Document title	Proposed assessment protocol for hazardous substances concentration core indicators
Code	4J-15
Category	DEC
Agenda Item	4J – HELCOM indicators and assessments
Submission date	14.10.2016
Submitted by	EN-Hazardous substances, indicator Lead Countries Sweden, Finland, Denmark, Poland
Reference	EN-HZ 4-2016 (point 13 of the outcome)

Background

The proposal to apply the OSPAR assessment method (referred to as the 'MIME R-script') also in HELCOM core indicators as the assessment protocol was first presented and discussed at the CORESET II hazardous substances and bio-effects meeting (CORESET II HZBF 2015) in February 2015 (paragraph 3.6-3.8 of the outcome). As a basis for the discussion, Sweden submitted a meeting document comparing the OSPAR approach to the HELCOM approach which is based on a national approach developed in Sweden. The CORESET II HZBF 2015 meeting concluded that the statistical evaluation methods produce comparable results. It was also noted that the OSPAR approach automates some of the calculation steps, and is thus labour saving.

Applying the OSPAR assessment method to the HELCOM core indicators was further discussed at the first meeting of the HELCOM expert network on hazardous substances (EN-HZ 1-2016) held in conjunction with the BalticBOOST HZ WS 1-2016 (points 6-10 of the outcome). The method comparison document from 2015 was updated with further clarifying information (meeting document 1-5 of the EN-HZ 1-2015 meeting). It was concluded that the methods are comparable and that the assessment outcome should not be significantly affected irrespective of which method is applied. The workshop recommended using the method developed in OSPAR MIME as the HELCOM core indicator assessment protocol (point 7 of the outcome), while identifying that further intersessional work is needed to clarify some remaining details regarding e.g. data checks and formats.

Two intersessional online meetings were consecutively organized to discuss the detailed questions and arrange for testing. The progress made during the intersessional work was presented at the EN-HZ 2-2016 (meeting document 3-1 of EN-HZ 2-2016). The intersessional work focussed on testing the OSPAR MIME R-script on two HELCOM core indicators using the OSPAR criteria with the main aim to identify any problems with applying the script to HELCOM COMBINE data. It should thus be noted that the tests have not used HELCOM GES boundaries (test results included in this document), however the script can be adapted to assess the HELCOM GES boundaries. The OSPAR assessment criteria excluded a significant amount of data in the first test run and were considered to be too strict. Based on a second extraction from COMBINE, the tests were re-run with more lenient criteria considered to be relevant for the HELCOM area. Among the details on these criteria it was noted that the script requires data points to be identified to a station name in the ICES station dictionary, and that in the future it is proposed that countries report uncertainty values related to each measurement.

The EN-HZ 4-2016 meeting concluded that the OSPAR assessment method can be used for the core indicators related to concentration of contaminants and agreed that the script should be modified so that the HELCOM GES boundary values are applied (point 13 of the outcome).
The assessment protocol would thereby apply to the following core indicators:

- Hexabromocyclodecane (HBCDD)
- Metals (Hg, Cd, Pb)

- Polybrominated biphenyl ethers (PBDE)
- Perfluorooctane sulphonate (PFOS)
- Polyaromatic hydrocarbons (PAH) and their metabolites
- Polychlorinated biphenyls (PCB) and dioxins and furans
- TBT and imposex

This document presents the proposal on applying the OSPAR assessment system for concentration based HELCOM core indicators as the indicator assessment protocol. The document is structured to first provide a summary of the protocol and an overview of the main steps (1-3) after which examples of the outputs are presented. At the end of the document the main steps of the protocol are described in technical detail.

Action requested

The Meeting is invited to:

- consider the expert level proposal to use the OSPAR assessment approach ('MIME R-script') as the assessment protocol for the concentration core indicators,
- endorse the use of this approach for the specified HELCOM core indicators as the assessment protocol.

Proposed assessment protocol for hazardous substances concentration core indicators

COMBINE data extraction and pre-processing

Data on contaminant concentration are retrieved from the HELCOM COMBINE database, hosted at ICES. The retrieval is carried out using an extraction table, where the relevant parameters for the core indicators are specified. The extraction table is specified for each core indicator.

The COMBINE reported data may require treatment before the statistical evaluation using the R-scipts is carried out. The data needs to be expressed in the same unit and base, e.g. µg/kg wet weight consistently throughout the dataset. Adjustments for e.g. lipid- and dry weight content, or other adjustments of the data may be necessary to be able to conduct a relevant assessment to the agreed GES boundaries. These recalculations are done automatically if the relevant lipid or dry weight content are available for each sample, or if there is an agreed proxy value for the species and organ. Otherwise the sample will be excluded from the assessment.

Summary of the assessment protocol

The assessment protocol has undergone rigorous developments in the OSPAR framework. The HELCOM Expert Network on Hazardous Substances is of the general opinion that the statistical methods proposed are sound and relevant for the assessment of hazardous substances concentrations in the marine environment.

The assessment protocol is structured in three main parts, 1) a time series analysis per station, 2) a compliance check against the GES boundary per station and 3) a spatial aggregation per assessment unit.

It should be noted that the assessment protocol makes the assumption that monitoring data stems from the same monitoring stations during consecutive years. The stations used by the protocol are defined in the ICES station dictionary. Stations with similar station name is grouped together, but it is also possible to define a group of stations with different names to be defined as the same station in the station dictionary. Usually a station is defined in the station dictionary with coordinates and a valid box around these coordinates, but coordinates outside of the box will only give a warning when reporting the data, and are not used in the actual data extraction.

This document is structured so that an overview of the main steps in the protocol are first described in order to provide an overview. At the end of the document each stage is described in more technical detail. There are detailed descriptions available both for the cases where all the concentration measurements are above the detection limit and a section describing how the methodology is adapted when there are ‘less-than’ measurements (i.e. measurements below the Limit of Quantification).

Overview of the assessment protocol for contaminants in biota and sediment

1) Overview of the assessment protocol for time series trend analysis per station

Time series of contaminant concentrations are assessed in four stages:

1. The concentration measurements from each year are summarised by an annual contaminant index. A weight is assigned to each index, scaled to lie between 0 and 1, that incorporates information about the analytical quality of the index.
2. The scaled weights are converted into statistical weights that account for the relative magnitudes of the analytical and environmental variability in the data.

3. A weighted regression model is fitted to the annual contaminant indices. The type of model depends on the number of years of data:
    - 1-2 years: no model
    - 3-4 years: mean
    - 5-6 years: linear trend
    - 7+ years: smoother

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. It should be noted that in order to assess a specific time period such as the period 2011-2016 as agreed to be used for the second HELCOM holistic assessment in the HOLAS II project, a time trend analysis is always carried out as a basis for the most recent 20 years, if more than 5 years of data exists, and the final point is within 2011 – 2016.. The robustness, i.e. confidence, of the trend improved by a longer time trend. The modelled value for the most recent year, and the upper 95% confidence limit for this year, is used in the next step.

2) Overview of the assessment protocol for the compliance check to GES boundary per station

The fitted models (from the time trend analysis) are used to assess environmental status against available assessment criteria. Compliance checking against the environmental status is assessed by comparing the upper one-sided 95% confidence limit on the fitted value in the most recent monitoring year to the available assessment criteria as a default.

For example, if the upper confidence limit is below the GES boundary, then the mean contaminant index in the most recent monitoring year is significantly below the GES boundary and concentrations are said to be in GES.

3) Spatial aggregation per assessment unit

For spatial aggregation only stations that are representative of general conditions are considered and those stations impacted due to a point source and baseline stations where trends would not be expected are excluded. The analysis is further restricted to sub-regions where there are at least three stations with trend information and where those stations have reasonable geographic spread.

The trend of each compound at each station is summarised by the estimated annual change in log concentration, with its associated standard error. The annual change in log concentration is then modelled by a linear mixed model with fixed effects.

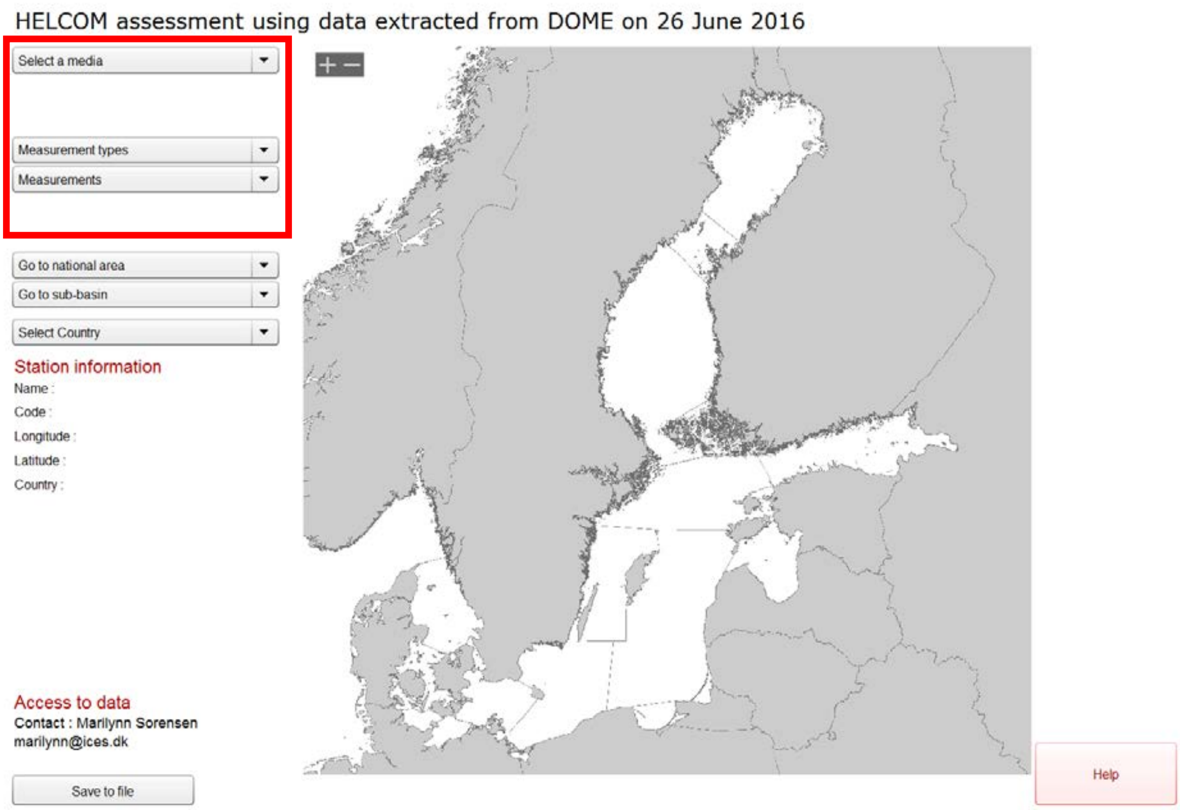
Similar analyses explores status at the sub-regional scale.

Test results showing the output of the scripts

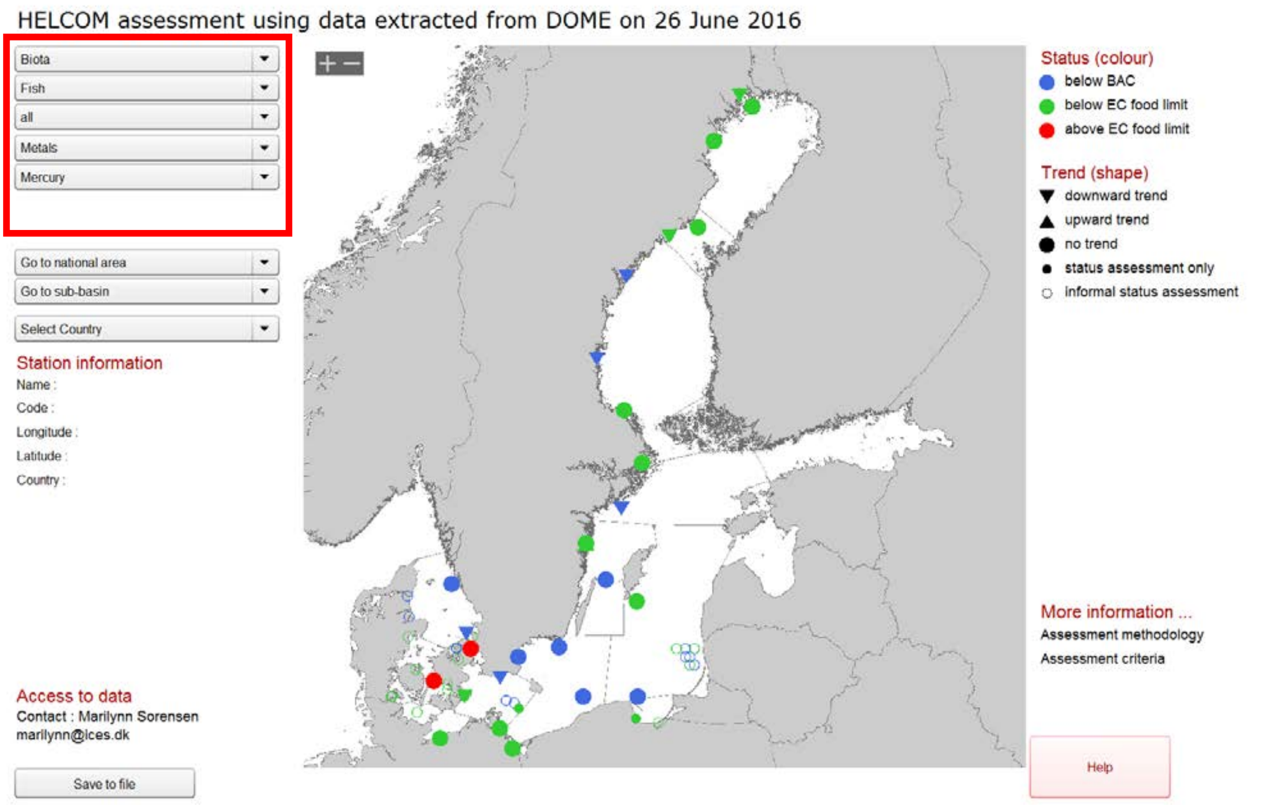
The test results can be viewed online showing the results per station in the output designed for OSPAR purposes

To view the results:

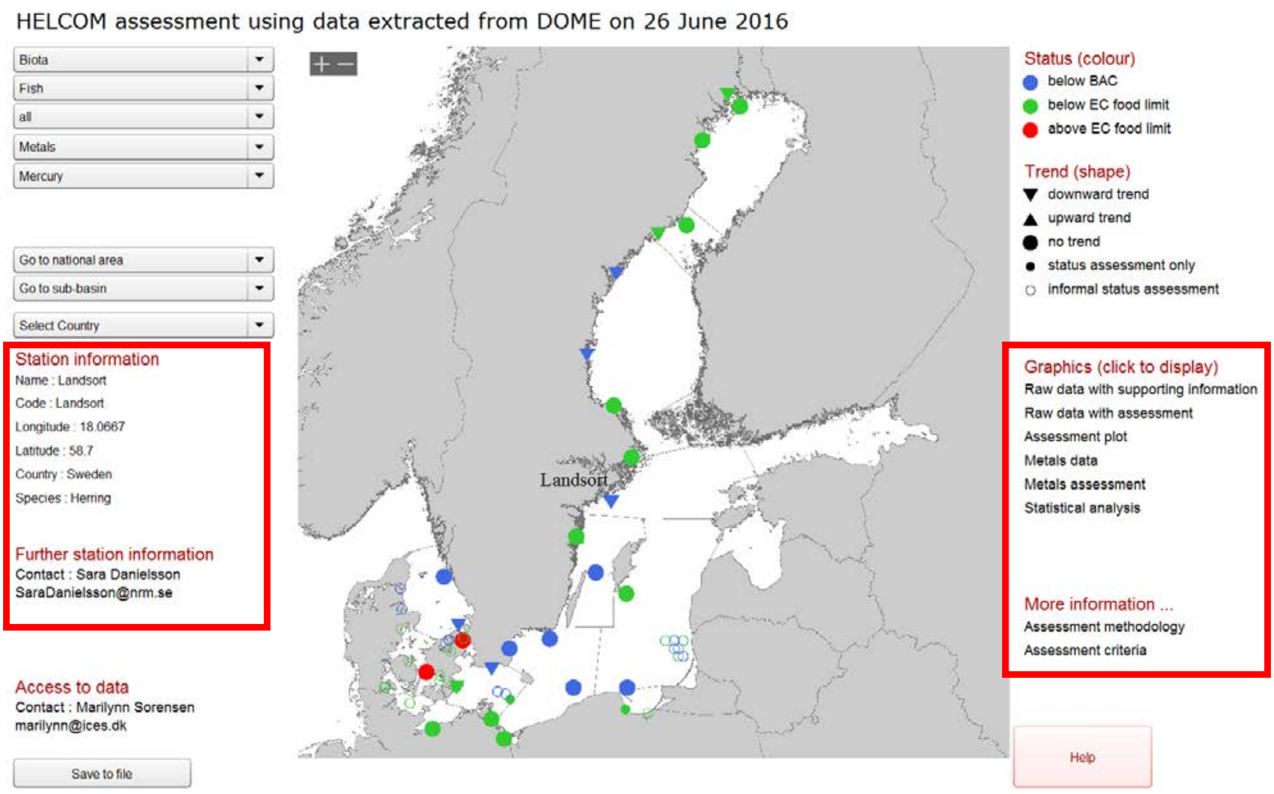
Select from the curtains on the menu to the left. It is important make a selection for each curtain.



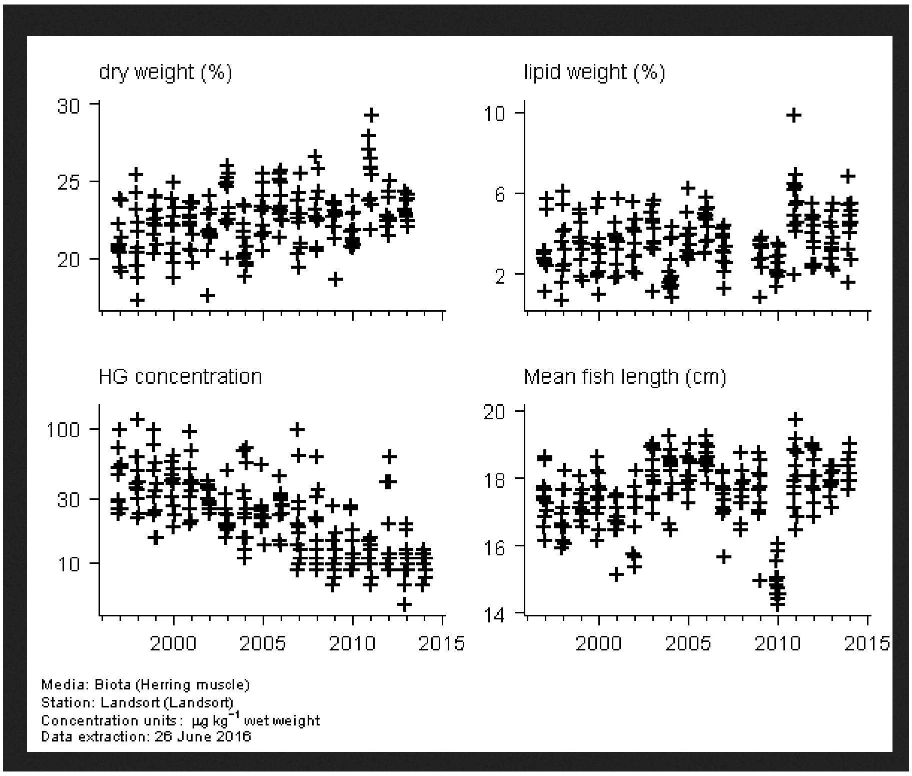
When the selection is made the results of the assessment evaluating trends and the environmental status against available GES boundaries are visible for each station (this test-version have not used HELCOM GES boundaries but OSPAR BAC and food limits). More detailed information can be viewed by choosing a specific station.



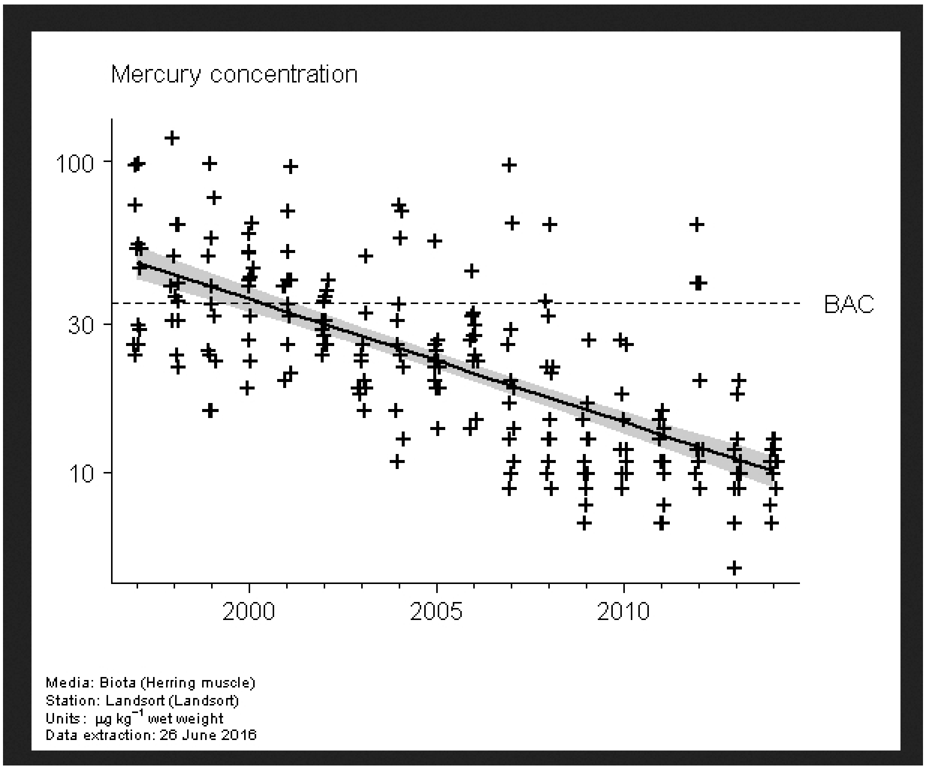
To be able to see detailed output for the specific stations click on a station, then station information will appear on the left hand side and station specific assessment output data will appear on the right hand side together with assessment methodology and assessment criteria.



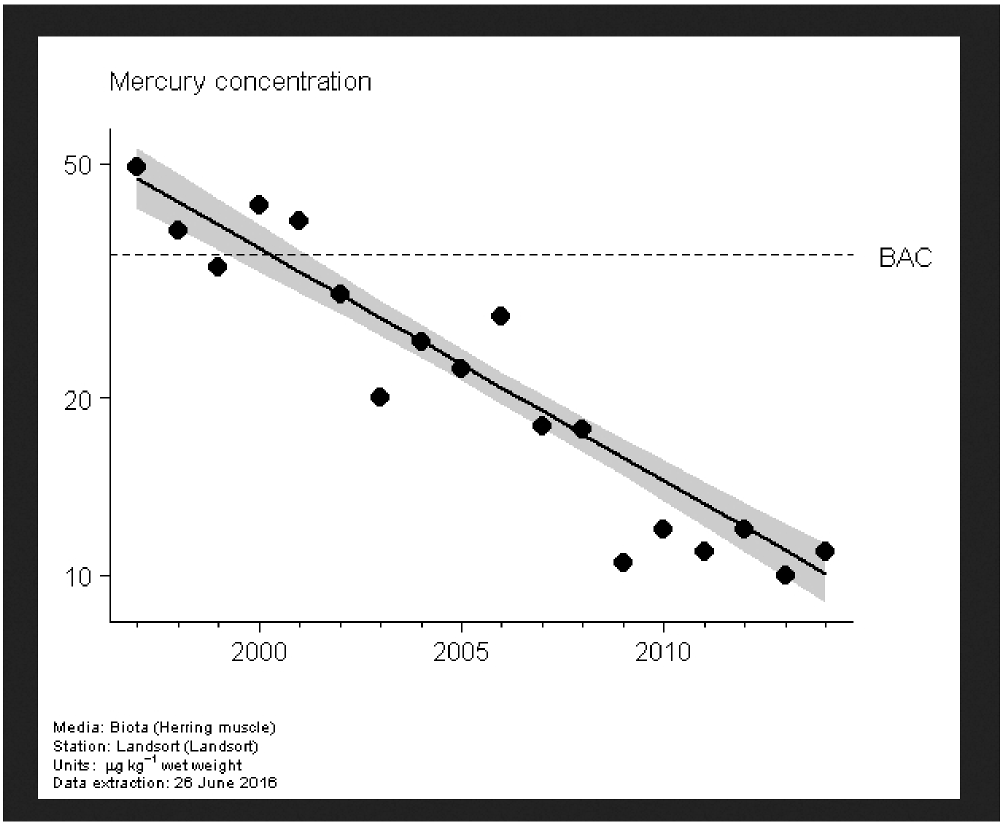
Raw data with supporting information



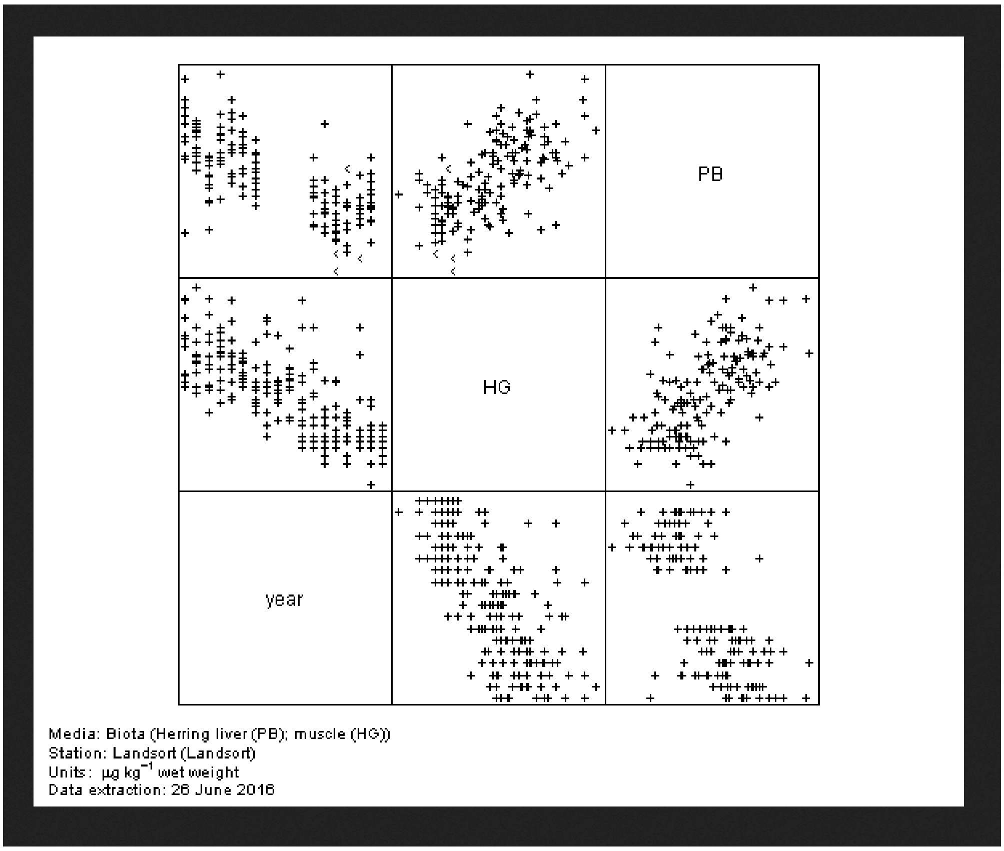
Raw data with assessment



Assessment plot



Metals data



Technical description of the protocol

1) Technical description: Calculating annual contaminant indices when all measurements are above the detection limit

Let c_{ti}, $i = 1 \ldots n_t$ be the concentrations measured in year t, $t = 1 \ldots T$. The annual contaminant index in year t is the median log concentration:

$$y_t = \text{median}\{\log(c_{ti}), i = 1 \ldots n_t\}$$

The scaled weight associated with y_t is based on the available quality assurance information and reflects the analytical quality of the index. A weight of 1 indicates good quality and 0 indicates atrocious quality. However, the weight is 'intuitive', rather than arising from some formal statistical procedure. Some justification can be found in Nicholson & Fryer (2002) and arises from the need to deal with historic data for which there is no analytical quality information. Statistical weights that provide a sensible balance between the scaled weights and the environmental variability in the data are found before fitting the temporal trend models.

The scaled weight is based on the analytical standard deviations of the concentration measurements, also known as the uncertainties, when these are reported to the ICES database. Let σ_{ti} be the analytical standard deviation of c_{ti}. The scaled weight in year t is then:

$$w_t = \begin{cases} 1 & \text{if } \max\{\frac{\sigma_{ti}}{c_{ti}}, i = 1 \ldots n_t\} < 0.125 \\ 0.2 & \text{otherwise} \end{cases}$$

This can be interpreted as the index having good analytical quality if the coefficient of variation of each measurement is less than 12.5%.

When the analytical standard deviations are unavailable (reporting them has only been mandatory since 2011), the scaled weights are based on QUASIMEME Z-scores supplied to ICES on CD by analytical laboratories and Certified Reference Material (CRM) concentrations held in the ICES database. The QUASIMEME Z-scores (for the contaminant, matrix and year) are considered to indicate acceptable performance (pass) if

$$\sum_i Z_i^2 < \chi^2_n(0.95)$$

where n is the number of Z-scores and $\chi^2_n(0.95)$ is the 95 percentile of a χ^2 distribution on n degrees of freedom. Similarly, the CRM concentrations are converted to Z-scores, assuming a coefficient of variation of 12.5%, as

$$Z_i = \frac{M_i - A_i}{0.125 A_i}$$

where M_i are the measured concentrations and A_i are the certified reference values, and considered to indicate acceptable performance if

$$\sum_i Z_i^2 < \chi^2_n(0.95)$$

The scaled weight is then given by

Performance		
QUASIMEME	CRM	scaled weight w_t
pass	pass	1.0

pass	fail / absent	1.0
absent	pass	1.0
fail	pass	0.7
fail / absent	fail / absent	0.2

The minimum scaled weight assigned is 0.2, and corresponds to poor analytical quality in the reported quality assurance information, or no quality assurance information at all (typically the case for historic data, submitted before the reporting of quality assurance information became mandatory).

Nicholson MD, Fryer RJ, 2002. Weighted smoothers for assessing trend data of variable analytical quality. ICES Working Group on Statistical Aspects of Environmental Monitoring 2002. ICES CM 2002/E:04.

Converting scaled weights to statistical weights

The contaminant time series are assessed for temporal trends by fitting a weighted regression model to the annual contaminant indices. Doing so is straightforward if the statistical weights are known beforehand. The statistical weights should be inversely related to the total environmental and analytical variance each year. However, many time series are not supported by analytical quality information for all years, so optimal weights can not be calculated.

The absence of quality assurance information is one of the reasons for constructing the scaled weights. The scaled weights reflect the analytical quality of the annual contaminant indices, but the actual values chosen are arbitrary and take no account of the relative importance of the analytical variance to the total environmental and analytical variance. Using the scaled weights directly in a trend assessment might be inappropriate. For example, all data with 'poor' analytical quality will have the same statistical weight, even though such data should be down-weighted less when the environmental variance dominates the analytical variance (when poor analytical quality doesn't matter so much). An iterative procedure is used to convert the scaled weights to statistical weights that account for the relative magnitudes of the environmental and analytical variances.

Assume that the contaminant time series can be described by the model:

$$yt=f(t)+\varepsilon t$$

where y_t is the annual contaminant index in year t, $f(t)$ is a smooth function of time (possibly linear) describing the underlying trend in contaminant levels, and ε_t is the 'noise' in year t from both environmental and analytical variation. Further, assume that the noise can be decomposed into two terms:

$$\varepsilon t=\tau t+\delta t$$

where τ_t is the noise due to analytical variation and δ_t is the noise due to all remaining sources of environmental variation. Finally, assume that the noise terms are mutually independent and normally distributed:

$$\tau t\sim \mathrm{N}(0,\sigma 2twt)$$

$$\delta\sim \mathrm{N}(0,\sigma 2\delta)$$

where w_t is the scaled weight for year t. Given this model, the appropriate statistical weights are:

$Wt=(\sigma 2\delta+\sigma 2\tau)(\sigma 2\sigma+\sigma 2\tau wt)-1$

The statistical weights provide an appropriate balance between the two variance components and satisfy $0 \le w_t \le W_t \le 1$.

The variance components $\sigma 2\delta$ and $\sigma 2\tau$ are unknown, so must be estimated to give the statistical weights. The approach used relies on the fact that the residuals r_t from an unweighted fit to the data should become more variable as the scaled weights decrease (for example as analytical quality degrades). To a first approximation, the squared residuals $r2t$ have mean $\sigma 2\delta+\sigma 2\tau/wt$. Thus, if the squared residuals $r2t$ are regressed against $1 / w_t$ the intercept and slope should provide estimates of $\sigma 2\delta$ and $\sigma 2\tau$ respectively. Formally, the regression is done using a generalised linear model with gamma errors and identity link. The estimates of $\sigma 2\delta$ and $\sigma 2\tau$ are then plugged into the formula for the statistical weights. Sometimes, the approach will give negative variance estimates, in which case the relevant estimates are taken to be zero.

Modelling the annual contaminant indices

The annual contaminant indices are modelled as:

$yt=f(t)+\varepsilon t$

where y_t is the annual contaminant index in year t, $f(t)$ is a smooth function of time (possibly linear) describing the underlying trend in contaminant levels, and ε_t is an error term assumed to be independent and normally distributed with variance σ^2 / W_t, where W_t are appropriate statistical weights.

The form of $f(t)$ 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 $f(t) = \mu$
>
> 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 $f(t) = \mu + \beta t$
>
> the contaminant indices vary linearly with time; the fitted model is used to assess status and evidence of temporal change

7+ years

> smooth model $f(t)$ = smooth function of time
>
> the contaminant indices vary smoothly over time; the fitted model is used to assess status and evidence of temporal change

A loess smoother is used to estimate the smooth function of time. The amount of smoothing is controlled by the neighbourhood of contaminant indices that is used to estimate each $f(t)$ as t runs from 1 to T. A neighbourhood of 9, for example, uses the 9 indices that are closest to t to estimate $f(t)$. A sequence of neighbourhoods are considered (7, 9, 11 up to T, if T is odd, or $T + 1$, if T is even) with the final choice based on Akaike's Information Criterion corrected for small sample size (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 $f(t) = \mu + \beta t$ is used instead.

Weighted linear regression is described by e.g. Draper & Smith (1998). Loess smoothers were developed by Cleveland (1979). The application of loess smoothers to contaminant time series is described by Fryer & Nicholson (1999).

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

Draper NR & Smith H, 1998. Applied regression analysis, 3rd edition. Wiley

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.

2) Technical description
Assessing environmental status and temporal trends

Environmental status and temporal trends are assessed using the model fitted to the annual contaminant indices.

Environmental status is assessed by comparing the upper one-sided 95% confidence limit on the fitted value in the most recent monitoring year to the available assessment criteria. For example, if the upper confidence limit is below the Background Assessment Concentration (BAC), then the mean contaminant index in the most recent monitoring year is significantly below the BAC and concentrations 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 comparing the contaminant index (1 year) or the larger of the two contaminant indices (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), there is evidence of a temporal trend if the slope β of the linear regression of y_t on t is significant at the 5% level. When a smooth model has been fitted, the fitted smoother is used to test for evidence of any systematic change in contaminant levels over time; this test is also decomposed into both a nonlinear and linear component. The results for each time series can be found in the statistical analysis output on the right hand side of the map under Graphics. Details of the methodology are in Fryer & Nicholson (1999). However, the summary maps focus on changes in contaminant levels in the most recent twenty monitoring years (i.e. between 1994 and 2013, since the assessment only includes data up to 2013). For this, the fitted value of the smoother in 2013 is compared to the fitted value in 1994 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 2013, then the fitted value in the last monitoring year is used instead. Similarly, if the time series starts after 1994, 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.

3) Technical description
Spatial aggregation measurements
Described as examples from the OSPAR assessment for ‘metals’ and ‘PAH’

Example for ‘metals’ from the OSPAR assessment
For spatial aggregation only stations that are representative of general conditions are considered and those stations impacted due to a point source and baseline stations where trends would not be expected are excluded. The analysis is further restricted to sub-regions where there are at least three stations with trend information and where those stations have reasonable geographic spread.

The trend of each compound at each station is summarised by the estimated annual change in log concentration, with its associated standard error. The annual change in log concentration is then modelled by a linear mixed model with fixed effects:

~ metal : sub-region

and random effects:

~ station + metal : station + within-series variation

The choice of fixed and random effects was motivated by the assumption that the metals could have very different trends as they have different sources and are metabolised differently. Thus, the fixed effects measure the trend in each metal in each sub-region and the random effects measure variation in trends:

- between stations common across metals (station)
- residual variation (metal : station + within-series variation)

There are two residual terms. Within-series variation is the variation associated with the estimate of the trend from the individual time series and is assumed known (and given by the square of the standard error). Metal: station allows for any additional residual variation.

Evidence of trends in metal concentrations at the sub-regional scale was then assessed by plotting the estimated fixed effects with point-wise 95 % confidence intervals.

Similar analyses explored status at the sub-regional scale. Two summary measures were considered: the log ratio of the fitted concentration in the last monitoring year to the EC; and the log ratio of the fitted concentration in the last monitoring year to the BAC. Baseline stations were also included in these analyses.

Example for PAHs from the OSPAR assessment

For the regional assessment of trends two of these conditions are more stringent:

- the monitoring station are classified as ‘representative’; ‘baseline’ stations (near pristine conditions or only subject to very diffuse inputs) are also omitted because any trends there will likely be caused by different processes to those at representative stations
- there are at least five years of data (over the whole time series)

Regional assessment - methods

Tabulating the number of time series in each status category by biogeographic region provides a quick summary of the individual time series results, but it does not provide an objective regional assessment of status. Similarly, tabulating the number of time series with an upward or downward trend does not provide an objective regional assessment of trend. This section describes how the individual time series results can by synthesised in a meta-analysis to assess both status and trend at the biogeographic regional level. The following sections show the results of the meta-analysis.

For a regional trend assessment, the trend in each time series is summarised by the estimated change in log concentration over the last twenty years (or shorter if the time series doesn't extend that far back). Regional trends are then estimated by fitting the following linear mixed model by restricted maximum likelihood:

- response: trend (yearly change in log concentration)
- fixed model: biogeographic region
- random model: PAH + biogeographic region.PAH + station + trend estimation variation + residual variation

The fixed model means that a trend is estimated for each biogeographic region. Trends in the different PAHs are assumed to be broadly similar, because inputs and control measures are likely to affect all PAHs in a similar way. Each regional trend can thus be interpreted as the common trend in PAH concentrations in that region (i.e. the trend averaged across all PAHs).

The random model has five terms:

- PAH allows for variation in trend between PAHs common to all regions
- biogeographic region.PAH allows for additional variation in trend between PAHs within biogeographic regions
- station allows for variation in trend between stations common to all PAHs
- trend estimation variation is the variance of the trend estimates from the individual time series analysis, assumed known and fixed
- residual variation is the variation that cannot be explained by any of the fixed effects or the other random effects

The meta-analysis is restricted to:

- biogeographic regions with at least three trend stations with good geographic spread. Three stations is considered the minimum required to provide some sort of evidence base at the biogeographic regional level
- PAHs with trend information for at least three stations (over all regions): PAHs are treated as random effects in the meta-anlaysis, so information about trends in one PAH can be 'borrowed' from the trend information in the other PAHs

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 PAHs 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: biogeographic region
- random model: PAH + biogeographic region.PAH + 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. Again, the meta-analysis is restricted to biogeographic regionsnations with at least three status stations with good geographic spread and to PAHs with status information for at least three stations (over all regions):

The results of both the trend and status regional assessments are back-transformed for presentation. The estimated regional trend is then interpreted as the percentage yearly change in concentration and the estimated regional status as the ratio of the mean regional concentration to the assessment concentration.

To illustrate the latter, a value of 1 indicates that the mean regional concentration is equal to the assessment concentration; a value of 0.5 indicates that it is half the assessment concentration, and a value of 2 indicates that it is twice the assessment concentration.

Treatment of 'less-than' measurements

Overview

Time series with 'less-than' measurements present all sorts of statistical problems, particularly if the limit of detection varies over time. The appropriate way of dealing with them depends on the proportion of measurements that are less-thans and the distribution of less-thans through the time series (i.e. all at the start, or spread evenly through the time series). Efficient statistical methods can often be developed for time series with full and reliable information. However, these methods can be complex and difficult to apply in large-scale assessments where the quality of information can degrade going back in time. Consequently, pragmatic adjustments to the standard methodology have been made to allow sensible, but sub-optimal, assessments of all time series that contain less-thans. There are adjustments to the calculation of the annual indices and to the length of the time series that is used to assess temporal trends. When most of the data are less-thans, a non-parametric test is used to compare levels with assessment criteria.

Calculating annual indices

The calculation of the annual indices is unaffected by less-thans. However, an annual index is treated as a less-than index if, that year, any measurement greater or equal to the median value is a less-than.

Formally, let c_{ti}, $i = 1 \ldots n_t$ be the measurements in year t, $t = 1 \ldots T$. Further, assume the measurements are ordered within-years so that $c_{t1} \leq c_{t2} \leq \ldots \leq c_{tnt}$. The annual index in year t is then regarded as a less-than if any of the c_{ti}, $i = m \ldots n_t$ is a less-than, where $m = (n_t + 1)/2$ if n_t is odd or $m = n_t/2$ if n_t is even.

Length of time series

When some of the annual indices are less-thans the time series is truncated to ensure that the less-thans do not unduly influence the assessment of temporal trends. The three rules are that:

1. The time series is truncated from the left; i.e. early years are omitted, later years are retained.
2. The time series cannot start with a less-than.
3. The time series cannot contain more than one less-than.

Let y_t be the annual index in year t, $t = 1 \ldots T$, and let $z_t = 1$ if y_t is a less-than and $z_t = 0$ otherwise. Then the time series that is assessed for temporal trends is y_t, $t = \mathrm{m} \ldots T$, where m is chosen to give the longest time series subject to the constraints that $z_m = 0$ and

∑*mTzt*≤1.

Non-parametric assessment of environmental status

If the length of the truncated time series is 2 years or less, then there are insufficient years to fit a parametric model and make a formal assessment of environmental status. However, if the original time series has more than five years of data, a one-sided sign test is used instead to provide a non-parametric test of status. All the indices in the last twenty years (the same period used to assess recent trends) are used to test the null hypothesis: H_0: $\mu \geq AC$ against the alternative: H_1: $\mu < AC$, where μ is the mean level and AC is the assessment criterion.