

Agreement on contaminants’ criteria and methods for the Intermediate Assessment 2017

# (OSPAR Agreement 2017-01)[[1]](#footnote-1)

# Contents

Annex 1 1

Contents 1

Assessment methodology used in OSPAR assessment 2016 3

Assessment methodology for contaminants in biota 3

Overview for status assessments 3

Overview for time series 3

Modelling changes in log concentration over time 4

Assessing environmental status and temporal trends 6

Assessment criteria: Metals in biota 7

Assessment criteria: PAHs in biota 8

Assessment criteria: CBs in biota 9

Assessment criteria: Pesticides in biota 10

Assessment criteria: Organo-metals in biota 11

Assessment criteria: Organo-bromines in biota 12

Assessment criteria: Dioxins and Furans in biota 12

Assessment criteria: Persistent organic pollutants in biota 12

Assessment methodology for biological effects 12

Assessment criteria: Biological effects 13

Assessment criteria: EROD 13

Assessment criteria: Bile metabolites 14

Assessment criteria; Scope for growth 15

Assessment methodology for imposex 15

Proportional odds model 15

Form of f(t) 16

Estimating the cut-points 19

Estimating the mean VDS class 20

Assessing environmental status and temporal trends 20

Assessment criteria: Imposex 21

Assessment methodology for contaminants in sediment 22

Overview 22

Normalisation 22

Modelling changes in log concentration over time 23

Assessing environmental status and temporal trends 25

Assessment criteria: Metals in sediment 26

Assessment criteria: PAHs in sediment 28

Assessment criteria: CBs in sediment 29

Assessment criteria: Organo-metals in sediment 29

Assessment criteria: Organo-bromines in sediment 30

Appendix 1 31

Changes to the assessment methodolgy 31

2016 Assessment Modelling of contaminants and biological effects 31

2015 Assessment 32

Determinands 33

Assessment criteria 33

Assessing status of imposex 33

Appendix 2 34

Factors for converting the basis of assessment concentrations in biota 34

Appendix 3 35

Pivot values for normalisation of metals in sediment 35

Appendix 4 36

Species area combinations used for cut-point estimation 36

# Assessment methodology used in OSPAR assessment 2016

This document is collected from the help file in the OSPAR web tool (<http://dome.ices.dk/OSPARMIME2016/main.html>), and describes the steps and statistics behind the assessment results.

The help files are updated every year if changes to the assessments are made (mainly updating the assessment criteria), so when looking at a particular dataset, the associated helpfiles are the relevant description. This document describes the status in 2016 and the dataset used for the IA2017. The statistical methods are stable and have been used for many years, it is mainly small tweaks in the conversion factors (appendix 2) and the changes are documentet from year to year (see appendix 1).

# Assessment methodology for contaminants in biota

## Overview for status assessments

For a regional status assessment, the status of each time series is summarised by the difference between the estimated mean log concentration in the final monitoring year and the log assessment concentration. This ensures that status is always measured on the same scale, even though the assessment criterion might vary between metals and time series. Essentially the same linear mixed model as for trends is then fitted:

* response: status (mean log concentration - log assessment concentration)
* fixed model: region \* metal
* random model: station + status estimation variation + residual variation

where status estimation variation is the variation in the status estimates from the individual time series analysis, assumed known and fixed.

There are no restrictions on the time series used in the status meta-analysis based on the classification of the monitoring station; time series from baseline, representative and impacted stations are all included. However, the few time series with a non-parametric assessment of status must be excluded, because there is no summary measure of status to use in the mixed model.

Again, the meta-analysis is restricted to regions and metal combinations with at least three status stations with good geographic spread.

The statistics used for status assessments will be further developed during the 2017 MIME meeting.

## Overview for time series

Time series of contaminant concentrations in biota are assessed in two stages:

The concentrations are log transformed and changes in the log concentrations over time are modelled using linear mixed models. The type of temporal change that is considered depends on the number of years of data:

*1-2 years:* no model is fitted because there are insufficient data

*3-4 years:* concentrations are assumed to be stable over time and the mean log concentration is estimated

*5-6 years:* a linear trend in log concentration is fitted

*7+ years:* more complex (smooth) patterns of change over time are modelled

The fitted models are used to assess environmental status against available assessment criteria and evidence of temporal change in contaminant levels in the last twenty years

These stages are described in more detail below. Other help files describe how the methodology is adapted when there are ‘less-than’ measurements, i.e. some concentrations are reported as below the detection limit, and missing uncertainties, i.e. the analytical variability associated with some of the concentration measurements was not reported. Changes to the methodology since the 2014 assessment are described in appendix 1.

## Modelling changes in log concentration over time

The log concentrations are modelled by a linear mixed model of the form:

response: log concentration

fixed: f(year)

random: year + sample + analytical

The fixed effects model describes how log concentrations change over time (year), where the form of f(year) depends on the number of years of data (described in the next paragraph). The random effects model has three components:

*year:* random variation in log concentration between years. Here, year is treated as a categorical variable

*sample:* random variation in log concentration between samples within years. When there is only one sample each year, this term is omitted and implicitly subsumed into the between-year variation

*analytical:* random variation inherent in the chemical measurement process. This is assumed known and derived from the the ‘uncertainties’ reported with the data. Specifically, if ui, i=1...n, are the uncertainties associated with concentrations cici (expressed as the standard deviations of the concentration measurements), then the standard deviations of the log concentration measurements log ci are taken to be ui/ci. Measurements with ui>ci (i.e. an analytical coefficient of variation of more than 100%) are omitted from the time series.

The model is fitted by maximum likelihood assuming each of the random effects are independent and normally distributed (on the log concentration scale)[[2]](#footnote-2).

The form of f(year) depends on the number of years of data:

*1-2 years*

no model is fitted as there are too few years for formal statistical analysis

*3-4 years*

mean model f(year)=μ

there are too few years for a formal trend assessment, but the mean level is summarised by μ and is used to assess status

*5-6 years*

linear model f(year)=μ+βyear

log concentrations are assumed to vary linearly with time; the fitted model is used to assess status and evidence of temporal change

*7+ years*

smooth model f(year) = s(year)

log concentrations are assumed to vary smoothly over time; the fitted model is used to assess status and evidence of temporal change.

The last case requires more explanation. When there are 7-9 years of data, both a linear model and a smoother (thin plate regression spline) on 2 degrees of freedom (df) are fitted to the data. Of these, the model chosen to make inferences about status and temporal trends is the one with the lower Akaike’s Information Criterion corrected for small sample size (AICc)[[3]](#footnote-3). When there are 10-14 years of data, a linear model and smoothers on 2 and 3 df are fitted, with the chosen model that with the lowest AICc. And when there are 15+ years of data, a linear model and smoothers on 2, 3, and 4 df are fitted, with model selection again based on AICc. Effectively, the data determine the amount of smoothing, with AICc providing an appropriate balance between model fit and model parsimony[[4]](#footnote-4).

## Assessing environmental status and temporal trends

Environmental status and temporal trends are assessed using the model fitted to the concentration data.

Environmental status is assessed by:

calculating the upper one-sided 95% confidence limit on the fitted mean log concentration in the most recent monitoring year[[5]](#footnote-5)

back-transforming this to the concentration scale

comparing the back-transformed upper confidence limit to the available assessment criteria

For example, if the back-transformed upper confidence limit is below the Background Assessment Concentration (BAC), then the median concentration in the most recent monitoring year is significantly below the BAC and concentrations are said to be ‘at background’. For an example, see Fryer & Nicholson (1999).

No formal assessment of status is made when there are only 1 or 2 years of data. However, an ad-hoc assessment is made by:

calculating the median of the log concentration measurements in each year back-transforming these to the concentration scale comparing the back-transformed median log concentration (1 year) or the larger of the two back-transformed median log concentrations (2 years) to the assessment criteria.

Temporal trends are assessed for all time series with at least five years of data. When a linear model has been fitted (i.e. when there are 5-6 years of data, or if there are 7+ years of data and no evidence of nonlinearity), the statistical significance of the temporal trend is obtained from a likelihood ratio test[[6]](#footnote-6) that compares the fits of the linear model f(year)=μ+βyear and the mean model f(year)=μ. The summary maps show a downward or upward trend if the trend is significant at the 5% significance level.

When a smooth model has been fitted, a plot of the fitted model is needed to understand the overall pattern of change. (This is available on the Raw data with assessment and Assessment pages on the right side of the summary map under Graphics.) The summary map focusses on just one aspect of the change over time: the change in concentration in the most recent twenty monitoring years; i.e. between 1995 and 2014 (the assessment only includes data up to 2014). For this, the fitted value of the smoother in 2014 is compared to the fitted value in 1995 using a t-test, with significance assessed at the 5% level. The correlation between the two fitted values is accounted for by the t-test. If the time series does not extend to 2014, then the fitted value in the last monitoring year is used instead. Similarly, if the time series starts after 1995, the fitted value in the first monitoring year is used.

*Reference:*

Fryer RJ & Nicholson MD, 1999. Using smoothers for comprehensive assessments of contaminant time series in marine biota. ICES Journal of Marine Science 56: 779-790.

## Assessment criteria: Metals in biota

Two assessment criteria are used to assess metal concentrations in biota: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**uropean **C**ommission food standard (EC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels.  Mean concentrations significantly below the BAC are said to be near background.
[ECs](http://dome.ices.dk/OSPARMIME2016/help_com_reg_2006_1881_en_consolidated_20121203.pdf) have been used in the absence of any satisfactory criteria for assessing the ecological significance of biota concentrations.  ECs are the maximum acceptable concentrations in food for the protection of public health.
BACs and ECs are available for the following metals:

|  | **BAC** | **EC** |
| --- | --- | --- |
|  | **Mussels** | **Oysters** | **Fish** | **All species** |
| Cadmium |   960 | 3000 | 26 | 1000 |
| Copper | 6000 | 6000 |  |  |
| Mercury |     90 |   180 | 35 |   500 |
| Lead | 1300 | 1300 | 26 | 1500 |
| Zinc | 63000   | 63000   |  |  |

Notes:
BACs for mussels and oysters are expressed as μg kg-1 dw and BACs for fish and ECs are expressed as μg kg-1 ww

* cadmium and lead are monitored in fish liver, for which no food standard exists; concentrations in fish liver are naturally higher than in fish muscle, so the food standards for fish muscle are not used; instead the food standards for shellfish are used as a proxy
* BACs and ECs are converted to other bases (wet, dry or lipid weight) using [species-specific conversion factors](http://dome.ices.dk/OSPARMIME2016/help_basis_conversion.html) (see appendix 2)

## Assessment criteria: PAHs in biota

Two assessment criteria are used to assess PAH concentrations in biota: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels. Mean concentrations significantly below the BAC are said to be near background.
EACs were developed by OSPAR and the [International Council for the Exploration of the Sea](http://www.ices.dk/) for assessing the ecological significance of biota concentrations. Concentrations below the EAC should not cause any chronic effects in marine organisms.

BACs and EACs are available for the following PAHs in mussels and oysters:

|  | **BAC** | **EAC** |
| --- | --- | --- |
| Naphthalene |  |  340 |
| Phenanthrene | 11.0  | 1700  |
| Anthracene |  |  290 |
| Fluoranthene | 12.2  |  110 |
| Pyrene |  9.0 |  100 |
| Benz[a]anthracene |  2.5 |    80 |
| Chrysene (Triphenylene) |  8.1 |  |
| Benzo[a]pyrene |  1.4 |  600 |
| Benzo[ghi]perylene |  2.5 |  110 |
| Indeno[123-cd]pyrene |  2.4 |  |

Notes:
all concentrations are expressed as μg kg-1 dw

* BACs and EACs are converted to other bases (wet, dry or lipid weight) using [species-specific conversion factors](http://dome.ices.dk/OSPARMIME2016/help_basis_conversion.html) (appendix 2)
* PAHs are not routinely monitored in fish, so no BACs and EACs for fish have been derived

## Assessment criteria: CBs in biota

Two assessment criteria are used to assess CB concentrations in biota: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels. Mean concentrations significantly below the BAC are said to be near background. EACs were developed by OSPAR and the [International Council for the Exploration of the Sea](http://www.ices.dk/) for assessing the ecological significance of biota concentrations. Concentrations below the EAC should not cause any chronic effects in marine organisms.

BACs and EACs are available for the following CBs:

|  | **BAC** | **EAC** |
| --- | --- | --- |
|  | mussels and oysters | fish | all biota |
| CB28 | 0.75 | 0.10 |   67 |
| CB52 | 0.75 | 0.08 | 108 |
| CB101 | 0.70 | 0.08 | 121 |
| CB105 | 0.75 | 0.08 |  |
| CB118 | 0.60 | 0.10 |   25 |
| CB138 | 0.60 | 0.09 | 317 |
| CB153 | 0.60 | 0.10 | 1585   |
| CB156 | 0.60 | 0.08 |  |
| CB180 | 0.60 | 0.11 | 469 |

Notes:
BACs are expressed as μg kg-1 dw for mussels and oysters and as μg kg-1 ww for fish

* EACs are expressed as μg kg-1 lw
* BACs and EACs are converted to other bases (wet, dry or lipid weight) using [species-specific conversion factors](http://dome.ices.dk/OSPARMIME2016/help_basis_conversion.html) (appendix 2)
* the EACs are based on partitioning theory and are sometimes known as EACpassive

## Assessment criteria: Pesticides in biota

Two assessment criteria are used to assess pesticide concentrations in biota: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels. Mean concentrations significantly below the BAC are said to be near background.

EACs were developed by OSPAR and the [International Council for the Exploration of the Sea](http://www.ices.dk/) for assessing the ecological significance of biota concentrations. Concentrations below the EAC should not cause any chronic effects in marine organisms.

BACs and EACs are available for the following pesticides:

|  |  |  |
| --- | --- | --- |
|  | **Mussels and oysters** | **Fish** |
|  | BAC | EAC | BAC | EAC |
| DDE (p,p') | 0.63 |  | 0.10 |  |
| Hexachlorobenzene | 0.63 |   | 0.09 |  |
| α-HCH | 0.64 |   |  |  |
| γ-HCH | 0.97 | 1.45 |  | 11 |

Notes:

* BACs and EACs are expressed as μg kg-1 dw for mussels and oysters and as μg kg-1 ww for fish
* BACs and EACs are converted to other bases (wet, dry or lipid weight) using [species-specific conversion factors](http://dome.ices.dk/OSPARMIME2016/help_basis_conversion.html) (appendix 2)
* γ-HCH is monitored in fish liver and the EAC is obtained by multiplying the EAC for whole fish by 10

## Assessment criteria: Organo-metals in biota

Two assessment criteria are used to assess organo-metal concentrations in biota: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels. Mean concentrations significantly below the BAC are said to be near background.

EACs were developed by OSPAR and the [International Council for the Exploration of the Sea](http://www.ices.dk/) for assessing the ecological significance of biota concentrations. Concentrations below the EAC should not cause any chronic effects in marine organisms.

BACs and EACs are available for the following organo-metals:

|  |  |  |
| --- | --- | --- |
|  | **Mussels and oysters** | **Fish** |
|  | BAC | EAC | BAC | EAC |
| Tributyltin | 5.0 | 12.0 |  |  |

Notes:

* BACs and EACs are expressed as μg kg-1 dw for mussels and oysters
* BACs and EACs are converted to other bases (wet, dry or lipid weight) using [species-specific conversion factors](http://dome.ices.dk/OSPARMIME2016/help_basis_conversion.html) (appendix 2)

## Assessment criteria: Organo-bromines in biota

Assessment criteria for organo-bromines in biota are under development.

## Assessment criteria: Dioxins and Furans in biota

Assessment criteria for dioxins and Furans in biota are under development.

## Assessment criteria: Persistent organic pollutants in biota

Assessment criteria for persistent organic pollutants in biota are under development.

# Assessment methodology for biological effects

The assessment methodology for biological effects measurements is essentially the same as that for [chemical concentrations in biota](http://dome.ices.dk/OSPARMIME2016/help_methods_biota.html). However, some modifications are required for glutathionine transferase, acetylcholine esterase activity, aminolevulinic acid dehydratase and scope for growth.

Low values of these variables indicate unhealthy organisms, so status is assessed using the *lower* one-sided 95% confidence limit on the fitted mean value in the most recent monitoring year. For example, if the lower confidence limit is above the Background Assessment Concentration (BAC), then the mean value in the most recent monitoring year is significantly above the BAC and levels are said to be ‘at background’.

Scope for growth

The measurements are not log transformed because scope for growth can be negative and because the data are approximately normally distributed on the untransformed scale. Consequently, all models are of temporal changes on the original scale. Further, low scope for growth indicates unhealthy organisms, so status is assessed using the *lower* one-sided 95% confidence limit on the fitted mean scope for growth in the most recent monitoring year. As the data have not been transformed before modelling, the lower confidence limit is compared directly to the assessment criteria; i.e. there is no need for any back-transformation.

## Assessment criteria: Biological effects

Two assessment criteria are used to assess biological effects: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

The assessment criteria were developed within the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) framework with scientific advice from the [International Council for the Exploration of the Sea](http://www.ices.dk/). Mean values significantly below the BAC are said to be near background. Values below the EAC indicate no chronic effects on the organisms concerned. Full details can be found in [Davies & Vethaak (2012)](http://www.ices.dk/sites/pub/Publication%20Reports/Cooperative%20Research%20Report%20%28CRR%29/crr315/CRR315_Integrated%20Monitoring_final.pdf) or [OSPAR (2013)](http://dome.ices.dk/OSPARMIME2016/help_bioeffects_background.pdf).

BACs and EACs are available for [EROD](http://dome.ices.dk/OSPARMIME2016/help_ac_biota_erod.html), [bile metabolites](http://dome.ices.dk/OSPARMIME2016/help_ac_biota_bile_metabolites.html) and [scope for growth](http://dome.ices.dk/OSPARMIME2016/help_ac_biota_scope_for_growth.html)

*References*Davies, I. M. and Vethaak, A. D. 2012. Integrated marine environmental monitoring of chemicals and their effects. ICES Cooperative Research Report No. 315. 277 pp.

OSPAR, 2013. Background documents and technical annexes for biological effects monitoring (Update 2013). OSPAR Commission, London. Publication 589, 238 pp.

## Assessment criteria: EROD

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Species** | **Latin name** | **Sex** | **Matrix** | **BAC** |
| Cod | *Gadus morhua* | Both | Liver microsome | 145 |
| Dab | *Limanda limanda* | Female | Liver S9 | 178 |
|  |  | Male | Liver S9 | 147 |
|  |  | Both | Liver microsome | 680 |
| Dragonet | *Callionymus lyra* | Both | Liver microsome | 202 |
| Flounder | *Platichthys flesus* | Male | Liver S9 | 24 |
| Four spotted megrim | *Lepidorhombus boscii* | Both | Liver microsome | 13 |
| Plaice | *Pleuronectes platessa* | Male | Liver S9 | 9.5 |
|  |  | Both | Liver microsome | 255 |
| Red mullet | *Mullus barbatus* | Male | Liver S9 | 208 |

Notes:

* BACs are expressed as pmol min-1 mg S9 protein-1 or pmol min-1 mg microsomal protein-1 for the liver S9 and liver microsome matrices respectively
* there are no EACs for EROD

## Assessment criteria: Bile metabolites

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Bile metabolite** | **Species** | **Latin name** | **Method** | **BAC** | **EAC** |
| 1-OH pyrene | Cod | *Gadus morhua* | HPLC-F | 21     |  |
|  |  |  | GC-MS |  | 483 |
|  | Dab | *Limanda limanda* | HPLC-F | 16     |  |
|  | Flounder | *Platichthys flesus* | HPLC-F | 16     |  |
|  | Haddock | *Melanogrammus aeglefinus* | HPLC-F | 13     |  |
| 1-OH pyrene equivalents | Cod | *Gadus morhua* | SSF | 1.1 |   35 |
|  | Dab | *Limanda limanda* | SSF |   0.15 |   22 |
|  | Flounder | *Platichthys flesus* | SSF | 1.3 |   29 |
|  | Haddock | *Melanogrammus aeglefinus* | SSF | 1.9 |   35 |
| 1-OH phenanthrene | Cod | *Gadus morhua* | HPLC-F | 2.7 |  |
|  |  |  | GC-MS |  | 528 |
|  | Dab | *Limanda limanda* | HPLC-F | 3.7 |  |
|  | Flounder | *Platichthys flesus* | HPLC-F | 3.7 |  |
|  | Haddock | *Melanogrammus aeglefinus* | HPLC-F | 0.8 |  |

Notes:

* HPLC-F is high performance liquid chromatography - fluorescence, GC-MS is gas chromatography - mass spectrometry, and SSF is synchronous scan fluorescence 341/383 nm
* BACs and EACs are expressed as ng ml-1 for HPLC-F, ng g-1 for GC-MS, and pyrene-type μg ml-1 for SSF
* the proliferation of methods and units are the origins of which are unclear

## Assessment criteria; Scope for growth

|  |  |  |
| --- | --- | --- |
| **Species** | **BAC** | **EAC** |
| Mussels | 25 | 15 |

Notes:

* BACs and EACs are expressed as J h-1 g-1
* High values of scope for growth indicate healthy mussels

# Assessment methodology for imposex

Overview

Ideally, imposex data are submitted as individual measurements; for example, as the Vas Deferens Sequence (VDS) of each female snail. This provides information about variation between individuals, and allows efficient statistical models to be fitted to assess trends and status. However, sometimes the data are submitted as an annual index; for example, as the Vas Deferens Sequence Index (VDSI), the arithmetic mean of the individual VDS measurements. A more ad-hoc modelling approach is then all that is possible. This help file describes the methodology for assessing time series of individual VDS measurements. Other help files describe the approach when VDS data are submitted as [annual indices](http://dome.ices.dk/OSPARMIME2016/help_methods_imposex_indices.html), or as a [mixture of individual measurements and annual indices](http://dome.ices.dk/OSPARMIME2016/help_methods_imposex_mixture.html).

For some species, imposex stage or intersex stage are reported rather than VDS, again either as individuals or annual indices[[7]](#footnote-7). However, for these measures, there is insufficient variation in stage between individuals to model the individual measurements and instead the [annual indices](http://dome.ices.dk/OSPARMIME2016/help_methods_imposex_annual) are assessed.

Changes to the methodology since the 2014 assessment can be found in appendix 1.

## Proportional odds model

The individual VDS measurements are modelled with a proportional odds model (McCullagh & Nelder, 1989). Let yijbe the VDS measurement of the jjth female snail in year ti,i=1...N, with yij∈{0,...,K} where K is the highest possible VDS class[[8]](#footnote-8). It is assumed that

logit(Prob(yij≤k))=f(ti)+θk

for 0≤k≤K−1, with Prob(yij≤K)=1. Here,  f(t) is a function that describes how imposex levels change over time (year). Various forms of f(t) are considered and these are discussed in the next section. The θk are cut points that measure the odds of being in a particular VDS class or below. Since the classes are ordered, the cut points are subject to the constraints:

θ0<θ1<...<θ K−1[[9]](#footnote-9).

The model is fitted by maximum likelihood. However, there are rarely sufficient data in a single time series to estimate the cut points precisely, so the cut points are first estimated from a saturated model fitted to multiple time series (described later) and are then assumed fixed and known[[10]](#footnote-10). The only parameters estimated when fitting the proportional odds model to a single time series are thus the parameters of f(t). Parameter standard errors are estimated from the Hessian matrix[[11]](#footnote-11).

McCullagh P & Nelder JA, 1989. Generalised Linear Models (second edition). Chapman & Hall, London.

## Form of f(t)

Several different candidate forms of f(t) are considered, depending on the length of the time series. As well as linear logistic and smooth trends, change-point models are considered. These are motivated by the patterns seen in many time series, where there are steep declines in VDS levels starting in the mid 2000s. These changes coincide with the introduction of EC Regulation 782/2003, which implemented the provisions of the International Maritime Organisation’s Antifouling Systems Convention (IMO, 2001) prohibiting application of TBT surface coatings to all vessels by 2003, and the global ban on TBT which came into force in September 2008. The steep declines usually cannot be described adequately by linear logisitic models, or even smoothers. However, change-point models provide a way of capturing the steep decline with relatively few parameters. The years 2004, 2005, 2006, 2007 and 2008 are regarded as potential change-years, since the environmental response to the TBT measures is likely to have started in this period.

Intuitively, the complexity of the candidate forms of f(t) should be based on the number of years of data, NN. For example, with 8 years of data, one might consider a linear model f(t)=μ+βt and a smoother f(t)=s(t) on 2 degrees of freedom (df) (analogous to the models fitted to contaminant time series). However, this runs into numerical difficulties when a time series starts with a series of years in which all VDS measurements equal the maximum value KK, or ends with a series of years in which all VDS measurements equal 0, as the amount of information in the data for estimating f(t) is then reduced[[12]](#footnote-12). Instead, the candidate forms of f(t) are based on NmidNmid, an approximate measure of the number of years of data that contain information about changes in VDS levels. Loosely, NmidNmid is the number of years in the ‘middle’ of the timeseries, where intermediate VDS levels are observed. Formally, NmidNmid is defined as follows. Let Ii = 1 if all the VDS measurements in year ti< equal K, -1 if all the VDS measurements in year ti equal 0, and 0 otherwise. Let

i1=1,N,min{i:Ii+1<1},if I1<1if Ii=1 ∀ I otherwise

i1={1,if I1<1N,if Ii=1∀imin{i:Ii+1<1}, otherwise

Similarly, let

Then {ti,i=i1,...,i2} are the ‘middle’ years of the time series and Nmid=i2−i1+1.

The linear and smooth candidate forms of f(t) are then

N≤2

no model is fitted

N≥3 and Nmid=1

mean model f(t)=μ

The VDS measurements in the entire time series either all equal K or all equal 0, so there is no trend.

N≥3 and Nmid=2,3 or 4

linear model f(t)=μ+βt

Nmid≥5

linear model f(t)=μ+βt and smooth model f(t)=s(t)

Smoothers on 2 degrees of freedom (df) are considered when 5≤Nmid≤7, on 2 and 3 df when 8≤Nmid≤10 and on 2, 3, and 4 df when Nmid≥1.

Change-point models are also considered provided that the time series starts before 2008 and that N≥3N≥3 and Nmid>1. Each change point model is of the form

f(t)={μ,μ+g(t), if t<tchange if t≥tchange

where tchange is the change year and g(tchange)=0 to ensure  f(t) is continuous. Let N∗mid be the number of ‘middle’ years from tchangetchange onwards; i.e. |{ti:ti≥tchange and i≥=i1}||{ti:ti≥tchange and i≥=i1}|. Then, similar to above, the form of g(t)g(t) depends on N∗mid.

N∗mid=2,3 or 4

linear change-point model g(t)=β(t− tchange )

N∗mid ≥5

linear change-point model g(t)=β(t−tchange) and smooth change-point model g(t)=s(t), with s(tchange) = 0

Smoothers on 2 degrees of freedom (df) are considered when 5≤N∗mid ≤7, on 2 and 3 df when 8≤N∗mid ≤10 and on 2, 3, and 4 df when N∗mid ≥11.

The change-point models are fitted for each change-year tchangetchange = 2004, 2005, 2006, 2007, 2008 in turn, provided that t1<tchange; i.e. the time series started before the change-year.

All the candidate models are fitted by maximum likelihood, with the final model chosen using AICc[[13]](#footnote-13). For some time series, there are many candidate models and there is a danger of over-fitting the data. However, this is mitigated somewhat by the fact that the models have been tailored to patterns of change seen in so many time series. It is also preferable to overfit rather than underfit for the purposes of assessing environmental status. Linear or smooth models often overpredict VDS levels in the final monitoring year, so if these are the only models considered, status will appear to be poorer than it should be.

## Estimating the cut-points

The cut-points θk,k=0,...,K−1 determine the probability of being in each VDS class given the underlying level of TBT contamination (represented by f(t) above). The cut-points can be thought of as measuring the eco-toxicological response of a species to TBT contamination and might reasonably be assumed to be constant over a wide area. The cut-points can therefore be estimated with good precision by fitting a ‘full’ model to the data from many time series collected over a wide area.

Suppose that, for a particular species and area, there are VDS time series at MM stations. With some abuse of notation, let ymij be the VDS measurement of the jth female snail in year tmi from station mm. Then the full model is

logit(Prob(ymij≤k))=μmi+θk

where μmirepresents the underlying level of TBT contamination in year tmitmi at station mm. The species area combinations used for estimating the cut-points are found in appendix 4.

## Estimating the mean VDS class

The mean VDS class in year tt, denoted v(t), is

where Yt is a random variable describing the VDS class of individual snails in year t. The probabilities are expressed in terms of f(t) and the cut-points through the relationships

 Prob(Yt=k)= Prob(Yt≤k)− Prob(Yt≤k−1)

=[exp(f(t)+θk)/(1+exp(f(t)+θk))]−[exp(f(t)+θk−1)/(1+exp(f(t)+θk−1))]

with Prob(Yt≤K)= 1 as before. The mean VDS class v(t) is then estimated by plugging the estimates of f(t) and the cut-points into these formulae.

Approximate confidence intervals on v(t) are obtained by simulating the distribution of the estimates of f(t) and the cut-points and hence the distribution of v(t)^[[14]](#footnote-14). In particular, an upper one-sided 95% confidence limit on v(t) is the 95% ordered value of the simulated distribution of v(t)^.

## Assessing environmental status and temporal trends

Environmental status and temporal trends are assessed using the model fitted to the VDS data

Environmental status is assessed by comparing the upper one-sided 95% confidence limit on the mean VDS class in the most recent monitoring year (see previous section) to the available assessment criteria. For example, if the upper confidence limit is below the Background Assessment Concentration (BAC), then the mean VDS class in the most recent monitoring year is significantly below the BAC and VDS levels are said to be ‘at background’.

No formal assessment of status is made when there are only 1 or 2 years of data. However, an ad-hoc assessment is made by computing an upper one-sided 95% confidence limit on the mean VDS class in the final monitoring year from the full model used to estimate the cut-points. This confidence limit is then compared to the assessment criteria.

Temporal trends are assessed for all time series with at least three years of data. When a linear or a linear change-point model has been fitted, the statistical significance of the temporal trend is obtained from an F test[[15]](#footnote-15) that compares the fits of the linear (change-point) model and the mean model f(year)=μf(year)=μ. The summary maps show a downward or upward trend if the trend is significant at the 5% significance level.

When a smooth or a smooth change-point model has been fitted, a plot of the fitted model is needed to understand the overall pattern of change. (This is available on the Raw data with assessment and Assessment pages on the right side of the summary map under Graphics.) The summary map focusses on just one aspect of the change over time: the change in f(t)f(t) in the most recent twenty monitoring years; i.e. between 1995 and 2014 (the assessment only includes data up to 2014). For this, the fitted value of the smoother in 2014 is compared to the fitted value in 1995 using a t-test, with significance assessed at the 5% level[[16]](#footnote-16). The correlation between the two fitted values is accounted for by the t-test. If the time series does not extend to 2014, then the fitted value in the last monitoring year is used instead. Similarly, if the time series starts after 1995, the fitted value in the first monitoring year is used.

## Assessment criteria: Imposex

Two assessment criteria are used to assess imposex in snails: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

The assessment criteria were developed within the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) framework with scientific advice from the [International Council for the Exploration of the Sea](http://www.ices.dk/). Mean values significantly below the BAC are said to be near background. Values below the EAC indicate no chronic effects of Tributyltin on snails. Full details can be found in [OSPAR (2013)](http://dome.ices.dk/OSPARMIME2016/help_bioeffects_background.pdf).

BACs and EACs are available for the following species and imposex measures:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Measure** | **Species** | **Latin name** | **BAC** | **EAC** |
| VDS | Dog whelk | *Nucella lapillus* | 0.3 | 2.0 |
| VDS | Red whelk | *Neptunea antiqua* | 0.3 | 2.0 |
| VDS | Netted dog whelk | *Nassarius reticulatus* |  | 0.3 |
| IMPS | Common whelk | *Buccinum undatum* |  | 0.3 |

*References*OSPAR, 2013. Background documents and technical annexes for biological effects monitoring (Update 2013). OSPAR Commission, London. Publication 589, 238 pp.

# Assessment methodology for contaminants in sediment

## Overview

Time series of contaminant concentrations in sediment are assessed in three stages:

1. The [concentrations are normalised](http://dome.ices.dk/OSPARMIME2016/help_methods_sediment_metals.html#Normalisation) to account for changes in the bulk physical composition of the sediment such as particle size distribution or organic carbon content. (This is not done for the Iberian Sea or the Gulf of Cadiz)
2. The (normalised) concentrations are log transformed and [changes in the log concentrations over time are modelled](http://dome.ices.dk/OSPARMIME2016/help_methods_sediment_metals.html#Mixed_model) using linear mixed models. The type of temporal change that is considered depends on the number of years of data:
	* 1-2 years: no model is fitted because there are insufficient data
	* 3-4 years: concentrations are assumed to be stable over time and the mean log concentration is estimated
	* 5-6 years: a linear trend in log concentration is fitted
	* 7+ years: more complex (smooth) patterns of change over time are modelled
3. The fitted models are used to [assess](http://dome.ices.dk/OSPARMIME2016/help_methods_sediment_metals.html#Assessment) environmental status against available assessment criteria and evidence of temporal change in contaminant levels in the last twenty years

These stages are described in more detail below. Other help files describe how the methodology is adapted when there are [‘less-than’ measurements](http://dome.ices.dk/OSPARMIME2016/help_methods_less_thans.html), i.e. some concentrations are reported as below the detection limit, and [missing uncertainties](http://dome.ices.dk/OSPARMIME2016/help_methods_missing_uncertainties.html), i.e. the analytical variability associated with some of the concentration measurements was not reported. Changes to the methodology since the 2014 assessment can be found in appendix 2.

## Normalisation

In most sub-regions, the concentrations are first normalised to account for changes in the bulk physical composition of the sediment such as particle size distribution or organic carbon content. (Concentrations from the Iberian Sea and the Gulf of Cadiz are not normalised.) Normalisation requires pivot values, estimates of the concentrations of contaminants and normalisers in pure sand. A normalised concentration is given by:

where

* css is the normalised concentration of the contaminant
* cm is the measured concentration of the contaminant
* cx is the pivot concentration for the contaminant
* nss is the reference concentration of the normaliser
* nm is the measured concentration of the normaliser
* nx is the pivot concentration for the normaliser

The analytical standard deviation uu of the normalised concentration is estimated from:

where uc and un are the analytical standard deviations of the contaminant and normalised concentration measurements respectively. These are submitted with the data where they are known as ‘uncertainties’.

Metal concentrations are normalised to a standard sediment with 5% aluminium. The pivot values cx and nx and reference concentration nss depend on the digestion method used in the chemical extraction and can be found in appendix 3. Organic concentrations are normalised to a standard sediment with 2.5% organic carbon content and, regardless of the digestion method, nssnss = 2.5. For organics, the contaminant and normaliser pivot values are both 0, so the formulae above simplify to:

and

## Modelling changes in log concentration over time

The log (normalised) concentrations are modelled by a linear mixed model of the form:

* response: log concentration
* fixed: f(year)
* random: year + sample + analytical

The fixed effects model describes how log concentrations change over time (year), where the form of f(year) depends on the number of years of data (described in the next paragraph). The random effects model has three components:

* year: random variation in log concentration between years. Here, year is treated as a categorical variable
* sample: random variation in log concentration between samples within years. When there is only one sample each year, this term is omitted and implicitly subsumed into the between-year variation
* analytical: random variation inherent in the chemical measurement process. This is assumed known and derived from the the ‘uncertainties’ reported with the data. Specifically, if ui, i=1...n, are the uncertainties associated with concentrations cici (expressed as the standard deviations of the concentration measurements), then the standard deviations of the log concentration measurements  log ci are taken to be ui/ci. Measurements with ui>ci (i.e. an analytical coefficient of variation of more than 100%) are omitted from the time series. When the concentrations are normalised, then the uncertainties are the analytical standard deviations of the normalised concentrations calculated in the previous section.

The model is fitted by maximum likelihood assuming each of the random effects are independent and normally distributed (on the log concentration scale)[[17]](#footnote-17).

The form of f(year) depends on the number of years of data:

*1-2 years*

no model is fitted as there are too few years for formal statistical analysis

*3-4 years*

mean model f(year)=μ

there are too few years for a formal trend assessment, but the mean level is summarised by μ and is used to assess status

*5-6 years*

linear model f(year)=μ+βyear

log concentrations are assumed to vary linearly with time; the fitted model is used to assess status and evidence of temporal change

*7+ years*

smooth model f(year) = s(year)

log concentrations are assumed to vary smoothly over time; the fitted model is used to assess status and evidence of temporal change

The last case requires more explanation. When there are 7-9 years of data, both a linear model and a smoother (thin plate regression spline) on 2 degrees of freedom (df) are fitted to the data. Of these, the model chosen to make inferences about status and temporal trends is the one with the lower Akaike’s Information Criterion corrected for small sample size (AICc)[[18]](#footnote-18). When there are 10-14 years of data, a linear model and smoothers on 2 and 3 df are fitted, with the chosen model that with the lowest AICc. And when there are 15+ years of data, a linear model and smoothers on 2, 3, and 4 df are fitted, with model selection again based on AICc. Effectively, the data determine the amount of smoothing, with AICc providing an appropriate balance between model fit and model parsimony[[19]](#footnote-19).

## Assessing environmental status and temporal trends

Environmental status and temporal trends are assessed using the [model fitted](http://dome.ices.dk/OSPARMIME2016/help_methods_sediment_metals.html#Mixed_model) to the concentration data.

Environmental status is assessed by:

* calculating the upper one-sided 95% confidence limit on the fitted mean log concentration in the most recent monitoring year[[20]](#footnote-20)
* back-transforming this to the concentration scale
* comparing the back-transformed upper confidence limit to the available assessment criteria

For example, if the back-transformed upper confidence limit is below the Background Assessment Concentration (BAC), then the median concentration in the most recent monitoring year is significantly below the BAC and concentrations are said to be ‘at background’. For an example, see Fryer & Nicholson (1999).

No formal assessment of status is made when there are only 1 or 2 years of data. However, an ad-hoc assessment is made by:

* calculating the median of the log concentration measurements in each year
* back-transforming these to the concentration scale
* comparing the back-transformed median log concentration (1 year) or the larger of the two back-transformed median log concentrations (2 years) to the assessment criteria.

Temporal trends are assessed for all time series with at least five years of data. When a linear model has been fitted (i.e. when there are 5-6 years of data, or if there are 7+ years of data and no evidence of nonlinearity), the statistical significance of the temporal trend is obtained from a likelihood ratio test[[21]](#footnote-21) that compares the fits of the linear model f(year)=μ+βyear and the mean model f(year)=μ. The summary maps show a downward or upward trend if the trend is significant at the 5% significance level.

When a smooth model has been fitted, a plot of the fitted model is needed to understand the overall pattern of change. (This is available on the Raw data with assessment and Assessment pages on the right side of the summary map under Graphics.) The summary map focusses on just one aspect of the change over time: the change in concentration in the most recent twenty monitoring years; i.e. between 1995 and 2014 (the assessment only includes data up to 2014). For this, the fitted value of the smoother in 2014 is compared to the fitted value in 1995 using a t-test, with significance assessed at the 5% level. The correlation between the two fitted values is accounted for by the t-test. If the time series does not extend to 2014, then the fitted value in the last monitoring year is used instead. Similarly, if the time series starts after 1995, the fitted value in the first monitoring year is used.

Fryer RJ & Nicholson MD, 1999. Using smoothers for comprehensive assessments of contaminant time series in marine biota. ICES Journal of Marine Science 56: 779-790.

## Assessment criteria: Metals in sediment

Two assessment criteria are used to assess metal concentrations in sediment: the

Background Assessment Concentration (BAC)

Effects Range Low (ERL)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels.  Mean concentrations significantly below the BAC are said to be near background.

[ERLs](http://www.nj.gov/dep/srp/guidance/ecoscreening/esc_table.pdf) were developed by the [United States Environmental Protection Agency](http://www.epa.gov/) for assessing the ecological significance of sediment concentrations.  Concentrations below the ERL rarely cause adverse effects in marine organisms.

BACs and / or ERLs are available for the following metals:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BAC** | **BAC** | **ERL** |
|  | All subregions exceptIberian Sea and Gulf of Cadiz | Iberian Seaand Gulf of Cadiz | All subregions |
| Arsenic | 25    |  |   8.2 |
| Cadmium |     0.31 |     0.129 |   1.2 |
| Chromium | 81    |  | 81  |
| Copper | 27    |  | 34  |
| Mercury |     0.07 |     0.091 |     0.15 |
| Nickel | 36    |  | 21  |
| Lead | 38    | 22.4    | 47  |
| Zinc | 122      |  | 150  |

Notes:

* all concentrations are expressed as mg kg-1 dw
* BACs are normalised to 5% aluminium in all subregions except the Iberian Sea and Gulf of Cadiz, where BACs are not normalised
* for arsenic and nickel, the ERLs are below the OSPAR Background Concentrations of 15 and 30 mg kg-1 respectively; concentrations are only assessed against the BAC
* for chromium, the ERL equals the BAC; concentrations are only assessed against the ERL

## Assessment criteria: PAHs in sediment

Two assessment criteria are used to assess PAH concentrations in sediment: the

Background Assessment Concentration (BAC)

Effects Range Low (ERL)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels.  Mean concentrations significantly below the BAC are said to be near background.
[ERLs](http://www.nj.gov/dep/srp/guidance/ecoscreening/esc_table.pdf) were developed by the [United States Environmental Protection Agency](http://www.epa.gov/) for assessing the ecological significance of sediment concentrations.  Concentrations below the ERL rarely cause adverse effects in marine organisms.

BACs and / or ERLs are available for the following PAHs:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BAC** | **BAC** | **ERL** |
|  | All subregions exceptIberian Sea and Gulf of Cadiz | Iberian Seaand Gulf of Cadiz | All subregions |
| Naphthalene |    8 |  | 160  |
| Phenanthrene |  32 |  7.3 | 240  |
| Anthracene |    5 |  1.8 |  85 |
| Dibenzothiophene |  |  | 190  |
| Fluoranthene |  39 | 14.4  | 600  |
| Pyrene |  24 | 11.3  | 665  |
| Benz[a]anthracene |  16 |  7.1 | 261  |
| Chrysene (Triphenylene) |  20 |  8.0 | 384  |
| Benzo[a]pyrene |  30 |  8.2 | 430  |
| Benzo[ghi]perylene |  80 |  6.9 |  85 |
| Indeno[123-cd]pyrene | 103  |  8.3 | 240  |

Notes:

* all concentrations are expressed as μg kg-1 dw
* BACs are normalised to 2.5% organic carbon in all subregions except the Iberian Sea and Gulf of Cadiz, where BACs are not normalised

## Assessment criteria: CBs in sediment

Two assessment criteria are used to assess metal concentrations in sediment: the

* **B**ackground **A**ssessment **C**oncentration (BAC)
* **E**nvironmental **A**ssessment **C**riteria (EAC)

BACs were developed by the [Oslo and Paris Commission](http://www.ospar.org/) (OSPAR) for testing whether concentrations are near background levels. Mean concentrations significantly below the BAC are said to be near background.

EACs were developed by OSPAR and the [International Council for the Exploration of the Sea](http://www.ices.dk/) for assessing the ecological significance of sediment concentrations. Concentrations below the EAC should not cause any chronic effects in marine organisms.

BACs and EACs are available for the following CBs:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BAC** | **BAC** | **EAC** |
|  | All subregions exceptIberian Sea and Gulf of Cadiz | Iberian Seaand Gulf of Cadiz | All subregions |
| CB28 | 0.22 |  |   1.7 |
| CB52 | 0.12 |  |   2.7 |
| CB101 | 0.14 |  |   3.0 |
| CB118 | 0.17 |  |   0.6 |
| CB138 | 0.15 |  |   7.9 |
| CB153 | 0.19 |  | 40.0 |
| CB180 | 0.10 |  | 12.0  |

Notes:

* all concentrations are expressed as μg kg-1 dw
* BACs are normalised to 2.5% organic carbon
* BACs are under development for the Iberian Sea and Gulf of Cadiz, where concentrations are only assessed against the EAC

## Assessment criteria: Organo-metals in sediment

Assessment criteria for organo-metals in sediment are under development.

## Assessment criteria: Organo-bromines in sediment

Assessment criteria for organo-bromines in sediment are under development.

# Appendix 1

## Changes to the assessment methodolgy

Changes made since the 2014 Assessment are described below:

* [2016 Assessment](http://dome.ices.dk/OSPARMIME2016/help_methods_changes.html#Assessment2016)
* [2015 Assessment](http://dome.ices.dk/OSPARMIME2016/help_methods_changes.html#Assessment2015)

Helpfiles for previous assessments can be found below:

2015 assessment

[assessment of contaminants in biota](http://dome.ices.dk/OSPARMIME2016/oldhelp/2015/methods_biota.html)

[assessment of contaminants in sediment](http://dome.ices.dk/OSPARMIME2016/oldhelp/2015/methods_sediment.html)

[assessment of biological effects](http://dome.ices.dk/OSPARMIME2016/oldhelp/2015/methods_biological_effects.html)

[assessment of imposex](http://dome.ices.dk/OSPARMIME2016/oldhelp/2015/methods_imposex.html)

2014 assessment

[assessment of contaminants in biota](http://dome.ices.dk/OSPARMIME2016/oldhelp/2014/methods_biota.html)

[assessment of contaminants in sediment](http://dome.ices.dk/OSPARMIME2016/oldhelp/2014/methods_sediment.html)

[assessment of biological effects](http://dome.ices.dk/OSPARMIME2016/oldhelp/2014/methods_biological_effects.html)

[assessment of imposex](http://dome.ices.dk/OSPARMIME2016/oldhelp/2014/methods_imposex.html)

## 2016 Assessment Modelling of contaminants and biological effects

There were major changes in the way contaminant and biological effects time series were assessed. These included

Modelling the original data, rather than annual indices derived from the data using a linear mixed model that estimated the variance components in the data, rather than a loess smoother applied to the annual indices

Correctly incorporating the analytical variation in the data (supplied as uncertainties), rather than using an ad-hoc ‘scaled weight’ to measure analytical quality

Adapting the likelihood so that less-than measurements are treated as left-censored observations

The changes are so wide-ranging that, to understand them properly, it is probably best to compare the current help files with the 2015 help files (which can be found in the links above).

Modelling of imposex (VDS)

There were also major changes in the assessment of imposex time series when submitted as individual VDS measurements. These included

Modelling the individual measurements, rather than annual indices, using a proportional odds model

Considering smooth changes in VDS levels over time

Considering change-point models in which VDS levels suddenly begin to change; the change-point is constrained to a year in the period 2004-2008, when the ban on the use of TBT was being implemented

Again, to undestand the changes properly, it is best to compare the current help files with the 2015 help file (which can be found in the links above).

## 2015 Assessment

Recent trends

The definition of recent trends was extended from 10 to 20 years for contaminants and biological effects (other than imposex) in biota. This brings it into line with the definition for contaminants in sediment and reflects the increasing use of year-skipping monitoring strategies, particularly for stations with low concentrations. A recent trend thus indicates a significant change in concentration in the period 1994 to 2013 (for the 2015 assessment).

Type and width of smoothing neighbourhood

Loess smoothers are used to model smooth changes in contaminant concentrations (for both biota and sediment) and biological effects measurements (apart from imposex) when there are 7+ years of data. The amount of smoothing is determined by the type and width of the neighbourhood of contaminant indices that is used to estimate each f(t) as t runs from 1 to T. Previously, a fixed-width neighbourhood (Fryer & Nicholson, 1999) was used with, for example, a width of 9 meaning that only the indices in the 9 years closest to t were used to estimate f(t). This worked well if there was annual monitoring, but was less effective when monitoring was less frequent since some parts of the fit were sometimes based on only a few indices. This has been replaced by a neighbourhood in which a fixed number of indices are used to estimate each f(t). For example, a neighbourhood of 9 now uses the 9 indices that are closest to t to estimate f(t). The fit in year tt can now be influenced by indices from years relatively distant to tt, but the fit is always based on the same number of indices. This type of neighbourhood was used in the original development of loess smoothers (Cleveland, 1979).

A greater range of neighbourhood widths are also now considered. Previously, widths of 7, 9, and 11 years were considered, with the final choice being the width giving the smallest Akaike’s Information Criterion corrected for small sample size (AICc). Now, widths of 7, 9, 11 up to T (if T is odd) or T+1 (if T is even) are considered, with the final choice again based on AICc. However, if there is no evidence of nonlinearity in the data (i.e. if the AICc of the linear model is lower than that of the best smoother) then the linear model f(t)=μ+βt is used instead.

Cleveland WS, 1979. Robust locally-weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74: 829-836.

Fryer RJ & Nicholson MD, 1999. Using smoothers for comprehensive assessments of contaminant time series in marine biota. ICES Journal of Marine Science 56: 779-790.

## Determinands

Persistent organic pollutants were introduced as a group of contaminants for biota.

Scope for growth and glutathionine transferase were introduced as biological effects for biota. For both, high values indicate healthy organisms. Glutathionine transferase is assessed in exactly the same way as chemical contaminants in biota, except that the lower confidence limit on the fitted value in the last monitoring year is used to assess status. For scope for growth, the annual indices are the median values of the scope for growth measurements in each year (there is no log transformation). This is because scope for growth can be negative (with negative values indicating bad status). The annual indices are then modelled in the same way as chemical contaminant indices, except that the lower confidence limit is used to assess status.

## Assessment criteria

The ERLs for C1-naphthalene, C2-naphthalene, C1-phenanthrene, C2-phenanthrene and C1-dibenzothiophene were not used as it was not possible to find sufficient justification for them in the literature.

## Assessing status of imposex

Previously, environmental status of imposex levels was assessed using the model fitted to the annual indices. The upper one-sided 95% confidence limit on the fitted value in the most recent monitoring year was compared to the available assessment criteria. However, in many time series, imposex levels have declined so rapidly that the linear models used to assess trends cannot track the change completely. The linear models correctly show evidence of a decline, but over-estimate imposex levels in the final monitoring year suggesting that environmental status is worse than it actually is. To overcome this, an alternative test of status is now used when there are individual measurements in the final monitoring year. A proportional odds model is fitted to the individual measurements and used to place an upper one-sided 95% confidence limit on the annual index in the final monitoring year. This confidence limit is then compared to the available assessment criteria.

# Appendix 2

## Factors for converting the basis of assessment concentrations in biota

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Species** | **Common name** | **% lipid weight****in muscle** | **% lipid weight****in liver** | **% dry weight****in soft body** | **% lipid weight****in soft body** |
| *Clupea harengus* | herring | 4.5 |    6.2 |  |  |
| *Gadus morhua* | cod |  | 45   |  |  |
| *Lepidorhombus whiffiagonis* | megrim |  | 23   |  |  |
| *Limanda limanda* | common dab |  | 16   |  |  |
| *Melanogrammus aeglefinus* | haddock |  | 65   |  |  |
| *Merlangius merlangus* | whiting |  | 45   |  |  |
| *Merluccius merluccius* | hake |  | 44   |  |  |
| *Molva molva* | common ling |  | 54   |  |  |
| *Perca fluviatilis* | European perch | 0.7 |    0.7 |  |  |
| *Platichthys flesus* | flounder |  | 13   |  |  |
| *Pleuronectes platessa* | plaice |  | 10   |  |  |
| *Zoarces vivparus* | eelpout | 0.6 |    0.7 |  |  |
| *Crassostrea gigas* | Pacific oyster |  |  | 19 | 1.8 |
| *Mya arenaria* | softshell clam |  |  | 14 | 0.6 |
| *Mytilus edulis* | blue mussel |  |  | 17 | 1.3 |
| *Mytilus galloprovincialis* | Mediteranean mussel |  |  | 19 | 2.0 |
| *Ostrea edulis* | native oyster |  |  | 22 | 1.8 |
| *Nucella lapillus* | dog whelk |  |  | 34 |  |

# Appendix 3

## Pivot values for normalisation of metals in sediment

**Metals**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Units** | **Digestion\*** | ***cx*** |
| Cadmium | mg kg-1 | Ps, Tot, Pw | 0.03 |
| Mercury | mg kg-1 | Ps, Tot, Pw | 0 |
| Lead | mg kg-1 | Ps, Pw | 2 |
| Lead | mg kg-1 | Tot | 9 |
| Arsenic | mg kg-1 | Ps | 3 |
| Arsenic | mg kg-1 | Tot | 5 |
| Arsenic | mg kg-1 | Pw | 1.5 |
| Chromium | mg kg-1 | Ps, Tot | 13 |
| Chromium | mg kg-1 | Pw | 10 |
| Copper | mg kg-1 | Ps, Pw | 1 |
| Copper | mg kg-1 | Tot | 3 |
| Nickel | mg kg-1 | Ps, Pw | 2.5 |
| Nickel | mg kg-1 | Tot | 4 |
| Zinc | mg kg-1 | Ps, Pw | 8 |
| Zinc | mg kg-1 | Tot | 13 |

**Aluminium (normaliser)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Units** | **Digestion\*** | ***nx*** | ***nss*** |
| % | Ps | 0.4 | 5.0 |
| % | Tot | 1.4 | 5.8 |
| % | Pw | 0.3 | 4.0 |

\* Digestion codes: Pw (Partial weak), Ps (Partial strong), Tot (Total)

# Appendix 4

## Species area combinations used for cut-point estimation

The cut-points are estimated for the following combinations of species and area. To make the estimation problem more tractable, the data are further split by contracting party. The value of *K* is the highest VDS class considered having combined the upper classes that have few observations.

html table generated in R 3.3.1 by xtable 1.8-2 package Mon Oct 24 09:00:19 2016

|  |  |  |  |
| --- | --- | --- | --- |
| **Species** | **Region** | **Country** | **K** |
| *Nucella lapillus* | Celtic Sea | Ireland | 5 |
| *Nucella lapillus* | Celtic Sea | United Kingdom | 4 |
| *Nucella lapillus* | Irish and Scottish West Coast | Ireland | 6 |
| *Nucella lapillus* | Irish and Scottish West Coast | United Kingdom | 5 |
| *Nucella lapillus* | Irish Sea | Ireland | 5 |
| *Nucella lapillus* | Irish Sea | United Kingdom | 4 |
| *Nucella lapillus* | Barents Sea | Norway | 4 |
| *Nucella lapillus* | Norwegian Trench | Norway | 4 |
| *Nucella lapillus* | Skaggerak | Norway | 4 |
| *Nucella lapillus* | Southern North Sea | The Netherlands | 4 |
| *Nucella lapillus* | Southern North Sea | United Kingdom | 4 |
| *Nucella lapillus* | Channel | United Kingdom | 4 |
| *Nucella lapillus* | Northern North Sea | United Kingdom | 5 |
| *Nassarius reticulatus* | Skaggerak | Sweden | 4 |
| *Nassarius reticulatus* | Southern North Sea | The Netherlands | 1 |

1. English only [↑](#footnote-ref-1)
2. Such models cannot be readily fitted in the R statistical environment becuase the analyticalanalytical variance is assumed know. Instead, the likelihood is maximised directly using the optim function. Ideally, the models should be fitted by restricted maximum likelihood (apart from when being used for likelihood ratio tests), but this has not been implemented yet. [↑](#footnote-ref-2)
3. AICc is a model selection criterion that gives greater protection against overfitting than AIC when the sample size is small. For contaminant time series, small sample sizes correspond to few years of data. AICc is not formally defined for mixed models, but the usual definition is adapted to give a sensible criterion for the models considered here. The usual definition of AICc is

AICc = - 2 log likelihood+2kn/(n−k−1)

where n is the sample size and kk is the number of parameters in the model. For a contaminant time series, the natural definition of the sample size is the number of years of data, N, say. The number of parameters in the number of fixed effects parameters, kfixed, plus the number of (unknown) variance parameters, krandom. For example, the linear model has kfixed = 2 and krandom = 2 (or 1 if the sample variance component is subsumed into the year variance component). This suggests using

AICc = - 2 log likelihood+2(kfixed+krandom)N/(N−kfixed−krandom−1)

However, the denominator now overly penalises models because the ‘sample size’ is the number of years and, whilst subtracting krandomkrandom correctly corrects for the year variance component, it also corrects for the sample variance component which measures within-year variation. (Indeed, the denominator = 0 if N = 5 and the linear model is fitted, or NN = 3 or 4 and the mean model is fitted). It therefore makes sense to take krandom in the denominator to be 1, corresponding to the year variance component, giving:

AICc = - 2 log likelihood+2(kfixed+krandom)N/(N−kfixed−2)

The denominator is now analogous to that used in a linear model with a single normally distributed error term. The AICc is still undefined when N = 3 and the mean model is fitted, but this doesn’t matter in practice. [↑](#footnote-ref-3)
4. Methods for estimating the smoothing degrees of freedom as part of the fitting process, for example by treating the amount of smoothing as an extra variance component, are available for several classes of models. However, such methods are not implemented in R for the case when the residual variance (the analyticalanalytical variance) is known. This is a topic for future development. [↑](#footnote-ref-4)
5. Approximate standard errors on the fixed effects parameter estimates are obtained from the Hessian matrix. These are used to estimate standard errors on the fitted values, with confidence intervals based on a t-distribution with NN - kfixedkfixed - 1 degrees of freedom. One-sided t-tests of whether the fitted value in the last monitoring year is below the assessment criteria can be found on the Statistical analysis page on the right hand side of the summary map under Graphics. The standard errors can be computed analytically (i.e. without using the Hessian), but this hasn’t been implemented yet. The degrees of freedom for the t-tests is a sensible approximation because, for time series models, the natural definition of the ‘sample size’ is NN, the number of years of data (see discussion on AICc above). However, if the year variance is small compared to the other variances, the degrees of freedom might be too small leading to a loss of statistical power. This is a topic for future development. [↑](#footnote-ref-5)
6. These tests have a type 1 error that is larger than the nominal value. For example, tests conducted at the 5% significance level will find ‘significant’ trends in more than 5% of time series, even when there are no trends. Using the standard error of the estimate of ββ from a restricted maximum likelihood fit of the linear model would be one way to improve the situation. Better still would be to use the Kenward Roger modification of F tests for linear mixed models (Kenward MG & Roger JH, 1997; Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, Biometrics 53: 983-997). [↑](#footnote-ref-6)
7. Imposex in Nassarius reticulatus, Neptunea antiqua, Nucella lapillus and Ocenebra erinaceus is assessed using VDS. Imposex in Littorina littorea and Buccinum undatum is assessed using intersex stage and imposex stage respectively.  [↑](#footnote-ref-7)
8. K = 6 for Nassarius reticulatus, Nucella lapillus and Ocenebra erinaceus and K = 4 for Neptunea antiqua.  [↑](#footnote-ref-8)
9. An additional constraint is necessary for identifiability since f(t) is also in the linear predictor. Typically, one of the intermediate cut-points is set to zero.  [↑](#footnote-ref-9)
10. Even with multiple time series, there are sometimes very few snails with VDS measurements in the highest class K. This can lead to difficulties estimating the highest cut-point, so the pragmatic decision is taken to combine the upper two classes with e.g. K reducing from 6 to 5. If there are still few snails in the (new) highest class, the process is repeated with e.g. K reducing to 4.  [↑](#footnote-ref-10)
11. Even with multiple time series, there are sometimes very few snails with VDS measurements in the highest class KK. This can lead to difficulties estimating the highest cut-point, so the pragmatic decision is taken to combine the upper two classes with e.g. K reducing from 6 to 5. If there are still few snails in the (new) highest class, the process is repeated with e.g. K reducing to 4.

5 The variance of the parameter estimates is obtained from the Hessian matrix in the usual way. If there is any evidence of over-dispersion (see footnote 7), the variance matrix is then multiplied by the estimate of the dispersion parameter.  [↑](#footnote-ref-11)
12. If all the VDS measurements for a series of years are equal to K, then there is negligible information with which to ‘anchor’ the estimates of f(t). All we know is that, on the logistic scale, f(t) could be anywhere between ‘large’ and infinite.  [↑](#footnote-ref-12)
13. AICc is a model selection criterion that gives greater protection against overfitting than AIC when the sample size is small. The usual definition of AICc is

AICc = - 2 log likelihood + 2pn/(n−p−1)

where n is the sample size and p is the number of parameters in the model. For a VDS time series, the natural definition of the sample size is the number of years of data, N. (One might consider using Nmid but things are complicated enough as it is.) Further, pp is the number of parameters associated with f(t) (with the cut-points ignored). For example, the linear model has p = 2. However, there is often evidence of over-dispersion and, although AICc is then formally undefined, a sensible adjustment can be made by dividing the log likelihood by an estimate of the dispersion parameter, and extending the second term to account for the additional (dispersion) parameter. This gives

AICc = - 2 log likelihood/ϕ + 2(p+1)N/(N−p−2)

where ϕ is the dispersion parameter. The value of ϕ is common to all the candidate models and is estimated by fitting all the candidate models in turn and comparing each to the fit of a full model. Let di be the deviance (- 2 log likelihood) of candidate model i and let pi be the corresponding number of model parameters. Further, let dfulldfull be the deviance of a full model in which f(t)=μt; i.e. there is a separate parameter estimated for each year. Then the dispersion parameter for model ii is estimated to be

ϕi=max(1,di−dfull/(N−pi))

and the dispersion parameter used in the AICc calculations is ϕ=min(ϕi). If N≤4, then AICc is undefined, and AIC is used instead where

AIC = - 2 log likelihood/ϕ + 2p [↑](#footnote-ref-13)
14. The estimates of the parameters of f(t) are assumed to be normally distributed with variance obtained from the Hessian matrix of the likelihood of the individual time series data (and multiplied by the over-dispersion parameter). The estimates of the cut-points are also assumed to be normally distributed with variance obtained from the Hessian matrix of the likelihood of the multiple time series data used to estimate the cut-points. For simplicity, the two sets of estimates are assumed to be independent. Typically, 1000 realisations are simulated.  [↑](#footnote-ref-14)
15. Let dfinal be the deviance (-2 log likelihood) of the linear (change-point) model, dmean be the deviance of the mean model, and ϕ be the dispersion parameter. Then F=(dfinal−dmean)/ϕ is referred to an F distribution on 1 and N−2 degrees of freedom.  [↑](#footnote-ref-15)
16. The t test has N−p degrees of freedom, where p is the number of parameters in f(t).  [↑](#footnote-ref-16)
17. Such models cannot be readily fitted in the R statistical environment because the analytical variance is assumed know. Instead, the likelihood is maximised directly using the optim function. Ideally, the models should be fitted by restricted maximum likelihood (apart from when being used for likelihood ratio tests), but this has not been implemented yet.  [↑](#footnote-ref-17)
18. AICc is a model selection criterion that gives greater protection against overfitting than AIC when the sample size is small. For contaminant time series, small sample sizes correspond to few years of data. AICc is not formally defined for mixed models, but the usual definition is adapted to give a sensible criterion for the models considered here. The usual definition of AICc is

AICc = - 2 log likelihood+2kn/(n−k−1)

where n is the sample size and k is the number of parameters in the model. For a contaminant time series, the natural definition of the sample size is the number of years of data, N, say. The number of parameters in the number of fixed effects parameters, kfixed, plus the number of (unknown) variance parameters, krandomkrandom. For example, the linear model has kfixed = 2 and krandom = 2 (or 1 if the sample variance component is subsumed into the year variance component). This suggests using

AICc = - 2 log likelihood+2(kfixed+krandom)N/(N−kfixed−krandom−1)

However, the denominator now overly penalises models because the ‘sample size’ is the number of years and, whilst subtracting krandomkrandomcorrectly corrects for the year variance component, it also corrects for the sample variance component which measures within-year variation. (Indeed, the denominator = 0 if N = 5 and the linear model is fitted, or N = 3 or 4 and the mean model is fitted). It therefore makes sense to take krandomkrandom in the denominator to be 1, corresponding to the year variance component, giving

AICc = - 2 log likelihood+2(kfixed+krandom)N/(N−kfixed−2)

The denominator is now analogous to that used in a linear model with a single normally distributed error term. The AICc is still undefined when N= 3 and the mean model is fitted, but this doesn’t matter in practice.  [↑](#footnote-ref-18)
19. Methods for estimating the smoothing degrees of freedom as part of the fitting process, for example by treating the amount of smoothing as an extra variance component, are available for several classes of models. However, such methods are not implemented in R for the case when the residual variance (the analyticalanalytical variance) is known. This is a topic for future development.  [↑](#footnote-ref-19)
20. Approximate standard errors on the fixed effects parameter estimates are obtained from the Hessian matrix. These are used to estimate standard errors on the fitted values, with confidence intervals based on a t-distribution with N - kfixed - 1 degrees of freedom. One-sided t-tests of whether the fitted value in the last monitoring year is below the assessment criteria can be found on the Statistical analysis page on the right hand side of the summary map under Graphics. The standard errors can be computed analytically (i.e. without using the Hessian), but this hasn’t been implemented yet. The degrees of freedom for the t-tests is a sensible approximation because, for time series models, the natural definition of the ‘sample size’ is N, the number of years of data (see discussion on AICc above). However, if the year variance is small compared to the other variances, the degrees of freedom might be too small leading to a loss of statistical power. This is a topic for future development.  [↑](#footnote-ref-20)
21. These tests have a type 1 error that is larger than the nominal value. For example, tests conducted at the 5% significance level will find ‘significant’ trends in more than 5% of time series, even when there are no trends. Using the standard error of the estimate of β from a restricted maximum likelihood fit of the linear model would be one way to improve the situation. Better still would be to use the Kenward Roger modification of F tests for linear mixed models (Kenward MG & Roger JH, 1997; Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, Biometrics 53: 983-997).  [↑](#footnote-ref-21)