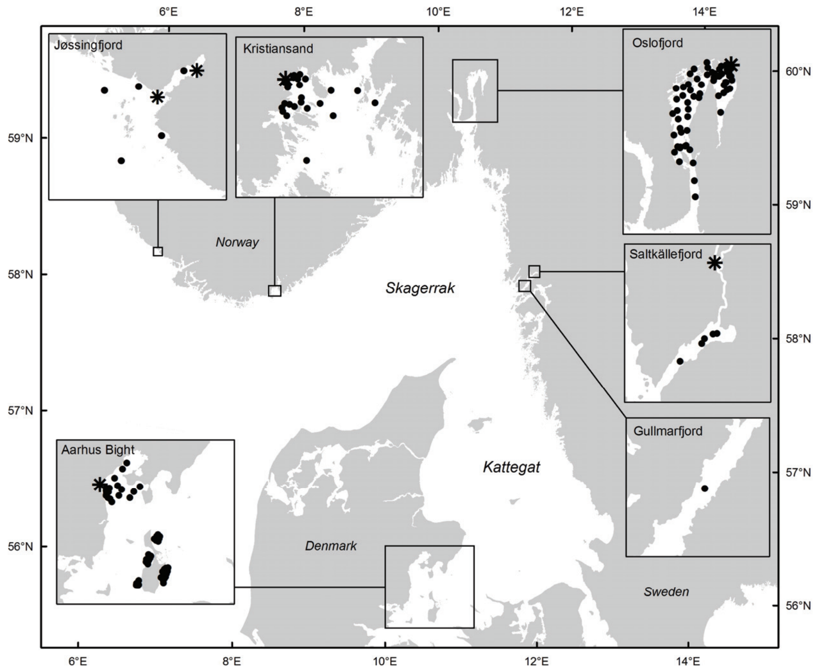

Figure 1b. Map showing investigated area with sites (dots) and location of tailings disposal site (asterisk) at Maarmorilik, West Greenland.

2.1.2 Heavy metal contamination gradient, Maarmorilik

The Maarmorilik gradient is situated in two inter-connected fjord arms, Affarlikassa and Quaamarujuk located in the inner part of the Uummannaq fjord system in West Greenland (Fig. 1b). The inner fjord arm has, in particular, received metal contaminated rock wastes and mine tailings from the “Black Angel Lead and Zink mine” (Fig.1b.) in the 1970s and 80s (Johansen et al., 2006). Being crushed rock, the gradient is likely a pure metal gradient with lead, zink and to some extent copper as dominating heavy metals. Tailings were disposed at 40 m depth in the inner fjord arm which had a maximum depth of 70 m. Macrobenthic species abundance and surface sediment metal concentrations where measured at 16 stations, in the tailings area and outside, before, during and 15 years after termination of tailings disposal activities, following a BACI sampling protocol (Johansen et al., 2006; Josefson et al., 2008) The resulting material used here was ca 150 paired faunal and sediment metal (Pb) samples.

2.1.3 Physical disturbance, Jøssingfjord

The Jøssingfjord area, Norway (Fig. 1a) has been used since 1960 for sea disposal of finely ground, inert tailings from a titanium mine. The discharged quantity was about 2 million tonnes/year in the 1980s. The submarine outfall was in 1984 relocated from the shallow Jøssingfjord to the deep basin Dyngadypet outside the fjord entrance. The relocation resulted in increased accumulation within 2–3 km from the new outfall. Tailings were detectable in the sediments as particles with high titanium content, i.e. 10–15% TiO_2 in tailings compared to a background level of 0.5–1%. Changes in the benthic fauna at varying distances from the old and new outfall sites have been followed, and biological impacts of mine tailings analysed (Olsgard and Hasle, 1993). A sedimentation of 4–5 cm of tailings per year at some locations resulted in changes in faunal composition, while at a rate of 1 mm per year, no impact was observed. This area is the most open and exposed area of the studied gradients.

2.1.4 Urban effluents gradient (PAH's, PCB, copper, ect.) in the Oslofjord

The Oslofjord penetrates inland over a distance of about 100 km from the open Skagerrak to the city of Oslo (Fig. 1a). Approximately 1 million people live in the area. The sediments in Oslo harbour and the innermost part of the fjord are contaminated as a result of industrial activities, boat traffic, urban road traffic, municipal wastewater, and small rivers draining from industrial areas. The fauna in the inner Oslofjord is affected by the urban effluent (Olsgard, 1995; Walday and Olsgard, 2004). Using PCB sediment concentrations as an indicator of urban effluents there is a clear gradient with high concentrations close to harbour of Oslo which are decreasing outwards in the fjord (Fig. 2). Sediment levels of copper, PAHs and other pollutants are high in the innermost part of the fjord and are correlated with PCB.

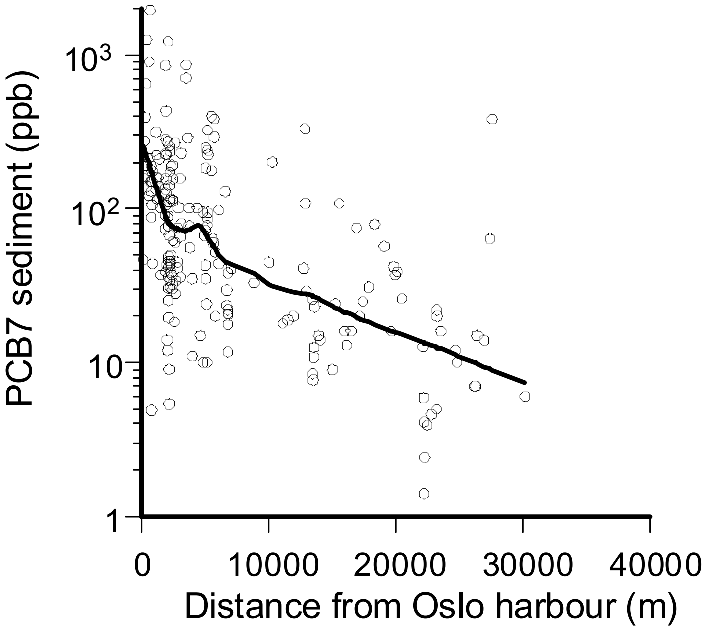

Figure 2. Distribution of sediment PCB/ concentrations with distance from Oslo harbour in the Oslofjord. Regression line obtained with LOWESS smoothing, tension=0.5.

2.1.5 Urban effluents gradient (nutrients, organic material and heavy metals) in a Danish bight

The Aarhus Bight on the east coast of Jutland (Denmark) (Fig. 1a) is connected to the Kattegat/Belt Sea area and has an average water residence time on the order of 2 weeks. The Bight receives effluents both from the city of Aarhus (ca 500 000 inhabitants) and the surrounding drainage area. The point of discharge is situated close to the harbour of Aarhus (Fig.1a). In addition to nutrients the effluent also contains organic material as well as various contaminants like heavy metals. Benthic macrofauna species and abundance and the sediment variables ignition loss and concentrations of *Clostridium* bacteria (CB) (a marker sewage effluents), were measured at 20 stations in the Aarhus Bight over the time period 1976 to 2005 (Josefson et al., 2006). Water depths at the stations varied between 12 and 17 m. The distances from investigated sites in the bight to the discharge point ranged from ca 1 km to 23 km. In addition, fauna samples were available from a reference area outside the Bight around the island of Samsø in the Belt Sea, some 50 km from the discharge point (Josefson et al., 2006). The total fauna data set with corresponding measurements of sediment variables was around 500 samples each corresponding to a bottom area of 0.1 m2. CB data indicated decreasing sewage influence from the discharge point with an inflexion point at ca. 11 km (Fig. 3).

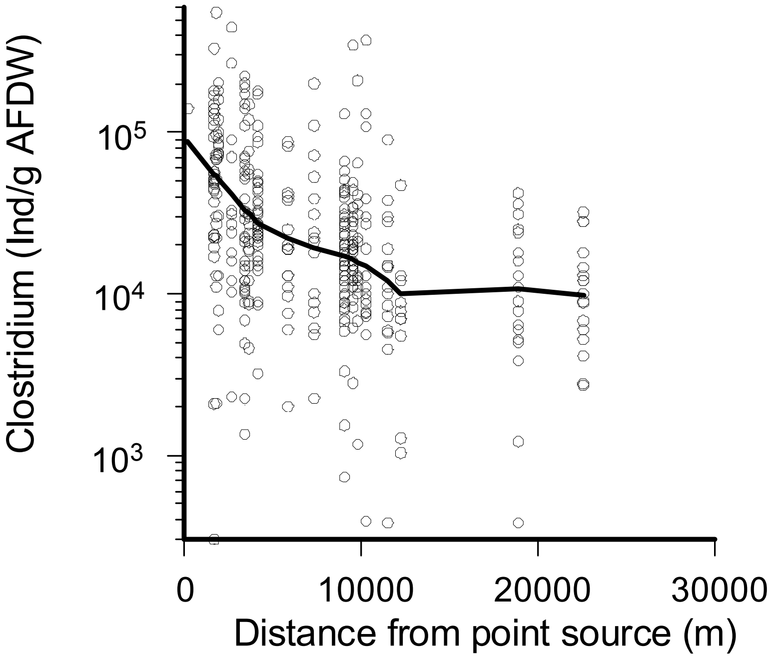

Figure 3. Plot of numbers of *Clostridium* bacteria in the surface sediments versus distance from point source in the Aarhus Bight. Solid line obtained from LOWESS smoothing, tension=0.5. Note the inflexion point at ca 11 km.

2.1.6 Temporal gradients after organic pollution from pulp mill effluents, Saltkällefjord

The benthic communities in Saltkällefjord (inner part of Gullmarfjord, Fig. 1a) were impacted by organic waste from a pulp mill up to 1966, when the factory closed down. The fauna was eliminated at the innermost stations and reduced in the inner half of the fjord. The recovery of the benthic fauna was followed in a number of studies (Rosenberg, 1972, 1973, 1976), and the communities had more or less completed the succession to "normal" conditions in the inner half of the fjord after eight years.

2.1.7 Temporal gradients following oxygen depletion events in a deep fjord on the Swedish west coast

The Gullmarfjord is open towards the west and the Skagerrak with a sill at the entrance of about 40 to 50 m depth and 118 m deep basin (Alsbäck) (Fig. 1a). The deep water in the fjord is generally exchanged once annually, but low oxygen concentrations occur in some years because of limited or no water exchange. The benthic community reduction and recovery at Alsbäck in relation to changing oxygen conditions have been studied by Josefson and Widbom (1988), Nilsson and Rosenberg (2000) and Rosenberg et al. (2002). The available material includes fauna and bottom water oxygen concentration time series over 31 years (1977–2007; Fig.4g).

2.2 Multi-metric indices

The three Nordic indices are multi-metric and share the same underlying rationale, that the quality of the faunal communities is reflected by both the contribution of sensitive and tolerant species and the species diversity. However, the relative weighting of the diversity and sensitivity components are different, and the indices uses two different classifications systems of the sensitivity of the community, and three different measures of species diversity. Very low numbers (i.e. < 10) of individuals in the samples are in all cases considered as a low quality feature and this is accounted for by using different factors to modify the indices against the number of individuals in the samples. All three indices are calculated on data from a sample size of the bottom of 0.1 m^2.

2.2.1 The Danish index

The Danish index, DKI described in Borja et al. (2007), uses the AMBI index where species are classified according to sensitivity/tolerance to disturbance (Borja et al., 2000), the Shannon diversity (H; Shannon and Weaver, 1963) and a factor compensating for both low species and density numbers. The value of AMBI is calculated from the proportions of individuals of sensitive/tolerant species within a sample. The diversity component (H’) and the sensitivity component (AMBI) are both normalised to attain a value between 0 and 1, where the diversity is normalised against the highest diversity observed in the sea area. The two components are weighted equally in the calculation of DKI, and the index-value is modified by the factor compensating for low species and density numbers:

$$DKI = \left(\frac{\left(1 - \frac{AMBI}{7}\right) + \left(\frac{H'}{H'_{\max}}\right)}{2} \right) \times \left(\frac{\left(1 - \frac{1}{N}\right) + \left(1 - \frac{1}{S}\right)}{2} \right)$$

where H´ the Shannon index with log base 2, and H_{max} the reference value which H is normalised against and is the highest value that H reaches in undisturbed conditions. H_{max} was set to 5.6 which was the highest value observed in this material, except for Maarmorilik where H_{max} was set to 4.0, N the number of individuals and S the number of species.

DKI can attain value between 0 and 1. For very high number of species the value of DKI approaches the mean of the normalised sensitivity and diversity components. If S=1 then $0 < DKI < 0.25$ and if $S = 1$ and $N = 1$ then $DKI = 0$

DKI scales linearly with the AMBI-value of the samples for all combina-tions of S and N

2.2.2 The Norwegian index

The Norwegian index, NQI, also uses the AMBI index (Borja et al., 2000) as a measure of sensitivity and is normalised to attain values between 0 and 1 in the same way as in DKI index. The diversity component is described by a factor, SN (Rygg, 2006) which is normalised to attain values between 0 and 1 and the diversity component is also modified by a factor to compensate for low densities. The normalised sensitivity and diversity components are weighted equally in NQI as described by:

$$NQI = 0.5 * \left(1 - \frac{\mathrm{AMBI}}{7}\right) + \left[0.5 * \frac{SN}{2.7} * \frac{N_{tot}}{N_{tot} + 5}\right]$$

where AMBI is the sensitivity component and SN = ln(S)/ln(ln(N)) is a diversity index. N is the number of individuals in the sample and S the number of species in the sample. The diversity component is normalised with a factor of 2.7, which is the maximum value of SN observed in the samples.

NQI can be calculated for values of $N > 1$ and $S < N$ and can attain val-ues between 0 and 1. However, for $S = N$ and $S < 4$ the SN is not a continuous function and calculation of NQI result in outlying index-values. Calculation of the diversity component is independent of the relative dominance (unevenness) of species in the sample.

2.2.3 The Swedish index

In the Swedish index, BQI (Rosenberg et al., 2004; Blomqvist et al., 2006; Anon., 2008), the sensitivity component is based on a classification of sensitive/tolerant species which is used to calculate a weighted average of the sensitivity of a species assemblage. The sensitivity value of 375 species was in this study objectively calculated from species-specific abun-dance distributions from a large Nordic data set based on 0.1 m^2 grab data from 800 stations.. Species richness is accounted for in BQI by a factor which scales with logarithm to the number of species and BQI is modified for low densities:

$$BQI = \left[\sum_{i=1}^{S_{classified}} \left(\frac{N_i}{N_{totalclassified}} * Sensitivityvalue_i \right) \right] * \log_{10}(S+1) * \left(\frac{N_{total}}{N_{total}+5} \right)$$

where $S_{classified}$ denoted the number of species classified. Sensitivity $value_i$ denotes the sensitivity of the ith species which ranges between 1 and 15. Low values indicate a high proportion of tolerant species, and high values indicate a high proportion of sensitive species (Rosenberg et al., 2004). S is the number of species (including not classified) per sample. N_i = the number of individuals of the i´th species, $N_{totalclassified}$ = the total number of classified individuals, N_{total} = the total number of individuals per 0.1 m^2.

BQI may be calculated for all values ofS, $S_{classified}$ and N. Theoretically, there is no upper limit for the BQI-value, which increases with the number of species. In practice, S in a sample is never expected to exceed 100. For $N = 1$ BQI ranges between 0.3 and 0.75. For $S = 1$ and high values of N then BQI ranges between 0.3 and 5 depending one the specific sensitivity value. For $S = 0$ then $BQI = 0$.

2.2.4 Method to determine a G/M border and statistical methods

The G/M border should ideally be determined in relation to some reference situation (Pollard and van de Bund, 2005). In practice, however, reference (i.e. = no or little human influence) situations are difficult to find, and some investigators have instead determined the G/M border from data coming from environments assumed to be in at least Good status (GIG NEA 8 for example, Carletti and Heiskanen draft). At least

Good status was then assessed when fauna data were on the less impacted side of the threshold values of impact (e.g. Birk and Hering, 2009). Here we adopt a somewhat similar approach and screen data from pressure gradients for data of *Good* and *High* status. We have graphically analysed where abrupt changes, or thresholds, occur along different pressure gradients in the benthic community structure estimated by the indices. Index values were regressed against the pressure variable using the smoothing method "Locally weighted least square regression" (LOWESS, Cleveland, 1979), which is an appropriate moving average method when smoothed data points not are equidistant, as often is the case here. For this purpose we used the LOWESS option in the statistical computer package SYSTAT ver. 10. The point on the regression line where rate of change shows a threshold is identified. Index values on the less impacted side of this point are considered to be in a at least "*Good*" ecological status according to the WFD. Regarding these index values as coming from a distribution of values in an environment in at least Good status, the G/M border is obtained by taking the 5th percentile of the data on the less impacted side of the threshold after bootstrapping. Bootstrapping (Efron and Tibshirani, 1993) is a powerful method to produce estimates of parameters taken from an unknown probability distribution, which is the case here. The 5th percentile was selected because that is a statistical praxis for one-sided tests. The proceedure was as follows: The sample of at least *Good* data was sampled randomly with replacement with the same number as the sample size. This was repeated 9999 times and finally the median 5th percentile and the 95% confidence limits around the median were estimated.

To assess if index values on the good side of the threshold points were different from values on the worse side of the thresholds, data were tested for difference with Kruskal-Wallis ANOVA with two groups.

3. Results

3.1 Identification of pressure thresholds

Relations between the three indices and pressures in the different gradients are shown in Figure 4 analysed with the smoothing method LOWESS.

Overall the figures showed distinct change of all three indices in each of the seven different gradients. In the gradients of metal contaminated sediments outside the Kristiansand Nickel smelter, the three indices all decreased markedly in relation to sediment concentrations of both nickel (Fig. 4a) and lead (the latter not shown in figure because patterns were similar to nickel). Inflexion points in the LOWESS regression lines occurred at Ni concentrations between 500 and 1500 mg/kg depending on index. For lead the threshold values were the same for all indices at around 200 mg/kg, and thus similar to the observed thresold in the Greenland lead gradient.

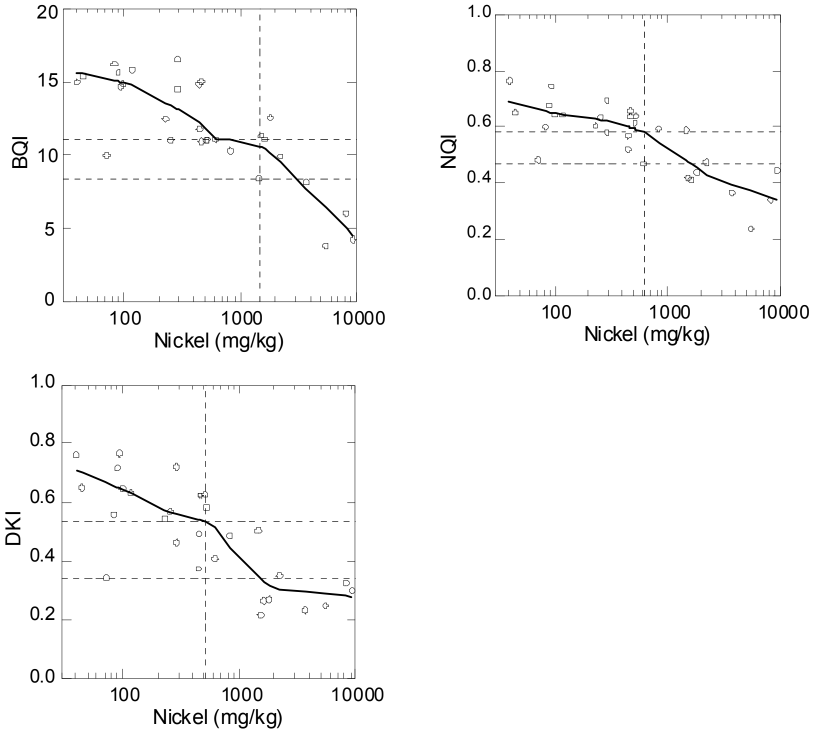

Figure 4 a. Plots of index values against sediment nickel concentrations in the Kristiansand gradient analysed with LOWESS smoothing non linear regression (solid curves, tension=0.5.Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the sediment lead gradient at the Lead and Zink mine at Maarmorilik, West Greenland all three indices showed dramatic declines in response to increasing lead concentrations (Fig. 4b). Similarly, the lead gradient in the Kristianssand Fjord (Southern Norway) showed inflexion points in the LOWESS curves around lead concentrations of 200 mg/kg.

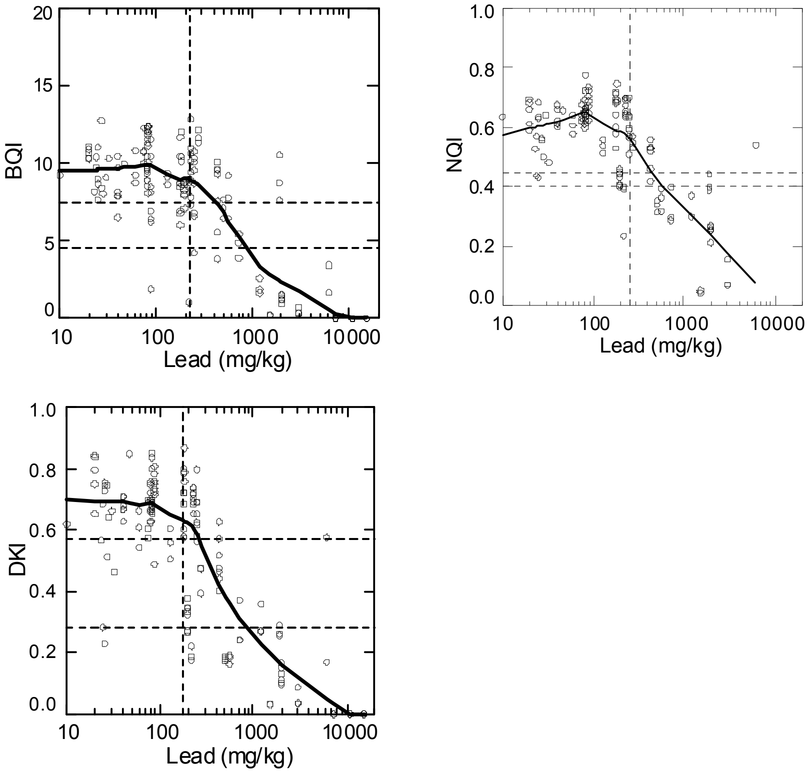

Figure 4 b. Plots of index values against sediment lead concentrations in the Maarmorilik gradient analysed with LOWESS smoothing non-linear regression (solid curves, tension=0.5).Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the spatial gradient of tailing deposition in the Jössingfjord area all three indices decreased in relation sediment concentration of TIO_2 which reflected the amount of tailings deposited (Fig. 4c).The main significance of its presence was considered as indicator of physical disturbance by tailing deposition. Although the changes in the indices along the gradient were small, there seemed to be a change in rates of decrease in the concentration interval 3.9–5.5% of sediment TIO_2.

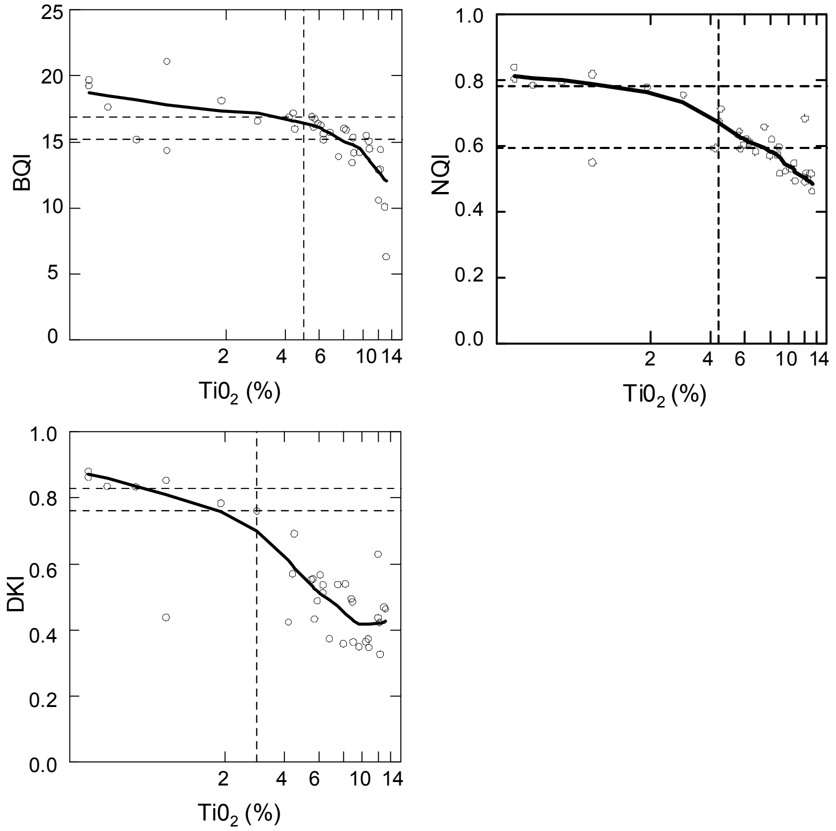

Figure 4 c. Plots of index values against sediment TiO2 in the Jössingfjord analysed with LOWESS smoothing non linear regression (solid curves, tension=0.5. Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the Oslofjord there was a spatial gradient in urban/industrial effluents with inputs of organics, heavy metals as well as polychlorinated hydrocarbons, PCB (Olsgard, 1995; Walday and Olsgard, 2004) with the main point sources located in the harbour area of Oslo (Fig. 2). Plotting the three indices against distance from the point source showed a clear increase of the values with increasing distance from the harbour (Fig. 4d). Inflexion points in the LOWESS curves, where the index values levelled off on a high level occurred around 20km from the harbour (Fig. 4d).

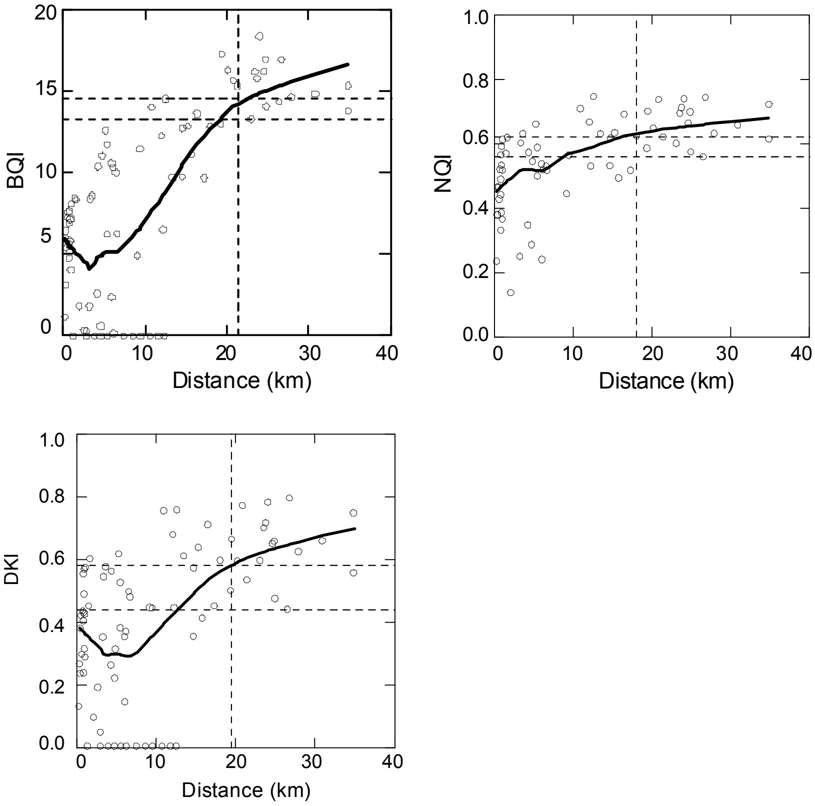

Figure 4 d. Plots of index values against distance from the hatbour in the Oslofjord impact gradient analysed with LOWESS smoothing non linear regression (solid curves, tension=0.5).Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the other urban/industial effluent gradient, the Aarhus Bight in Denmark, there was a clear change in sewage influence, measured by *Clostridium* bacteria concentrations in the sediment (Fig. 3), with increasing distance from sewage point source (Fig.4e). All three indices showed a consistent pattern; an increase and subsequent levelling off, with increasing distance from the sewage point source. The inflexion points occurred around 10 km from the harbour. Unlike the Oslo gradient, the water depth was fairly constant along the gradient, (15–17m) and mixing with water masses in the open part of the Belt Sea is probably more intensive due to the short residence time in the Aarhus Bight.

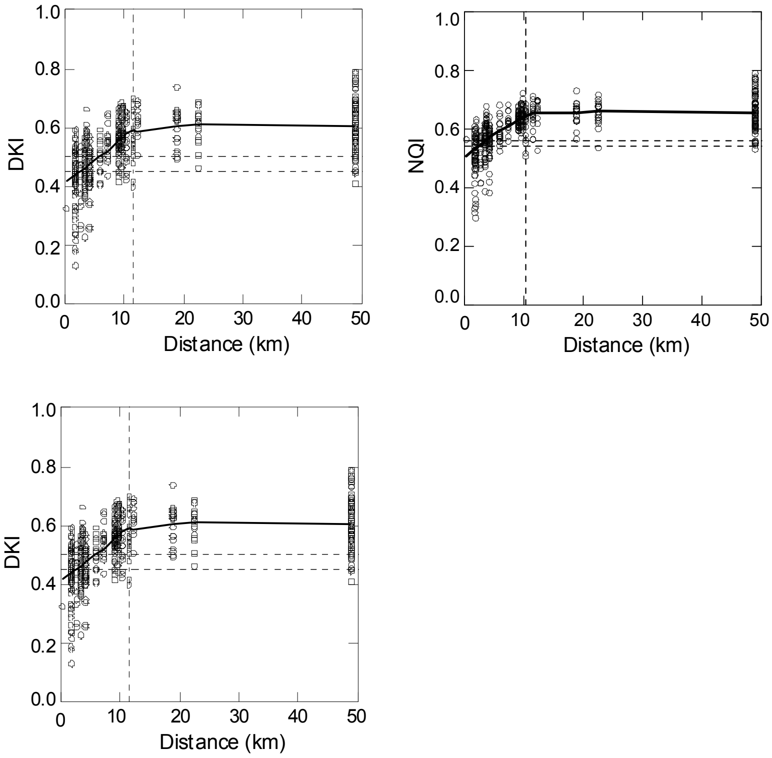

Figure 4 e. Plots of index values against distance from point source in the Aarhus Bight gradient analysed with LOWESS smoothing non linear regression (solid curves, tension=0.5.Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the Saltkällefjord, where the pressure factor was organic material from pulp mill effluents and the gradient is shown in the temporal dimension, the index values started at a very low level in the 1960s concomitant with the closure of the mill (Fig. 4f). At this point nearly azoic bottoms prevailed in the area close to the discharge. A recovery process continued until the mid 1970s when the indices stabilised and thereafter remained at a more constant level. Here the inflexion point, demarking the phase in the recovery process where the community begin to stabilise, was set to year 1977 for all three indices.

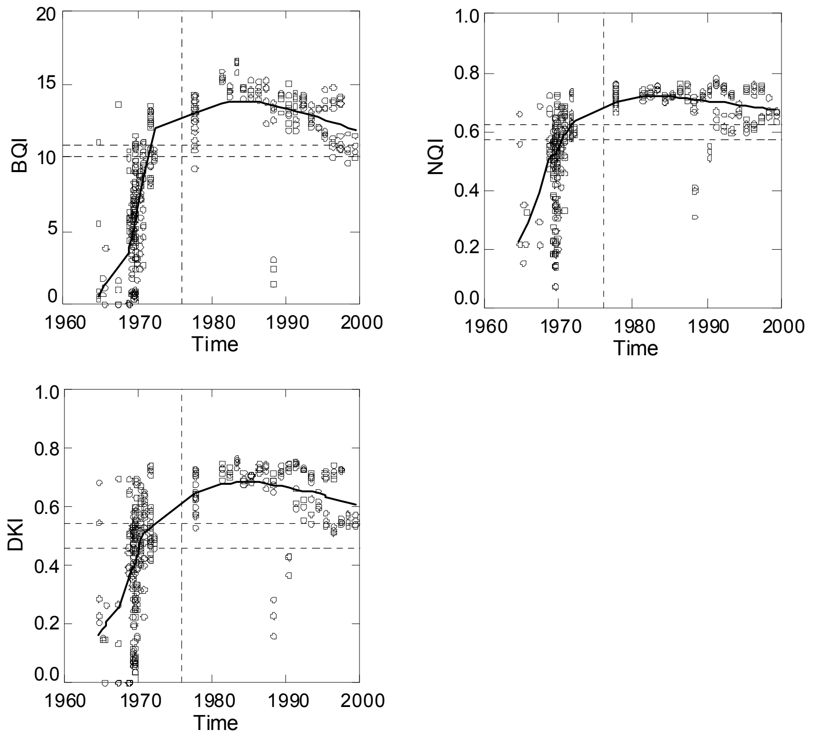

Figure 4 f. Plots of index values against time in the Saltkällefjord analysed with LOWESS smoothing non-linear regression (solid curves, tension=0.5). hreshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

In the Gullmarfjord deep basin (Figure 4g,h and i), repeated hypoxia with varying time intervals has precluded the straightforward approach used in the other gradients. Here we used two approaches to screen for at least *Good* data: 1. Selection of data from time periods with stable high index levels without occurrence of low oxygen values in the bottom water i.e 1977–80 and 1999–2007 (Fig. 4h).

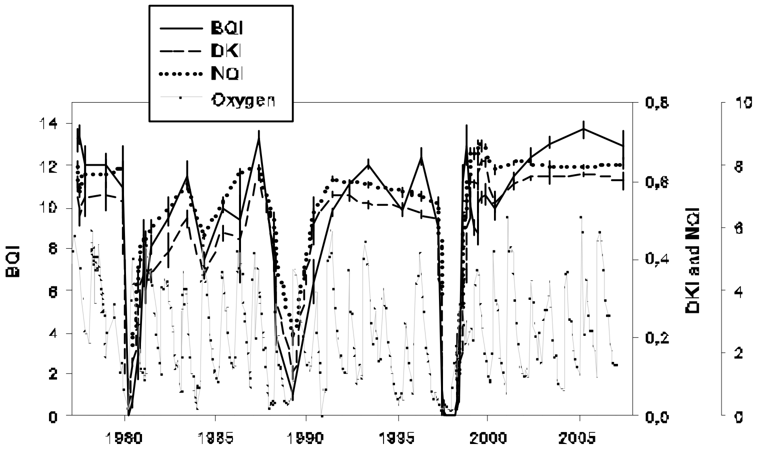

Figure 4 g. Long-term development of the three indices, BQI, NQI and DKI in the Gullmarfjord Basin. Superimposed is the development of oxygen concentrations in the near-bottom water.

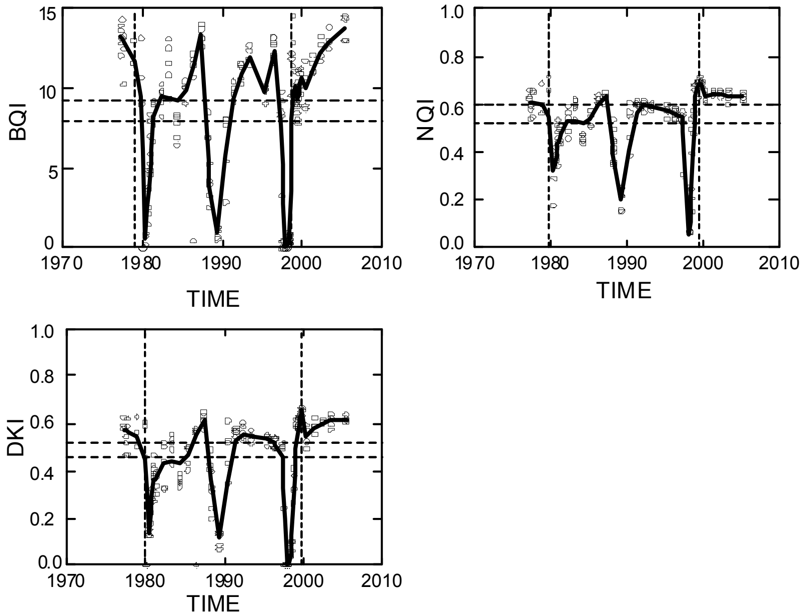

Figure 4 h. Plots of index values against time in the Gullmarfjord basin analysed with LOWESS smoothing non linear regression (solid curves, tension=0.2). Data points outside the vertical dashed lines were assumed to come from at least Good status conditions. Horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped values of these data.

2. Harmonisation of all index data against time elapsed from servere hypoxic conditions resulting in defaunation and thereby pooling the evolution of the recovery phases following each of the severe hypoxic events (Fig. 4i). All major events of defaunation followed after oxygen depletion events with oxygen concentrations below 0.5 mg/l over some weeks. The effect of the oxygen depletion event on the fauna community depends on the duration and due to the sampling frequence it was not always possible to determine the exact duration of the hypoxia. Therefore index values were plotted against elapsed time since oxygen was below 0.5 mg/l and an effect was observed in the fauna community during that particular event. This resulted in pooling of the succession sceemes of the three recovery events. In the latter approach LOWESS smoothing showed an inflexion point of the recovery phases at about 500 days ~ 2 years after the hypoxia event. In both approaches at least *Good* data were selected from the less impacted side of thresholds obtained with LOWESS smoothing (Figure 4h,i).

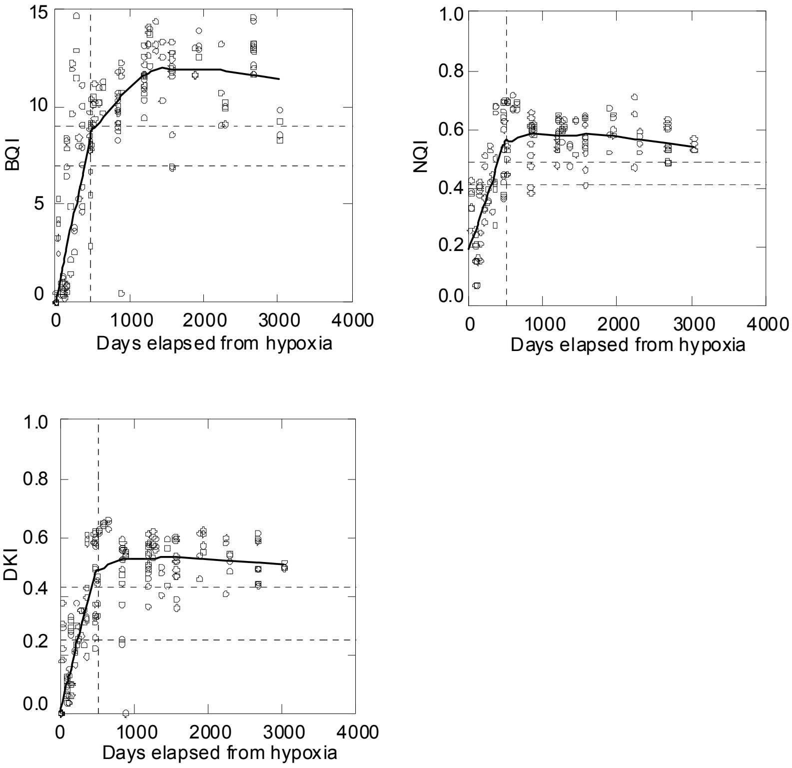

Figure 4 i. Plots of index values against time elapsed from defaunation due to hypoxia in the Gullmarfjord basin analysed with LOWESS smoothing non linear regression (solid curves, tension=0.5). Threshold value of pressure factor indicated by vertical dashed line, and horizontal dashed lines denote 95% confidence interval around the median 5th percentile (G/M border) of bootstrapped data on the less impacted side of the threshold.

To assess if index values on the less impacted side of the threshold points were different from values on the more impacted side of the thresholds, data were tested for statistical difference (Table 1). In all instances, both for indices and gradients, P was < 0.001 showing significant difference. The single exception was for the gradients in Kristiansand, with only some P values < 0.01. This shows that the data from at least *Good* status were significantly higher than data classified as worse status (*Moderate*, *Poor* and *Bad*).

Table 1 Statistical comparison of index values of at least *Good* EcoQS with values at worse status delimited by treshold values of pressure variables using Kruskal Wallis ANOVA. Median values for the two groups of data are shown.

Gradient/index	Median good	Median worse	n good	N worse	MWU	P
Kristiansand Ni / BQI	14.45	8.99	21	8	141	0.005
Kristiansand Ni / NQI	0.64	0.43	19	10	179	0.000
Kristiansand Ni / DKI	0.62	0.30	18	11	188	0.000
Kristiansand Pb / BQI	14.79	9.58	13	8	84	0.020
Kristiansand Pb / NQI	0.62	0.42	13	8	93	0.003
Kristiansand Pb / DKI	0.62	0.27	13	8	93	0.003
Maarmorilik Pb / BQI	9.37	4.17	94	55	4193	0.000
Maarmorilik Pb / NQI	0.64	0.36	101	36	3405	0.000
Maarmorilik Pb / DKI	0.68	0.27	72	77	4700	0.000
Jøssingfjord / BQI	17.19	15.12	11	24	228	0.001
Jøssingfjord / NQI	0.78	0.58	10	25	222	0.000
Jøssingfjord / DKI	0.83	0.47	8	27	201	0.000
Oslofjord / BQI	15.30	6.88	14	67	899	0.000
Oslofjord / NQI	0.66	0.53	19	48	797	0.000
Oslofjord / DKI	0.65	0.39	17	64	965	0.000
Aarhus Bight / BQI	8.91	7.02	169	317	42919	0.000
Aarhus Bight / NQI	0.65	0.59	169	311	41942	0.000
Aarhus Bight / DKI	0.61	0.51	151	335	40727	0.000
Saltkällefjord / BQI	13.57	6.19	128	234	28732	0.000
Saltkällefjord / NQI	0.72	0.55	128	222	26453	0.000
Saltkällefjord / DKI	0.70	0.46	128	234	27282	0.000
Gullmarfjord I / BQI	11.58	8.10	57	132	6084	0.000
Gullmarfjord I / NQI	0.65	0.54	44	125	4848	0.000
Gullmarfjord I / DKI	0.60	0.41	41	148	5298	0.000
Gullmarfjord II / BQI	11.40	3.29	107	82	8115	0.000
Gullmarfjord II / NQI	0.58	0.39	102	67	5710	0.000
Gullmarfjord II / DKI	0.54	0.23	103	86	7745	0.000

MWU = Mann Whitney U-statistic, n = number of values, P = likelihood of a true 0-hypothesis (no difference).

3.2 Comparison of indices

The three indices showed a high degree of covariation as they responded to the gradient of the various pressure factors in a very similar way (Figure 4a-i). Regressions between them (Fig. 5), although positive and highly significant, indicated differences between them. NQI and DKI were more similar than either of them compared to the BQI index.

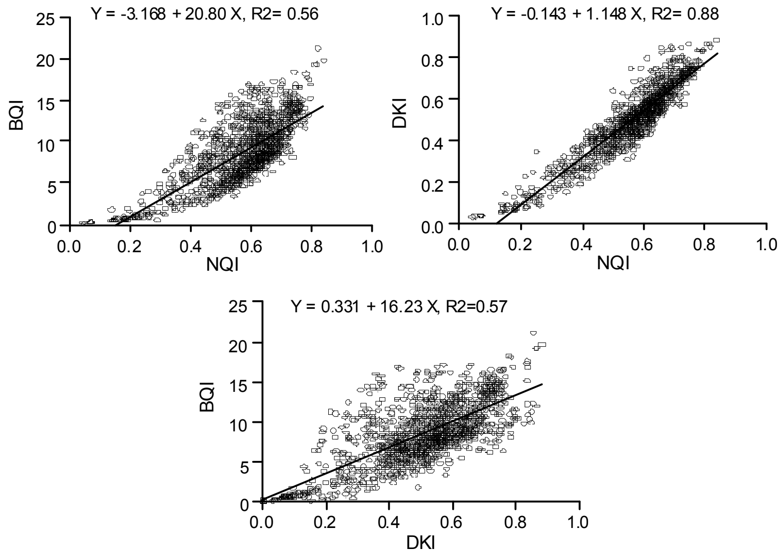

Figure 5. Plots of the three indices against each other with linear regression lines based on the total material (n~1400).

3.3. Estimating the G/M border

Based on the threshold values identified from the inflexion points, the index values in the *Good* end of the gradients were regarded as representing at least *Good* or *High* status. The lower border of these selected data should represent the *Good-Moderate* border according to the WFD. The estimated G/M borders, 5th percentiles of bootstrapped data, are shown in Table 2 and Figure 4 and 6 for each index and gradient. Apart from the gradient in physical disturbance in the open-coast area outside Jøssingfjord, which showed elevated values for all indices, most estimates of the G/M borders were similar between the gradients (Fig 6).

Table 2. Treshold limits of pressure factors and 5th percentile estimates of the *Good/Moderate* border based on bootstrapped data from the less impacted side of the treshold.

Gradient	Pressure proxy	Treshold value	Median G/M_{5perc}	Lower 95% CL	Upper 95% CL	Nr of values
BQI						
Kristiansand	Sed Pb	202	10.57	9.96	13.18	13
Kristiansand	Sed Ni	1462	9.96	8.35	10.98	21
Maarmorilik	Sed Pb	227	6.49	4.56	7.44	94
Jøssingfjord	Sed TiO2	5.0	15.53	15.16	16.90	11
Oslofjord	Distance m	21365	13.58	13.27	14.51	14
Aarhus Bight	Distance m	10290	5.96	5.27	6.49	169
Saltkällefjord	Year	1977	10.52	9.99	10.85	128
Gullmarfjord I			8.68	7.99	9.29	57
Gullmarfjord II	Days	489	8.26	6.96	8.99	107
NQI						
Kristiansand	Sed Pb	202	0.48	0.47	0.57	13
Kristiansand	Sed Ni	618	0.48	0.47	0.58	19
Maarmorilik	Sed Pb	247	0.41	0.40	0.45	101
Jøssingfjord	Sed TiO2	4.4	0.62	0.59	0.78	10
Oslofjord	Distance m	18016	0.57	0.56	0.62	19
Aarhus Bight	Distancem	10290	0.55	0.54	0.56	169
Saltkällefjord	Year	1977	0.61	0.57	0.62	128
Gullmarfjord I			0.55	0.52	0.60	44
Gullmarfjord II	Days	517	0.46	0.41	0.49	102
DKI						
Kristiansand	Sed Pb	202	0.36	0.34	0.48	13
Kristiansand	Sed Ni	524	0.37	0.34	0.53	18
Maarmorilik	Sed Pb	177	0.50	0.28	0.57	72
Jøssingfjord	Sed TiO2	2.9	0.77	0.76	0.83	8
Oslofjord	Distance m	19422	0.46	0.44	0.58	17
Aarhus Bight	Distance m	11515	0.46	0.45	0.50	151
Saltkällefjord	Year	1977	0.53	0.46	0.54	128
Gullmarfjord I			0.47	0.46	0.52	41
Gullmarfjord II	Days	517	0.36	0.25	0.43	103

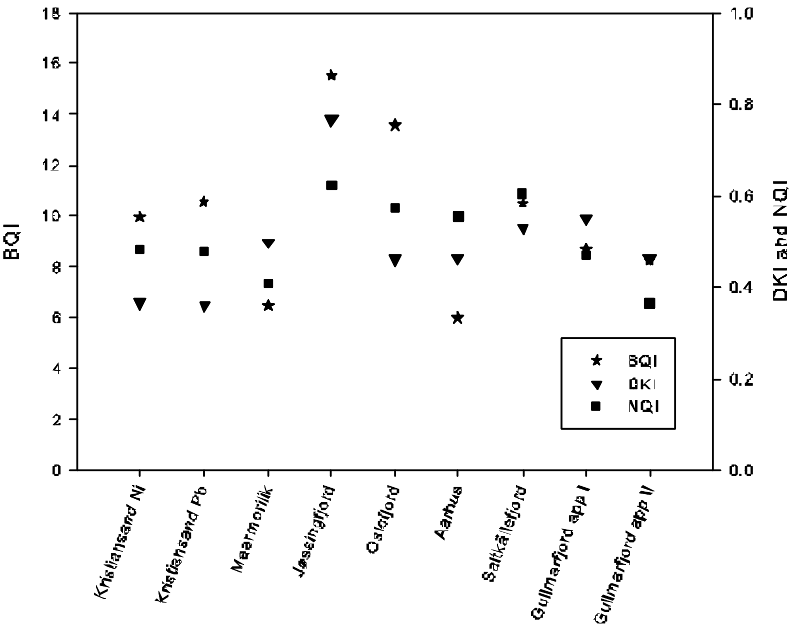

Figure 6. Estimates of *Good-Moderate* border values for the three Nordic indices in seven pressure gradients. Values are estimated 5th percentiles of at least Good data and Higher status.

4. Discussion

Since the end of the last century several marine scientists have presented different indices for the assessment of benthic habitat quality in European waters. Here we present comparative analyses of three indices from Denmark, Norway and Sweden applied on materials from coastal areas affected by various degrees and types of disturbance. The disturbances include combinations of organic enrichment, hypoxia, heavy metal contamination and physical disturbance by tailings deposition. Despite difference in impact factors, the response by the benthic fauna structure, measured by the indices, shows high similarities. Overall the values of the indices decrease with increasing values of the impact factors. Furthermore, it appears that the respons of indices along the various pressure gradients is typically non-linear. At some threshold value of the impact factor there is often a marked change in change rates of the indices, indicating an increased deterioration of the benthic fauna structure. When the pressure exceeds the resistance of the community, for instance considerable species loss will occur. Such threshold responses of benthic macrofauna has previously been described in relation to hypoxia (Diaz and Rosenberg, 1995; Levin and Gage, 1998), organic enrichment in general (Hyland et al., 2005), and sediment metal contamination (Burd, 2002; Josefson et al., 2008).

4.1. Comparison of the three indices

The multi-metric indices include different aspects of the benthic community. One basic element in the Norwegian and Danish indices, is the classification of species according to their sensitivity or degree of opportunism in relation to disturbance, the AMBI index. The AMBI classification method, based on the Pearson and Rosenberg (1978) classification in relation to organic pollution, has proven useful in describing also other impacts (Borja et al., 2003; Muxika et al., 2005). The Swedish index uses a different classification system of the species sensitivity (Rosenberg et al. 2004). The other basic element in the indices used is some measure of species diversity, a measure that has a quality value in itself. Diversity is quantified in terms of species richness in the Swedish index, BQI. SN is used in the Norwegian index and Shannon H in the Danish index. These differences are likely reflected by the variation in the regressions between indices where for instance the smaller variation between NQI and DKI probably is because both indices include the same AMBI index (Fig. 5). The differences between AMBI and the Swedish sensitivity component are illustrated by Figure 7, where the scatter around the overall positive relations is considerable.

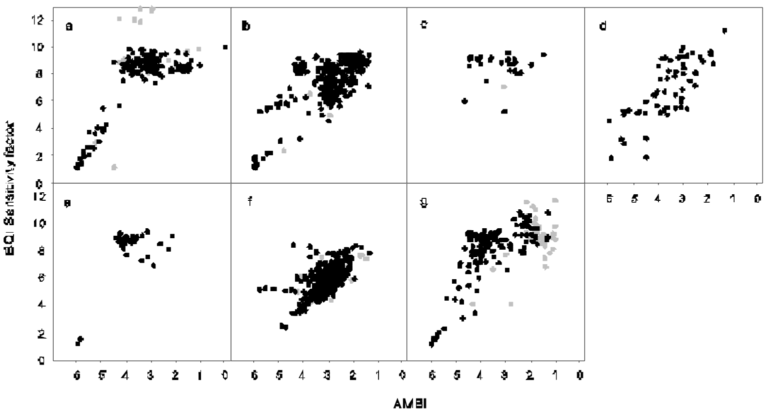

Figure 7. Plots of the two sensitivity factors used in the indices for each gradient. Gullmarfjord (a), Saltkällefjord (b), Kristiansand (c), Oslofjord (d), Jøssingfjord(e), Aarhus Bight (f) and Maarmorilik (g). Grey dots represents grabs where either BQI or AMBI have more than 20% of the individuals not assigned a sensitivity value.

However, despite these differences there was generally an excellent agreement between the three indices in reflecting the response to the pressure factors. This is clearly evident from the close agreement of the estimated threshold values for all gradients (Table 2). The use of different measures of species diversity has advantages and disadvantages and this affects the applicability of the indices to different enviroments. The DKI and NQI incorporate diversity measures normalised to the highest values of Shannon or SN diversity within that environment. This is an advantage if these reference values are known. However, such reference conditions may rarely exist. The use of SN-diversity requires species richness values of more than 4 species per sample. The use of species numbers in the Swedish method does not depend on either knowledge of a reference of species richness or a minimum number of species. The formulation in this index means that the index value, in theory, has no upper limit.

4.2. Methodological issues

The G/M border should ideally be determined in relation to some reference situation (Pollard and van de Bund, 2005). In this case some reference or high status values were defined and the values between the *High*/*Good* border and the lowest value divided into four classes where the G/M border was set between class 1 and 2 from the H/G border. In practice, however, reference (= no human influence) situations are difficult to find, and the G/M border is determined either arbitrarily or in relation to a situation as unaffected as possible (the best you can get). The latter option includes at least two ways. One is to use threshold values of pressure variables to delimit sites/environments that should be in at least *Good* EcoQS and use faunal data from these environments (Birk and Hering, 2009). Another option might be to deduce at least good data from the faunal response along pressure variable gradients. In the present study we explore the latter possibility further by looking for threshold change in different indices along different kinds of pressure/pollution gradients. The basic idea is that the integrity of an ecosystem to some degree can resist pressures or disturbance until a certain point where system changes occur and that this point mark the border between an acceptable status, i.e. Good status or higher and an unacceptable status, i.e Moderate or worse status, using the WFD terminology. This point should be reflected by a major change in structure and function of the system such as diversity or species composition, and consequently also in the indices. Therefore, we examined regressions between indices and pressure factors for points where change-rates were increased or decreased. Subsequently we regarded the values on the good side of the threshold points as being from an at least Good status situation. Some lower percentile of these data should be an estimate of the border between *Good* and *Moderate* EcoQS according the WFD.

4.3. Assessment of the border between Good and Moderate EcoQS/comparison with previous studies.

Previous studies in the North-East Atlantic area have arrived at index values of DKI between 0.53 and 0.63 (Borja et al., 2007; Carletti et al., draft) as estimates of the G/M border. However, the absolute values of the DKI index is dependent on the value of H_{max}, the maximum Shannon diversity in the actual environment. The value of H_{max} will vary between environments and therefore will depend somewhat on delineation of typology. In Borja et al. (2007) as well as in Carletti and Heiskanen (draft) H_{max} was set to 5.0, whereas for most of the materials in this study H_{max} was 5.6, the highest value found around Skagerrak/Kattegat. The effect on the estimates of the G/M border of DKI in this study of using H_{max}=5 instead of H_{max}=5.6 is an increase of the index value by 5–10%, depending on the area studied. The span of G/M values for DKI with H_{max}=5 is approximately 0.40–0.59 to be compared with 0.58 (Borja et al., 2007) and 0.53 (Carletti et al., draft), which suggests reasonable agreement between the studies. The G/M border for DKI should be somewhere in the interval 0.4–0.65. This is a rather broad interval and it may be questioned how good the definition is of “at least Good status”, because the H_{max} value refers to pristine conditions which is unknown because most all coastal environments are nowadays affected by human activities. However, the Greenland data may be an exception where we can exclude human impact in the pre-mining period in the Greenlandic fjord when truly pristine conditions were likely to have prevailed.

Previous studies in the North-East Atlantic area have arrived at an index value of NQI of 0.63 as an estimate of the G/M border, when using 2.7 as a maximum value (normalisation factor) for the diversity component SN. This approach is analogous to the use of H_{max} in the DKI index. A further normalisation factor of 0.78 (assumed reference value) for the whole NQI index is applied to obtain the EQR value (Carletti and Heiskanen, in press).

Rosenberg et al 2004 suggested two different G/M borders, one above and one below the halocline in the Kattegat/Skagerrak area of the North-East Atlantic. By separating the BQI-span into 5 equal classes they arrived at G/M borders of 10.8 above and 12 below the halocline. In a more recent approach based on assessment with the lower confidence limit of the median of all BQI values from a water body Anon (2008) arrived at G/M borders for the same area of 10.3 above and 12 below the halocline. These results agree well with the BQI G/M borders (between 6 and 15.6) suggested for the different gradients in this study.

Thus determination of G/M border from disturbance gradients is one way of deducing undisturbed condition when these are unknown, but this should ideally be done separately for different physical environments (water types according to the WFD). Yet, as seen in Figure 6, the estimates of the G/M borders in the Greenland material are not very different from the estimates in the more southern fjord systems.

5. Conclusions

- A new procedure is described to estimate the border between *Good* and *Moderate* EcoQS in benthic marine coastal environments.
- Benthic macrofaunal data from seven marine coastal areas with gradients of different pollution sources have been analysed with three Scandinavian methods to assess ecological benthic quality according to the WFD.
- The indices/methods responded in a similar way to the pressure factors, irrespective of the pressure being organic pollution or heavy metal contamination. Correlations between the three indices were generally high.
- Changes in the slopes of non-linear regressions between indices and pressure variables were used to identify thresholds where significant deterioration of the faunal communities started.
- Based on these threshold values, suggested to correspond to the point of change from *Good* to *Moderate* status, the *Good-Moderate* border for the indices was estimated.
- The estimated *Good*/*Moderate* borders showed similar values between different gradients and it is suggested that the border for the Norwegian and the Danish indices, except for the open coast gradient, should be in the interval 0.36 – 0.61 and the Swedish index 6 – 15.6.
- These former results compare well with results from other investigations in the North Sea area.

References

Anonymous, (2008). Naturvårdsverkets föreskrifter (2008:1) och allmänna råd om klassificering och miljökvalitetsnormer avseende ytvatten. http://www.naturvardsverket.se/Documents/foreskrifter/nfs2008/nfs_2008_01.pdf (in Swedish)

Bagge, P., (1969). Effects of pollution on estuarine ecosystems. I. Effects of effluents from wood-processing in-dustries on the hydrography, bottom and fauna of Saltkällefjord (W Swe-den). II. The succession of the bottom fauna communities in polluted estua-rine habitats in the Baltic-Skagerrak region. Merentutkimuslait. Julk. / Havsforskningsinst. Skr. 228, 3–118.

Berge, J., Bjerkeng, B., Næs, K., Oug, E., Ruus, A., (2007). Undersøkelse av miljøtilstanden i Kristiansandsfjorden 2006. Miljøgifter i sediment og organismer og sammensetning av bløtbunnsfauna. NIVA 5506.

Birk, S., Hering, D., (2009). A new procedure for comparing class boundaries of biological assessment methods: A case study from the Danube Basin. Ecological Indicators 9, 528–539.

Blomqvist M., Cederwall H., Leonardsson K., Rosenberg R. (2006). Bedömningsgrunder för kust och hav. Bentiska evertebrater 2006. Rapport till Naturvårdsverket 2006–03–21, 70 pp. (In Swedish with English summary).

Borja, A., Franco, J., Perez, V. (2000). A marine biotic index to establish the ecological quality of soft-bottom benthos within European Estuarine and Coastal Environments. Marine Pollution Bulletin 40, 1100–1114.

Borja, A., Muxika, I., Franco, J. (2003). The application of a marine Biotic Index to different impact sources affecting soft-bottom benthic communities along European coasts. Marine Pollution Bulletin 46, 835–845.

Borja, A., Josefson, A.B.,. Miles, A, Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G., Rygg, B. (2007). An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, ac-cording to the European Water Framework Directive. Marine Pollution Bulletin 55, 42–52.

Borja, A., Dauer, D.M., (2008). Assessing the environmental quality status in estuarine and coastal systems: Comparing methodologies and indices. Ecological Indicators 8, 331–337.

Burd, B., Macdonald, R., Boyd, J., (2000). Punctuated recovery of sediments and benthic infauna: a 19-year study of tailings deposition in a British Columbia fjord. Marine Environmental Research 49, 145–175.

Burd, B.J., (2002). Evaluation of mine tailings effects on a benthic marine infaunal community over 29 years. Marine Environmental Research 53, 481–519.

Carletti, A., Jowett, D.,Heiskanen, A.-S., draft. Water Framework Directive intercalibration technical report. Part III Coastal and Transitional waters. JRC Scientific and Technical Reports.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of American Statistical Association 74, 829–836.

Creutzberg, F., Wapenaar, P., Duineveld, G., Lopez, N. (1984). Distribution and density of the benthic fauna in the southern North Sea in relation to bottom characteristics and hydrographic conditions. RappP-v- Réun, ConsintExplorMer 183,101–110.

Diaz, R.J., Rosenberg, R. (1995). Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. Oceanography and Marine Biology: An Annual Review 33, 245–303.

Diaz, R.J., Rosenberg, R., 2008. Spreading dead zones and consequences for marine ecosystems. Science 321, 926–929.

Diaz, R.J., Solan, M., Valente, R.M. (2004). A review of approaches for classifying benthic habitat and evalu-ating habitat quality. Journal of Environmental Management 73, 165–181.

Hurlbert, S.N. (1971). The non-concept of species diversity. Ecology 53, 577–586.

Efron, B., Tibshirani, R.J. (1993). An introduction to the bootstrap. New York, Chapman & Hall

Fallesen, G. (1992). How sewage discharge, terrestrial run-off and oxygen deficiencies affect the bottom fauna in Århus Bay, Denmark. In: Colombo G, Ferrari I, Ceccherelli VU, Rossi R (eds) Marine eutrophication and population dynamics. Olsen & Olsen, Fredensborg, p 29–33.

Hyland, J., Balthis, L., Karakassis, I., Magni, P., Petrov, A., Shine, J., Vestergaard, O., Warwick, R. (2005). Organic carbon content of sediments as an indicator of stress in the marine benthos. Marine Ecology Progress Series 295, 91–103.

Johansen, P., Riget, F.F., Asmund, G., Josefson, A.B.,Hansen, J.L.S. (2006). Miljøundersøgelser ved Maarmorilik 2005. Danmarks Miljøundersøgelser. Faglig rapport fra DMU 605 (In Danish, elektronic)

Josefson, A.B., Widbom, B. (1988). Differential response of benthic macrofauna and meiofauna to hypoxia in the Gullmar Fjord basin. Marine Biology 100, 31 40.

Josefson, A., Hansen, J., Sejr, M.K., Sømod, B. (2006). Bundfauna i Århus Bugt 1976–2005 Økologisk status og udvikling. Teknisk Rapport til Aarhus Amt, Dec. 2006, 80 pp,(in Danish)

Josefson, A.B., Hansen, J..LS., Asmund, G., Johansen, P. (2008). Threshold response of benthic macro-fauna integrity to metal contamination in West Greenland. Marine Pollution Bulletin 56, 1265–1274.

Leppäkoski, E. (1975). Assessment of degree of pollution on the basis of macrozoobenthos in marine and brackish-water environments. Acta Academia Aboensis Ser. B 35, 1–90

Levin, L.A., Gage, J.D. (1998). Relationships between oxygen, organic matter and the diversity of bathyal macrofauna. Deep-Sea Research II 45, 129–163.

Lyche, A., Andersen, T., Brettum, P., Bækken, T., Bongard, T., Moy, F., Kroglund, T., Olsgard, F., Rygg, B., Oug, E. (2004). BIOKLASS – Klassifisering av økologisk status i norske vannforekomster: Forslag til aktuelle kriterier og foreløpige grenseverdier mellom god og moderat økologisk status for utvalgte elementer og påvirkninger. NIVA report 4860–2004. 63 pp.

Mirza, F.B., Gray, J. (1981). The fauna of benthic sediments from the organically enriched Oslofjord, Norway. Journal of Experimental Marine Biology and Ecology 54, 181–207.

Muxika, I., Borja, A., Bonne, W. (2005). The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Eco-logical indicators 5, 19–31.

Nilsson, H.C., Rosenberg, R. (2000). Succession in marine benthic habitats and fauna in response to oxygen deficiency: analysed by sediment profile imaging and by grab samples. Marine Ecology Progress Series 197, 139–149.

Olsgard, F., Hasle, J.R. (1993). Impact of waste from titanium mining on benthic fauna. Journal of Experimental Marine Biology and Ecology 172, 185–213.

Olsgard, F. (1995). Overvåking av forurensningssituasjonen i indre Oslofjord. Undersøkelse av bløtbunnsfauna 1993. Overvåkingsrapport 622\95 TA nr. 1258\1995.

Oug, E., Ruus, A., Håvardstun, J. (2004). Miljøtilstanden i Hanneviksbukta og Vesterhavn, Kristiansandsfjorden, før tildekking av forurensede bunnsedimenter. Bunnfauna og miljøgifter i organismer. NIVA 4915

Pearson, T.H., Duncan, G., Nuttall, J. (1982). The Loch Eil project: Population Fluctuations in the Macrobenthos. Journal of Experimental Marine Biology and Ecology 56, 305–321.

Pearson, T.H., Rosenberg, R. (1978). Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology: An Annual Review 16, 229–311.

Pinto, R., Patricio, J., Baeta, A., Fath, B.D., Neto, J.M. (2008). Review and evaluation of estuarine biotic indices to assess benthic conditions. Ecological Indicators 9, 1–25.

Pollard, P., van de Bund, W. (2005). Template for the development of a boundary setting protocol for the purposes of the intercalibration exercise.

Common Implementation Strategy - Working Group A ECOSTAT. 28 pp.

Puente, A. and Diaz, R. J. (2008). Is it possible to assess the ecological status of highly stressed natural estuarine environments using macroinvertebrates indices? Marine Pollution Bulletin 56: 1880–1889.

Rosenberg, R. (1972). Benthic faunal recovery in a Swedish fjord following the closure of a sulphite pulp mill. Oikos 23, 92–108.

Rosenberg, R. (1973). Succession in benthic macrofauna in a Swedish fjord subsequent to the closure of a sulphite pulp mill. Oikos 24, 244–258.

Rosenberg, R. (1976). Benthic faunal dynamics during succession following pollution abatement in a Swedish estuary. Oikos 27, 414–427.

Rosenberg, R., Agrenius, S., Hellman, B., Nilsson, H.C., Norling, K.(2002). Recovery of benthic habitats and fauna in a Swedish fjord following improved oxygen conditions. Marine Ecology Progress Series 234, 43–53.

Rosenberg, R., Blomqvist, M., Nilsson, C.H., Cederwall, H., Dimming, A. (2004). Marine quality assessment by use of benthic species-abundance distributions; a proposed new protocol within the European Union Water Framework Directive. Marine Polution Bulletin 49, 728–739.

Rygg, B. (1985). Effects of sediment copper on benthic fauna. Marine Ecology Progress Series 25, 83–89.

Rygg, B. (2002). Indicator species index for assessing benthic ecological quality in marine waters of Norway. NIVA report 4548–2002. 32 pp.

Rygg, B. (2006). Developing indices for quality-status classification of marine soft-bottom fauna in Norway. NIVA report 5208–2006. 33 pp.

Shannon, C.E., Weaver, W. (1963). The Mathematical Theory of Communication. University of Illinois Press, Urbana, 117 pp.

Swartz, R., Schults, D., Ditsworth, G., DeBen, W., Cole, F. (1985). Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall, Vol 865. American Society for testing and materials, Philadelphia

Van Hoey, G., Rees, H. L., and Vanden Berghe, E. (2007). A comparison of indicators reflecting the status of the North Sea. Pp. 116–127 In: ICES Cooperative Resaerch Report No. 288.

Walday, M., Olsgard, F. (2004). Ny senketunnel i Bjørvika: Biologiske førundersøkelser i november 2003. Bunnforhold kartlagt med ROV og bunnfauna med grabb. NIVA 4832.

Sammenfatning

I dette arbejde sammenligner vi tre Skandinaviske multimetriske indeks for bundfauna struktur for at vurdere marin bentisk økologisk status i syv forskellige forureningsgradienter. Indeksen måler på forskellige måder artsdiversitet og artssammensætning med hensyn til sensitivitet/tolerance ovenfor forstyrrelser. Presfaktorer i forureningsgradienterne omfattede organisk belastning, hypoxi, metalforureninger i sediment, forureninger i almindelighed fra byer, havne og industrier, og ren fysisk forstyrrelse fra tailing deposition. Indeksen svarede på nogenlunde den samme måde ovenfor ændringer af presfaktorer, dette uanset om faktoren var organisk forurening eller metalforurening. Indeksen viste i de forskellige gradienter mer eller mindre tydelige tærskel responser.

Grænsen mellem *God* og *Moderat* økologisk status ifølge Vandrammedirektivet (VRD), G/M grænsen, skal idealt bestemmes som en afvigelse fra en reference situation, d.v.s. fra en situation med minimal menneskelig påvirkning. Reference data er dog svære at finde. En alternativ procedure for at bestemme G/M grænser bliver derfor beskrevet i dette arbejde, en procedure der ikke kræver naturlige naturlige reference værdier. Tærskel værdier af presfaktorer, bortom hvilke bundfaunaens struktur begynder at markant forandres eller kollapse, blev identificeret fra ikke-lineare regressioner mellem indeksen og presfaktorerne. Indeks værdier fra før starten af struktur forandringen blev antaget komme fra et miljø i *God* eller *Høj* økologisk tilstand. Den nedre grænse af disse data, estimeret ved 5te percentilen af data efter boot-strapping, blev antaget være et estimat af G/M grænsen. G/M grænser for vert af de syv gradienter blev beregnet og sammenlignet med tidligere undersøgelser.