSEA benchmark introduced an extension to allow for varying correlation between different ages by setting the correlation of the log F annual increments to be a simple function of the age difference (AR(1) process over the ages). By doing this, individual log F processes will develop correlated in time, but in such a way that neighbouring age classes have more similar fishing mortalities than more distant ones. This correlation structure does not introduce additional parameters to the model, and is referred to below as an AR correlation structure (see Nielsen and Berg, 2014 for more details). This approach is used in the turbot assessment as well.

C.1. Model used

SAM

Software used: Source code and all scripts of the Interbenchmark are freely available at https://github.com/ices-eg/wg_IBPTur.27.4 as well as full diagnostic plots. Stock = '**TUR-nsea**'

Model Options chosen:

A configuration file is used to set up the model run once the data files, in the usual Lowestoft format, have been prepared. During the 2018 IBP, the SAM model configuration for the turbot assessment was evaluated. Sensitivity runs were carried out with various combinations of parameter bindings. The initial binding was set to provide the model parameters most possible freedom, followed by binding parameters together where needed when the AIC criteria indicated a more restrictive model was prefered. In total, 41 sensitivity runs were performed to find an optimum in number of free parameters versus model fit (AIC) and Mohn's rho (both being minimized). All outputs can be found at the Github repository of the 2018 Inter-benchmark. The final SAM model configuration has the following form:

SAM configuration file

```
# Min Age
1
# Max Age
8
# Max Age considered a plus group (0=No, 1=Yes)
1
# The following matrix describes the coupling  of fishing mortality STATES
# Row represent Catch, Columns represent ages.
1	2	3	4	5	6	7	7
# Use correlated random walks for the fishing mortalities
# ( 0 = independent, 1 = correlation estimated, 2=AR1)
2
# Coupling of catchability PARAMETERS (Surveys)
# Row represent fleets (SNS and BTS only; lpue age-aggregated), Columns represent ages.
1	1	2	3	3	3	0	0
4	4	5	5	6	6	6	0
7	0	0	0	0	0	0	0
# Coupling of power law model EXPONENTS
(not used)
# Coupling of fishing mortality RW VARIANCES
1	2	3	3	4	4	5	5
# Coupling of log N RW VARIANCES
1	2	2	2	2	2	2	2
# Coupling of OBSERVATION VARIANCES
# Row represent fleets (Catch, SNS, BTS, lpue age-aggregated), Columns represent ages.
1	2	3	3	4	4	5	5
6	6	7	8	8	8	0	0
9	9	9	10	11	11	11	0
12	0	0	0	0	0	0	0
# Coupling of SURVEY CORRELATION CORRECTION BY AGE
# Row represent fleets (Catch, SNS, BTS, lpue age-aggregated), Columns represent corre-
lated ages.
NA	NA	NA	NA	NA	NA	NA	NA
1	1	1	1	1	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA
# Stock–recruitment model code (0=RW, 1=Ricker, 2=BH, ... more in time)
0
# Indicator for LPUE time series (biomass treatment) (0 = SSB, 1 = catch, 2 = exploitable
biomass)
2
# Years in which catch data are to be scaled by an estimated parameter
(Catch not scaled)
# Define FBAR range
2–6
```

This configuration has the following features:

- The assessment time-series starts in 1981.
- The plus group of the stock assessment is set to 8+.
- Fishing mortality states are the same for ages 7+. There is no survey data for ages older than 8 and the catch numbers at these ages are low and highly variable (in part due to limited samples used in raising the data).
- The correlated random walks for the fishing mortalities was set at 2 (= cor-rela-tion between ages but declines following a power function when dis-tance be-tween ages increases). This run provided the best AIC.
- The catchability of the two youngest and three oldest ages of the two age-structured surveys are linked. For the BTS catchabilities for ages 3 and 4 are linked as well.
- The variance on the fishing mortality random walks is estimated separately linking ages three and four, five and six and the ages 7+. This was done to account for the likely impact on F for this ages.
- The variance on the log N random walk for age 1 is estimated separately.
- Compared to the 2017 IBP there is one parameter extra freedom in the catch and SNS. Observation variance for the catch is estimated seperately for the young ages 1 and 2, linking ages three and four, five and six and the ages 7+. The observation variance for age 3 in the SNS and age 4 of BTS are esti-mated seperately, whilehe observation variance for the oldest ages are linked.
- LPUE time-series indicator: set to exploitable biomass
- No stock–recruit function is used in the model (recruit follows a random walk).
- No catch scaling is done.

Input data types and characteristics:

Type	Name	Year range	Age range	Variable from year to year Yes/No
Canum	Catch-at-age in numbers	1975–1978, 1981–1990, 1998, 2000–2002, 2004–now	1–10+	Yes
Weca	Weight-at-age in the commercial catch	1981–1990, 1998, 2000–2002, 2004–now	1–10+	Yes – modelled values used in assessment
West	Weight-at-age of the spawning stock at spawning time.	1981–1990, 1998, 2000–2002, 2004, 2007–now	1–10+	Yes – modelled values used in assessment
Mprop	Proportion of natural mortality before spawning	1975–now	1–10+	No, assumed 0
Fprop	Proportion of fishing mortality before spawning	1975–now	1–10+	No, assumed 0
Matprop	Proportion mature at age	1975–now	1–10+	No, assumed constant over years
Natmor	Natural mortality	1975–now	1–10+	No, assumed constant over ages and years

Tuning data:

Type	Name	Year range	Age range
Tuning fleet 1	SNS	2004–now	1–6
Tuning fleet 2	BTS ISIS	1991–now	1–7
Tuning fleet 3	NL Beam trawl fleet	1995–now	Exploitable biomass

Assessment settings used in the final assessment:

Year	2017 (IBPTurbot proposal)
Model	SAM
First tuning year	1981
Last data year	2016
Ages	1–8+
Plus group	Yes
Stock weights-at-age	von Bertalanffy growth curve with time varying Linf
Catch weights-at-age	von Bertalanffy growth curve with time varying Linf
Total Landings	Not used
Landings-at-age	1981–1990, 1998, 2000–present
Discards	Not used (assumed 0)
Abundance indices	BTS-Isis 1991–2017 SNS 2004–2017
	Standardized NL-BT2 lpue age-aggregated catchable biomass 1995–2017

The final assessment of Turbot in 4 is given below (Figure C.1). The assessment fits the Dutch LPUE and catch at ages 3 and 4 very well. The fit to other ages and survey data is markedly lower. The estimates are all associated with low uncertainty which is likely due to smoothed stock and catch weights at age and fixed maturity-at-age data. Process error is generally low and does not show any trends. No estimated parameters show inappropriate correlation structures which indicates that most parameters are estimated independently. The model converges easily and has no issues in running retrospective analyses. Furthermore, all parameters are estimated well and the uncertainty in the parameter estimates is low as well. Residual plots do not show clear patterns in either positive or negative residuals.

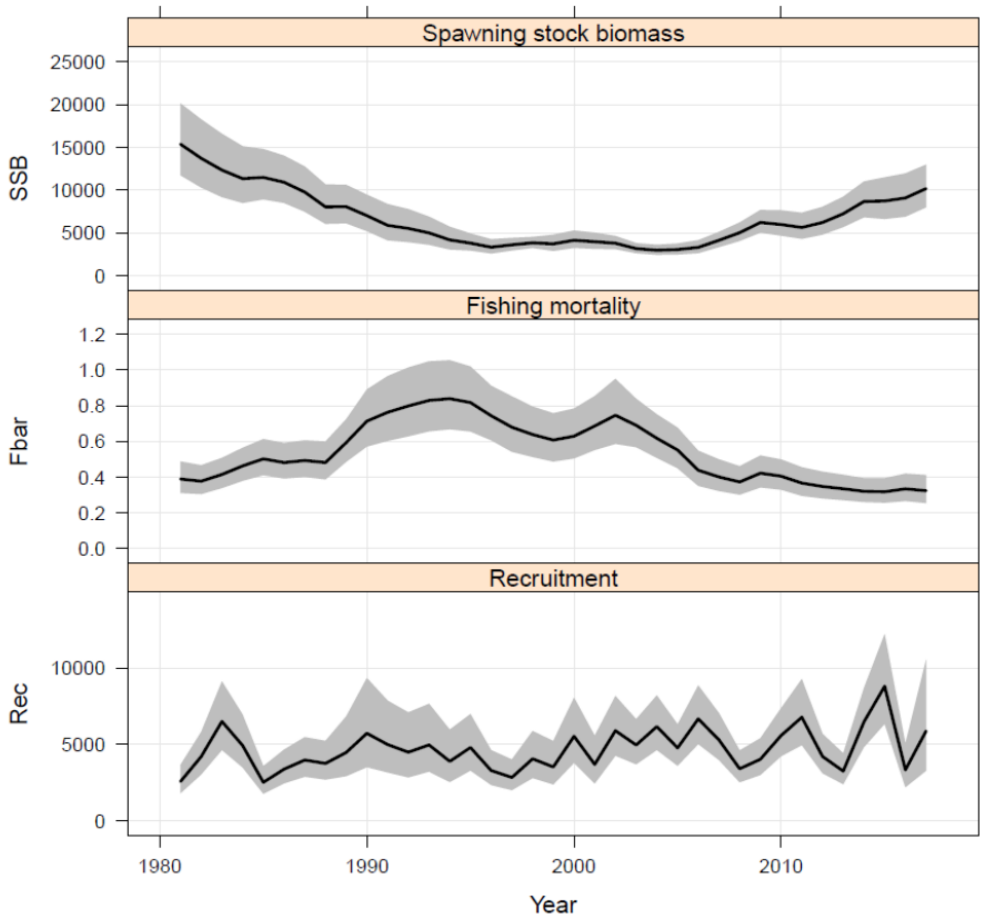

Figure C.1. Summary of the turbot in 4 assessment.

D. Reference points

Reference points were calculated using Eqsim software and ICES guidelines which was developed early 2018 by D.C.M. Miller to ensure a correct procedure in estimating reference points was followed. The script used to estimate reference points can be found at the [IBP Github page].

The simulations were executed with the entire time-series of Stock Recruitement (SR)-pairs. These includes the most recent estimate of recruitment given that the SNS is a dedicated survey on juvenile flatfish in the coastal areas and is hence expected to provide accurate estimates of recruitment (correlation ~0.83 between SNS age 1 index and estimated recruitment in the assessment). In the period 1981–1986, the productivity of the stock was markedly lower than in more recent years, but these years were included as it provided overall better fits to the stock-recruitment models. Although productivity (in recruit per spawner) has gone down in recent years, we do not assume the stock to have a lower productivity potential. The trends in R/SSB mainly show a strong negative density de-pendent effect of SSB on recruitment success.

Simulations were run with 200 iterations and applying a mixture of two SR-models, namely Segmented Regression and Ricker (sampling from 2000 fits) (Figure D.1). The fit to the Beverton-Holt SRR showed no decline towards the origin. Weight-at-age and selectivity at-age do show some trend in the past decade and hence the average over the 5 recent years were used in the simulations (excluding the most recent year), similar to the default settings. The cv on F, phi on F and cv on SSB were taken as the default values being in conformity with the WKMSYREF IV re-port (cv of F being 0.212 and phi F being 0.428, cv of SSB was set to 0).

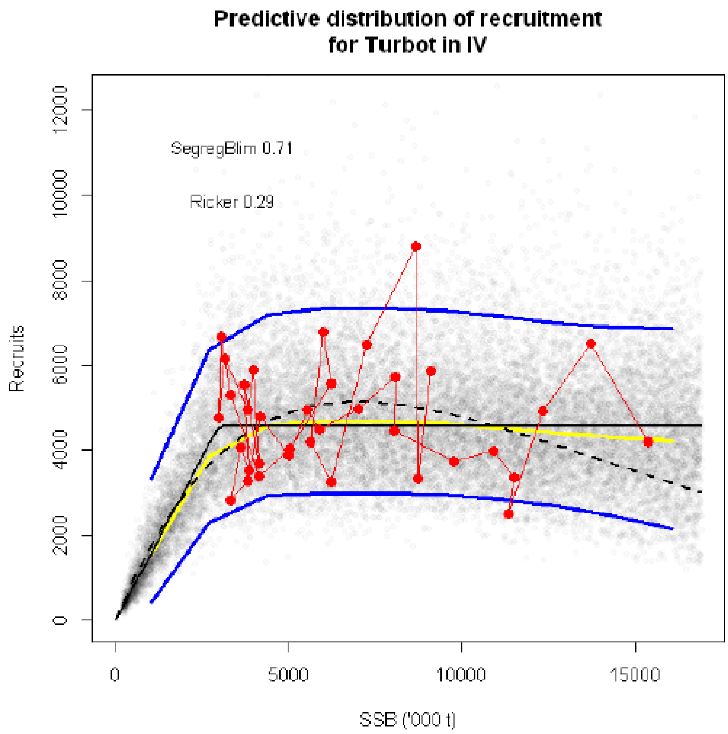

Figure D.1. Fitted combinations of stock recruitement fits to the SR-couples of turbot.

B_{lim} was set at B_{loss} since there are no indications that the stock has encountered impaired recruitment in the time-series. At very similar SSBs the stock has produced among the highest and lowest year-classes which shows that there is no distinct SSB ~R relationship. This is also true for SSBs near the lower end of its distribution, right where the breakpoint of the segmented regression is estimated. No auto-correlation in recruitment was detected. B_{pa} was derived multiplying B_{lim} with exponent of sigmaSSB * 1.645.

F_{lim} was derived from B_{lim} by simulating the stock with segmented regression SR function with the point of inflection at B_{lim}. F_{lim} = the F that, in equilibrium, gives a 50% probability of SSB > B_{lim}.

MSY $B_{trigger}$ was set to 0, Fcv, Fphi, SSBcv were set to 0 and rhoRec was set to FALSE. F_{pa} was derived multiplying F_{lim} with the exponent of -sigmaF * 1.645. Both sigmaF and sigmaSSB were set to the default values of 0.2045 (resulting into a multiplication factor of ~1.4 for B_{lim} and F_{lim} to derive B_{pa} and F_{pa}).

The initial F_{MSY} was calculated including stochasticity in the population and exploitation as well as assessment/advice error following WKLIFE IV with default values of 0.212 and 0.423 for Fcv and Fphi respectively. From this run, also $F_{MSY\ upper}$ and $F_{MSY\ lower}$ were obtained. MSY $B_{trigger}$ was set to zero while B_{lim} and B_{pa} were included. Since F_{MSY} was lower than F_{pa}, F_{MSY} was taken as the point estimate from the simulation.

MSY $B_{trigger}$ was taken as the 5th percentile of SSB at MSY which was higher than B_{pa}. Given that the stock has been fished at or below F_{MSY} since 2012, and no MSY $B_{trigger}$ value was defined before, MSY $B_{trigger}$ was set at this 5th percentile.

Finally, $F_{P.05}$ was evaluated using the MSY $B_{trigger}$ estimate from the previous analysis. This value (0.86) was higher than $F_{MSY\ upper}$ (0.48) so a modification of the F_{MSY} range was not needed.

The table below shows the estimated reference points using the final IBP 2018 assessment.

Reference point	Estimate
1. MSY $B_{trigger}$	6353
2. B_{pa}	4163
3. B_{lim}	2974
4. F_{pa}	0.43
5. F_{lim}	0.61
6. $F_{P.05}$	0.86
7. F_{MSY}	0.36
8. $F_{MSY\ lower}$	0.25
9. $F_{MSY\ upper}$	0.48

E. short term forecast

It was decided to use the FLR Flash package using fwd-routines for short term forecasts. Terminal year estimates from the SAM assessment were used as starting conditions. Since there is no clear relationship between SSB and Rec, it was decided to assume recruitment to follow a geometric mean for the entire time-series, including the latest estimate. An analyses on bias and uncertainty in recruitment prediction was undertaken to justify this decision. Figure E.1 below shows the bias and uncertainty in recruitment prediction when 3, 5, 10, 15 or 20 historic years were used in predicting recruitment. An extra scenario including the entire time-series was evaluated as well. This analyses showed that the absolute deviation when using the entire time-series was lowest, as well as the bias over the years.

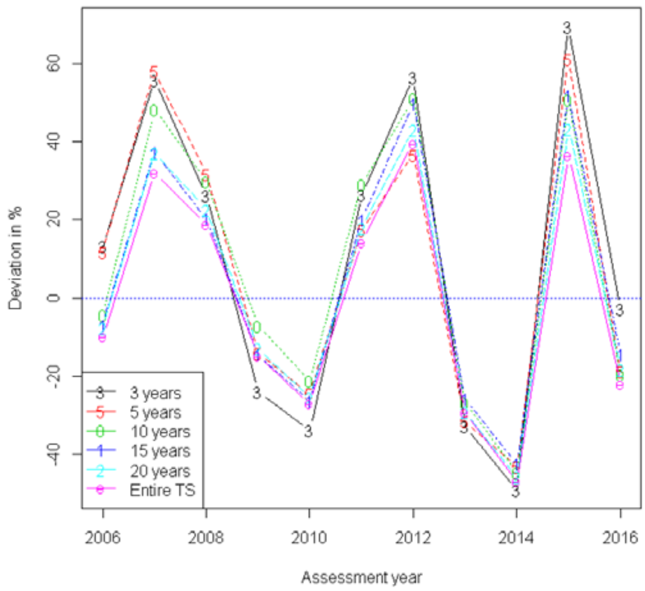

Figure E.1 Bias and deviation in recruitement prediction based on different assumptions on the number of historic years to be included in the geometric mean recruitement prediction.

Since stock and catch weight-at-age are modelled, we assume in the forecast that weights are identical to the weights used in the final assessment year. As such, we do not introduce a break in the smoothness of the weight-at-age time-series. Maturity at age and time of spawning are fixed over time, and these values are used in the forecast. Selectivity-at-age is with minimal trends in recent years, but has changed in the past decade. Hence, a 3-year average was used for future years in the simulations.

TACs for Turbot are agreed upon in combination with Brill. The proportion of Turbot landings out of this combined TAC was calculated and fluctuated around 65% without a clear trend, but with substantial variation. However, the TAC has in recent years never been exhausted and therefore using a % TAC was deemed inappropriate. Hence, the assumption for the intermediate year was made to not use a catch constraint but a status-quo F. This was also supported by the recent years in which F has been very stable around 0.34.

Table E.1. Turbot in Subarea 4. Assumptions made for the interim year and in the forecast.

Variable	Value	Notes
$F_{ages\ 2\text{–}6}$ (2018)	0.32	F status-quo from 2017 assessment result
SSB (2019)	10436	Short-term forecast (STF), in tonnes
R_{age1} (2018)	4503	Geometric mean (GM, 1981–2017), in thousands
R_{age1} (2019)	4503	Geometric mean (GM, 1981–2017), in thousands
Total catch (2018)	3490	Short-term forecast (STF), in tonnes, assuming an F status-quo

F. Other issues

After screening all available input data and work conducted during the inter-benchmark, the tur.27.4-assessment still has to be based on input data derived from limited sampling:

- Only an incomplete catch-at-age matrix is available that needs to be reconstructed by modelling for the years 1991 to 1997 and 1999 to 2001. Age in-formation for recent years is mainly available only from the Dutch fishing fleets representing more than 50% of total landings. Data from Denmark were available for the four most recent years. Danish data demonstrated a shift towards older fish compared to the Dutch data. This may also be true for other countries but no information is available. Although the inclusion of Danish data impacted the overall catch-at-age matrix only to a minor extent, this highlights the need to get catch-at-age data not only from the Netherlands but also from other countries. Also, retrospective data would be highly beneficial if available.
- The sampled mean weight-at-age matrix cannot be used directly because of too low sampling intensities and gaps in time-series. A smoothing method is applied; since this analysis is updated each time new data is added, this leads to an additional source of retrospective bias in biomass estimates.
- The available scientific surveys have a low internal consistency especially for older ages leading to a low ability to track cohorts over time. Because of this, the assessment is strongly influenced by a Dutch lpue index. This index has been standardized for changes in fishing areas and gears used (i.e. tra-ditional beam trawls vs. pulse trawls). It is also used as exploitable biomass index without age information, to avoid the catch-at-age matrix being used twice. However, a scientific survey with higher catch rates for turbot and a better internal consistency would be preferable.

Potential improvements which can be addressed in the next few years are:

- An UK lpue index was tested in the inter-benchmark but inconsistencies in the calculation procedures became obvious. In general, it would be beneficial to combine the Dutch lpue with lpue indices from other countries to derive an overall standardized lpue index. Obtaining standardised Belgian, UK and Danish LPUE data for use in the assessment model should be investigated.
- The Dutch LPUE data series receives a high weight in the assessment (higher than any other data source, and much higher than the survey indices of abundance); this weighting is, arguably, unrealistically high. The Dutch LPUE data are standardised by applying a statistical model that includes interactions in space, time and gear, and it may be possible to extract CVs associated with the estimates from this model. It is recommended that the use of such CVs in the SAM assessment be investigated to better deal with the weighting of the LPUE data series.
- There is a possibility of excluding age 1–2 from the NL lpue data. A first trial reduced the problematic retrospective pattern in F. However, currently this would mean to shorten the time-series of the lpue index considerably since disaggregated data to distinguish market categories/ages were not available before 2002 for this interbenchmark. Work on providing such data further back in time could be beneficial for the assessment.
- The Dutch LPUE data series (an aggregated biomass index) is associated with 60–70% of the total catch for turbot, but the current SAM assessment uses the selectivity estimated for the total catch to build an exploitable biomass estimate used to fit the Dutch LPUE data. This is not entirely representative and likely introduces some model misspecification. There is a fleet-based version of SAM that, given fleet-

based data could be used to deal with this problem. It is therefore recommended that the use of such fleet-based data and a fleet-based SAM version be investigated to pro-vide a more appropriate fit to the Dutch LPUE data.

- Currently, scientific surveys show relatively poor performance (due to low catch rates) in assessments of large flatfish. A new standardised survey with higher catch rates for large flatfish should be developed to improve assessments for these species.
- A delta GAM index combining different BTS surveys was tested. Currently such an index could not improve the assessment. However, age information in DATRAS was not available for the whole time-series, and errors seem to have occurred during the upload of additional data. Once the whole time-series of age information is available, a detailed analysis of delta GAM indices with various settings may be carried out.
- There is little knowledge of the natural mortality of this stock. For other flatfish species we have natural mortality estimates that are empirically derived from the cease in fishing during WWII. Using the statistical relationship as estimated by Gislason *et al.* (2010), we derived estimates for natural mortality that are higher than those for sole and plaice. The reason for these high estimates are the high K and L_∞. The benchmark group then decided to use M = 0.2 per year, as is used for many other fish in the ICES areas. Further exploration of M for turbot would improve the appropriateness of the ICES advice that will result from using the assessment.
- At present the EU provides a combined TAC for turbot and brill in the North Sea. This TAC seems largely ineffective in reducing F: increases in the stock at similar TACs lead to increased discarding. In addition, it is unclear how the quantitative single species advice for turbot and the qualitative single species advice for brill can/will be used to formulate a combined TAC for these two stocks. In this situation, improving the brill assessment may be necessary in order to ensure efficient management of both of these stocks. Ideally, a combined TAC is on that is not used.

G. References

Bedford, B.C., L.E. Woolner and Jones, B.W. 1986. Length–weight relationships for commercial fish species and conversion factors for various presentations. Data Rep., MAFF Direct. Fish. Res., Lowestoft, (10), 41 pp.

Boon, A.R., and Delbare, D. 2000. By-catch species in the North Sea flatfish fishery (data on turbot and brill) preliminary assessment DATUBRAS, study 97/078. Final RIVO report C020/00. 107 pp.

de Boor, C. 2001. A Practical Guide to Splines (Revised Edition), Springer-Verlag, New York.

Delbare, D. and R. De Clerck. 1999. Stock discrimination in relation to the assessment of the brill fishery. Study in support of the Common Fisheries Policy. Final Report EC-Study Contract DG XIV 96/001. 36pp.

ICES. 2015. Report of the Workshop to consider FMSY ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. 187 pp.

ICES 2016. Report of the Inter-Benchmark Protocol for Turbot in the North Sea (IBP Turbot), May 2015, By correspondence. ICES CM 2015/ACOM:51. 76 pp.

Nielsen, E.E., P.H. Nielsen, D. Meldrup and M.M. Hansen. 2004. Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. Molecular Ecology **13**, 585–595.

Nielsen, A., and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state–space models. Fisheries Research, 158: 96–101.

Vandamme, S. G., Maes, G. E., Raeymaekers, J. A. M. et al. 2014. Regional environmental pressure influences population differentiation in turbot (*Scophthalmus maximus*). Molecular Ecology 23(3): 618-636.

Weber, W. 1979. On the turbot stock in the North Sea. ICES C.M. 1979/G:12.