Benthic monitoring and sampling design and effort to detect spatial changes: a case study using data from offshore wind farm sites

Anita Franco^{1*}, Victor Quintino², Michael Elliott¹

1. Institute of Estuarine and Coastal Studies, University of Hull, UK; Mike.Elliott@hull.ac.uk

2. Department of Biology & CESAM, University of Aveiro, Portugal; victor.quintino@ua.pt

* Corr. Author:

A.Franco@hull.ac.uk; Tel. +44 (0)1482 466695; fax +44 (0)1482 466772

Fit-for-purpose benthic sampling at offshore wind farms

13 Abstract

14 The exploitation of renewable energies, in particular offshore wind farms (OWFs), is an expanding
15 sector which involves activities that may adversely affect the marine benthic ecology. Fit-for-purpose
16 monitoring is required with sufficient statistical power to detect ecologically meaningful changes,
17 but to date there have been no studies on the suitability of monitoring programmes applied to
18 OWFs. The theoretical relationship of sampling effort with precision in community estimates and
19 sensitivity of the analysis in detecting spatial changes was investigated, this latter assessed through
20 power analysis. Benthic community monitoring strategies and descriptors applied to UK OWFs were
21 used to interrogate real data variability in the marine environment. There was a general lack of
22 clarity in the survey rationale and hypotheses tested within OWF monitoring programmes hence a
23 lack of rigour in the survey design and statistical testing. Consequently the statistical properties of
24 monitoring strategies have been rarely assessed. Precision of mean estimates of benthic community
25 descriptors and the sensitivity to detect differences in the means increased with sampling effort. At
26 the average sampling effort applied in the OWF case studies (4 stations per impact type area and 3
27 replicates per station), the studies had a high probability of reliably detecting a $\geq 50\%$ change
28 between areas in mean benthic species richness (S; 5 species). Due to their higher variability than S,
29 more stations per impact type area were required to reliably detect a $\geq 50\%$ change between areas in
30 mean benthic abundance (N; 5 stations) and mean biomass (B; 10 stations). Higher sensitivity and
31 precision of estimates of S, N and B was achieved with transformation of data. Understanding the
32 general implications of monitoring design on the sensitivity of the detection of spatial changes is
33 important, particularly when monitoring effort has to be adjusted due to logistic and financial
34 constraints. Although there is no 'one-size-fits-all' approach to marine environmental data
35 acquisition, this study guides researchers, developers and regulators in optimising benthic
36 monitoring strategies at OWFs.

37 Keywords

38 Benthic community descriptors; Minimum detectable effect size; Precision of mean estimates;
39 Power analysis; Sampling effort

41 1. Introduction

42 All human activities in the marine environment have the potential to adversely affect the natural
43 system (Gray & Elliott, 2009). Renewable energy generating devices lessen the depletion of non-
44 renewable resources and have perceived lesser environmental effects (Gill, 2005). Offshore wind
45 generating capacity in particular is the most rapidly expanding sector of the renewable energy
46 industry (Wilson et al., 2010) and the UK is globally-leading this with as much capacity already
47 installed as the rest of the world combined (RenewableUK, 2014).

48 Offshore wind farms (OWF) produce 'green energy'. Their construction, operation and
49 decommissioning, however, may impact the composition and structure of benthic communities
50 through loss or change of habitat and physical disturbance of the seabed in ways that are difficult to
51 measure, minimize and mitigate (Gill, 2005; Wilson et al., 2010). Whether these effects constitute an
52 ecologically significant impact depends on their direction, duration, extent and magnitude, and on
53 the value and sensitivity of the receiving habitats and organisms (Boehlert & Gill, 2010; IEEM, 2010;
54 Wilson et al., 2010; Garel et al., 2014). Monitoring the condition of the benthos is a condition of the
55 operating license for an OWF. The developer has to prove that the OWF will not cause harm rather
56 than the regulator having to show that harm will occur (Gray & Elliott, 2009). Hence environmental
57 impact assessment (EIA) and response is urgently needed in the renewable energy sector (Inger et
58 al., 2009; Vaissière et al., 2014) despite there being large knowledge gaps (Garel et al., 2014).

59 The European Directive 2011/92/EU requires that an EIA is carried out for the consent of projects
60 having significant effects on the environment, including OWFs (CEFAS, 2004). In the resulting
61 Environmental Statement (ES), the main stressors and receptors should be identified and the
62 significance of potential impacts assessed. The consenting process should test impact hypotheses in
63 construction and operation and validate predictions (Judd, 2012). The existing guidance for
64 monitoring and assessment of potentially impacting activities in the marine environment, including
65 OWFs (CEFAS, 2004; Judd, 2012; IEEM, 2010), inevitably can only be generic rather than a highly
66 prescriptive methodology, largely because of site-specificity and the questions being asked
67 regarding habitat distribution, diversity and heterogeneity (CEFAS, 2004; Judd, 2012).

68 Environmental monitoring usually aims to investigate changes relative to a defined baseline
69 condition or set of parameters to quantify any impact. Changes are assessed before and after
70 construction, during construction vs. pre-construction, inside vs. outside the wind farm array, while
71 also accounting for temporal and spatial natural variability (Judd, 2012). Sampling programmes

should allow hypothesis-testing statistical techniques usually based on a Before-After Control-Impact (BACI) Paired-series approach or its modifications (Underwood, 1994; Ellis & Schneider, 1997). Whether the monitoring is aimed at assessing an impact or characterising spatial variability of baseline conditions, an adequate sampling effort is required (CEFAS, 2004; Judd, 2012) to quantify parameters with a certain level of precision and sufficient statistical power to detect the signal of change, minimise the risk of Type I and II errors and correctly reject the null hypothesis (Zar, 1999). Power analysis is capable of informing sampling design during the planning stage of a study (prospective power analysis; Cohen, 1988; Underwood & Chapman, 2003), and it can be applied also after the data have been collected and analysed to evaluate the adequacy of a specific design in detecting biologically meaningful patterns (retrospective power analysis; Andrew & Mapstone, 1987; Thomas, 1997).

CEFAS et al. (2010) recently reviewed UK FEPA (Food and Environment Protection Act 1985) OWF monitoring datasets to give preliminary recommendations on sampling adequacy, but to date there are no studies specifically appraising the suitability of monitoring programmes to detect variability in the status of the marine environment at OWF sites. The present study aims to integrate existing experience to guide suitable monitoring strategies of benthic communities at OWFs. Survey data and information from a selection of UK OWF monitoring studies were interrogated with the following objectives: (1) to review benthic monitoring strategies applied to OWFs in the light of existing guidance for EIA monitoring of benthic communities; (2) to assess the precision of mean estimates of benthic community descriptors in relation to the sampling effort at the station level, and (3) to apply power analysis in order to identify the overall most appropriate monitoring effort needed to detect spatial variability in benthic communities with a certain statistical power.

2. Materials and methods

2.1. Dataset, survey strategies and benthic variables

The UK offshore wind energy generating sector comprises several licensing phases co-ordinated by the Crown Estate (the landlord and owner of the seabed), with Round 1 launched in 2001, Round 2 in 2003 and Round 3 in 2010. Subtidal benthic survey data from a selection of Round 1 and Round 2 wind farms were compiled from ES and monitoring reports, the COWRIE (Collaborative Offshore Wind Research into the Environment) website (<http://www.offshorewindfarms.co.uk>) and also from individual developers (Table B.1).

OWF benthic sampling regimes were summarised using several parameters, including sampling method, number of surveyed stations and replicate samples collected per station. Non-parametric

analysis (Mann-Whitney U test) assessed differences between Round 1 and Round 2 OWF monitoring programmes.

Monitoring designs at the studied OWFs located sampling stations within and around development sites, often by distinguishing areas based on the expected distribution of impacts generated by the OWF. Criteria for station allocations to sampling areas were derived from the description of sampling regimes and survey maps as provided in the monitoring reports. According to these, stations were located within the OWF area and in some cases within the near-field area of the wind turbine foundations to determine scour effects. Stations were also often sited along the OWF cable corridor, around the development site within one tidal excursion from it (e.g., within the area affected by sediment transport and deposition; BOWind, 2007) or outside the tidal excursion to represent control areas. All these areas were classified in this study respectively as DS (development site), SA (scour assessment), CC (cable corridor), SI (secondary impact) and reference/control sites (RS). Survey strategies were reviewed in the light of existing guidance for monitoring benthic communities and for EIA of OWFs. Primary benthic community descriptors (mean species richness S , total benthic abundance N and biomass B) were derived from each dataset, depending on data availability.

2.2. Power analysis and precision assessment

Power analysis was employed to investigate the theoretical relationship between the sampling effort applied in monitoring designs and the size of the detectable change in mean S , N and B (Minimum Detectable Effect Size, MDES). The sample variances used in the power analyses were derived from ANOVAs on the benthic data collected at the studied OWF sites. By using data from a wide variety of case studies our findings apply as measures of central tendency for the group as a whole.

The ANOVA model for the OWF study design is a 2-level nested ANOVA (Sokal & Rohlf, 1995). It partitions the variance in the measured variables due to the main factor Area (factor A) and the nested factor Stations within Area (factor B(A)), and tests for differences among Areas. The null hypotheses tested in this study are no differences in the means of S , N and B among impact type areas (H_0 : DS SA CC SI RS). The minimum effect size (i.e., the difference between mean values of the analysed variable) that can be detected by the ANOVA was calculated as $MDES = \Phi^{-1}(P) \cdot (2 a s_Y^2)^{0.5}$ (Eqn 1; Ling & Cotter, 2003), where P is the power of the statistical test, Φ^{-1} is the inverse of the normal distribution function Φ , a is the number of areas (groups) compared in the analysis and s_Y^2 is the sample variance of group means. The term s_Y^2 was calculated for the 2 level nested ANOVA as the ratio between the mean square for the nested term ($MS_{B(A)}$) and the product of the number of stations per area (b) by the number of replicate samples per station (n)

(Ling & Cotter, 2003). After expressing $MS_{B(A)}$ as the ratio between the sum of squares for the nested term ($SS_{B(A)}$) and the associated degrees of freedom ($a(b-1)$), the resulting formula for the calculation of MDES was:

$$MDES = \Phi^{-1}(p) \cdot \sqrt{\frac{2 SS_{B(A)}}{n b (b-1)}} \text{ (Eqn 2).}$$

Equation 2 was used to identify the minimum effect size that could be detected with variable sampling effort (combinations of n and b , with b ranging from 2-100 and n ranging from 2-10) with an adequate level of power and significance. This was identified based on the five-eighty convention, whereby acceptable significance and power levels are set at 0.05 (i.e., 5% chance of making a Type I error) and 0.80 (i.e., 80% power or 20% chance of making a Type II error) respectively (Di Stefano, 2003). The values for $SS_{B(A)}$ were derived from the ANOVA tables obtained after applying the test to each dataset. Any decrease in the effect size detected by the test was considered as indicative of an increased sensitivity of the analysis.

Power analysis, as described above, assumes a balanced nested ANOVA design. When this condition was not fulfilled by the data, the design was balanced a posteriori by randomly selecting replicate stations (minimum 3 per area) with similar sampling replication per station at each case study. Before undertaking the analysis, the most appropriate transformation of the community variables (chosen between no transformation, square root and logarithmic) was applied independently to each case study in order to fulfil the assumptions of the applied parametric statistics (normality, homogeneity of variances and independence of errors; Sokal & Rohlf, 1995). Case studies where the data (transformed or not) did not fulfil the ANOVA assumptions were excluded from further analysis, whereas where multiple transformation options were effective, these were all taken into account.

The precision of mean S , N and B estimates (raw or transformed data) associated with sampling replication was calculated for each station (where replicate samples were available) and the overall mean precision was derived for each study (based on the balanced designs above). Precision was based on the ratio of the standard error to the mean (D ; Andrew & Mapstone, 1987) and was expressed as a percentage using the formula $100 \cdot (1-D)$. The relationship between precision and mean estimate was assessed by means of Spearman's correlation.

3. Results

3.1. Benthic survey strategies

Overall, data were obtained from 29 benthic monitoring programmes covering 19 OWF projects (Table B.1). Most programmes (20) were undertaken before construction, for environmental characterisation and baseline monitoring, whereas surveys undertaken during or post-construction respectively accounted for 2 and 7 studies. Macro-benthic data used here were obtained by either a Day or Hamon grab (the latter being used on coarser sediments), both of 0.1 m² sampling area, although ancillary sampling (using benthic dredges) was also undertaken in some cases (Table B.2). With few exceptions (SyS98 and SyS05), the same grab sampling method was used in pre- and post-construction monitoring at a single OWF site (Table B.2).

Most studies incorporated several benthic survey stations within and adjacent to the OWF, usually with several control stations located away from the wind farm. Round 1 OWFs generally had a smaller direct footprint and a smaller number of sampling stations compared to wider Round 2 OWFs (Tables B.1 and B.2), although the sampling density was higher in the former cases (1 to 6.5 stations per OWF area (km²); values mostly above 1.9) than in Round 2 projects (0.42 to 1.74 stations per OWF area; values mostly below 1.6). Differences between Round 1 and Round 2 OWFs were statistically significant for all these variables (Mann-Whitney U test $p < 0.001$).

Whilst most survey reports did not formally describe the type of sampling regime, the baseline (pre-construction) and following (during or post-construction) surveys generally had stations arranged as a systematic or random stratified pattern, with the strata including multiple impact areas (DS, SA, CC, SI) and control sites (RS). In some studies, stations were clearly and explicitly allocated to these areas (e.g., BOWind, 2007), including also a formal identification of control areas (e.g. HG04 included two control boxes with two replicated stations each; Table B.2). However, other survey reports provided only a generic explanation on station selection (e.g., stating that stations were selected with regard to the nature of the sediment in the area and the proposed position of the turbines, without any further information given in the text; Titan, 2002), and maps of the OWF boundaries and of the distribution of sampling stations had to be examined to allocate stations inside the development area, along the cable corridor or outside the OWF. In the absence of further information on the extent of the tidal excursion around the OWF, these latter stations could not be assigned into secondary impact and reference stations, and they were arbitrarily allocated to the SI category for the analysis in this study.

Individual stations typically had 3 replicates, except for stations in GS02 (2 replicates), SS02 (max. 2 replicate samples collected at only 4 stations) and SyS98 (single samples) (Table B.2). In some cases not all replicate samples were processed, hence did not contribute to the assessment. The survey designs used for Round 1 sites generally included replicate sampling at all or most of the stations, whereas the Round 2 sites almost exclusively used a higher number of single grab samples (to obtain wider spatial coverage) with a lower number of stations (11 to 30 stations per survey) with replicate sampling.

3.2. Sampling effort and detectable effect size

Two to 5 areas were surveyed in the OWF studies, with an overall average of 4 areas generally including DS, SI, CC and RS. When considering only stations with replicate samples, the mean number of stations per area (b) ranged 1 to 25 across the case studies, with also a marked variability observed among areas within a single case study (Table B.2). Sampling designs were therefore balanced and case studies LID01, SS02 and SyS98 were omitted from further analysis as no sufficient replication (within stations or areas) was available (Table B.2). The most appropriate transformation was applied to S, N and B data that allowed fulfilment of the ANOVA assumptions. As a result, the subsequent analyses (power and precision of estimates) were carried out on different subsets of case studies sharing similar data transformation and including between 2 and 13 case studies dependent on the variable considered (Table B.3). ANOVA was applied separately to S, N and B (raw or transformed data) in each of these case studies.

The MDES was calculated for variable combinations of n and b (sampling effort) by using sample variances derived from the ANOVA applied to individual case studies and the results were averaged across case studies (Fig. 1, Tables C.1, C.2 and C.3). The average MDES for each community descriptor decreased with increasing sampling effort (Fig. 1, Tables C.1, C.2, C.3 and C.4). For the purpose of interrogating the different community variables, the MDES calculated at the average sampling effort applied across the case studies (n=3 and b=4) has been considered below. With this sampling effort, the power analysis indicated a MDES between 18% and 113% (depending on the case study) for S (42% average across case studies, i.e. a detectable change of 5.3 species), whereas for N the MDES varied 49% to 68% (59% average, i.e. a change of 103.9 individuals 0.1m^{-2}). MDES could not be calculated for raw B, as these data did not fulfil the ANOVA assumptions. A higher sensitivity of the analysis was obtained with transformed data, with the average MDES decreasing to 35% and 28% for S and 50% and 61% for N, with square root and log transformation respectively. A higher mean MDES (98%) was associated with biomass data (after square root transformation) compared to the other community variables (Tables C.1, C.2 and C.3). The MDESs changed but the

patterns highlighted above were the same when considering examples of individual case studies where data from multiple variables and/or different transformations were available and suitable for the analysis (Table C.4).

Figure 1. Change in minimum detectable effect size (MDES) with sampling effort (n and b). MDES is expressed as percentage of the overall mean for the relevant community descriptor: (A) species richness (untransformed data), (B) total benthic abundance (untransformed data). The actual mean number of stations per area (b) in the case studies included in the analysis is shown by symbols below the horizontal axis. Power of the tests is set at 80% and significance level is 5%.

3.3. Sample mean estimates and precision

Sample means were estimated with a mean precision ranging 56% to 92% (values >75% in most studies) for S, whereas for N mean precision varied between 80% and 90% (Fig. 2). Data transformation generally resulted in increasing the precision of mean S and N estimates within single

case studies (Fig. 2). This pattern was evident for S also when considering the overall mean precision across the case studies, the increase in precision corresponding to a decrease in the variability of the transformed variables within stations (as expressed for example by the coefficient of variation of the data in Table C.1).

Figure 2. Precision of sample mean estimates: (A) species richness, S; (B) total benthic abundance, N; (C) total benthic biomass, B. Precision is shown for raw or transformed data, as per most appropriate dataset used in the analysis (see Table B.3). Mean $\pm$ SE for sample precision (bars and whiskers) and sample mean values (raw data, as points) are given for the different monitoring studies. Dataset ID codes are as in Appendix A.

4. Discussion

The reviewed documents on benthic monitoring at UK Round 1 and 2 OWFs emphasised that sampling was explicitly designed to assess potential impacts, with pre-construction monitoring

assessing baseline conditions and subsequent monitoring (Round 1 projects only) assessing changes in these conditions in response to the construction, presence and operation of OWF developments. Surveys were planned in consultation with regulatory authorities and generally fulfilled the existing guidance (particularly Boyd, 2002; Cefas, 2004; Ware & Kenny, 2011; Judd, 2012) and any recommendations within the licence required to construct and operate the OWF. Certain aspects of the benthic sampling were broadly standardised, e.g. the use of 0.1m² grabs is common practice for benthic monitoring, mostly dictated by a compromise between obtaining sufficient sediment and unwieldiness of the grab (Gray & Elliott, 2009), although other sample unit sizes may have been more cost effective (Ferraro and Cole, 1990). Other aspects of the survey design (e.g., number of sampling stations, spatial coverage) were case dependent, due to the variability of site-specific features of the development.

4.1. Limitations of applied designs for impact assessment

Lack of clarity with respect to the monitoring survey and sampling designs, objectives, and the hypotheses to be tested in OWF benthic surveys (CEFAS et al., 2010) hinder evaluation of their effectiveness (Judd, 2012).

Although the Before-After-Control-Impact (BACI) design is the most rigorous statistical design for detecting environmental impacts (Green, 1979; Underwood, 1994), it couldn't be applied in the studied OWF monitoring programmes due to the lack of before and after OWF construction data or clear distinctions between control and impact areas. Instead of multiple control areas, a key element of EIA survey designs (Underwood, 1994; Judd, 2012), the OWF survey designs investigated had reference stations located generally in the direction of the main tidal flow, sometimes with additional reference stations in other areas. The maximum distance of tidal excursion was generally used as a criterion to distinguish control and impact areas, as recommended by CEFAS (2004) and Judd (2012). However this information was not always given making it not possible to differentiate control from impact stations. Although it is not a regulatory requirement of EIA monitoring programmes, impact areas were further divided into subareas based on the expected distribution of impacts generated by the OWF. The location and extent of the turbine grid and cable corridors were generally used as criteria to identify near-field and far-field impact areas.

Seabed habitat type and its heterogeneity significantly affect natural variability in marine benthic communities (Gray & Elliott, 2009) therefore it should be factored into sampling designs (Boyd, 2002; CEFAS 2004; Ware & Kenny, 2011; Judd 2012). Furthermore, appropriate identification of control areas assumes this factor to be taken into account in the design, as controls are expected to represent the range of habitats occurring in the area of potential impact (Underwood, 1994). It is

assumed here that existing hydrographic and sedimentological data at the studied OWF sites were used to apportion sampling and place reference stations in broadly similar habitats to those inside the development site (Boyd, 2002), but this information was rarely included in the survey reports. Where geophysical surveys were used to inform on local habitat heterogeneity and further stratification of sampling stations by habitat was mentioned in the survey reports, no additional details were given on the strata or the station allocation (e.g., Titan, 2002; RPS, 2008). Consequently, this source of spatial variability could not be assessed, although its influence on the appropriateness of the design and its results cannot be excluded.

Sampling of the study areas was often achieved by variable replication at both area and station level, this variability occurring both between and within monitoring programmes, leading often to unbalanced survey designs, in contrast with recommendations for the application of parametric statistical analysis (Underwood, 1997). The density of stations was higher and sample replication more frequent in baseline studies undertaken in the smaller Round 1 OWFs compared to their Round 2 counterparts, where habitat characterization (through a higher number of stations with single grabs located in the wider area) was generally combined with the baseline monitoring (involving sample replication in a subset of stations). Morrissey et al. (1992) highlighted the importance of nested replicated sampling at various spatial scales in order to fully test for temporal or spatial patterns, and the sample replication within sampling stations is usually emphasised for environmental monitoring (Boyd, 2002; CEFAS, 2004). However, it is noted that favouring replication at the area level over that within stations is considered as the most efficient design, from both a statistical perspective (lowest variance and higher power) and when cost limitations influence the total number of observations allowed (Sokal & Rohlf 1995; Ling & Cotter, 2003). Where sample replication occurred, 3 replicates were almost exclusively used, as recommended in CEFAS (2004), although also five replicates are often used by regulatory bodies such as the Environment Agency and CEFAS (e.g., UK National Marine Monitoring Programme) as a compromise of effort vs. precision/cost effectiveness (e.g. Boyd, 2002). Overall, it was apparent that the choice of replication within the studied OWF monitoring programmes was mostly the result of a pragmatic decision for cost and effort as there was little or no use of pilot studies or power analysis to assess the necessary sampling effort required to detect a given degree of change (although such information might not have been included in the survey reports or ES examined).

4.2. Community estimate precision

The variability in quantitative estimates of the mean abundance or density of marine benthos is a major issue particularly if the variance of replicate samples is higher than may be expected from

random variation (Parsons et al., 1984). The importance of assessing precision of community estimates is related to the confidence in such estimates (Andrew & Mapstone, 1987).

The precision of mean community descriptors estimated at sampling stations at the analysed OWFs highly varied with the type of descriptor, its value at the study site and the transformation applied. The observed variability in the community parameters can be considered typical for benthic surveys (Elliott & O'Reilly, 1991), with precision decreasing from species richness to abundance to biomass estimates, reflecting the increase in the inherent variability of these descriptors (Tokeshi, 1993). That increasing variability reflects the control by the physico-chemical environment on the species richness compared to the control by biological interactions and recruitment, predation etc. on the abundance and biomass (Gray & Elliott, 2009).

A higher precision was associated when data were subject to transformation, this having an effect of reduction in the sample data variability and variance stabilisation (Zuur et al., 2009), as confirmed by the decrease in the coefficient of variability of transformed versus raw estimates. Again, the reduced precision in abundance and biomass estimates compared to species richness is also likely ascribed to the higher complexity of biological inter-relationships influencing the former variables compared to the predominant effect of number of available habitats and niches on the species richness (Gray & Elliott, 2009). Whilst there are no prescribed levels of precision recommended for benthic sampling, typically a standard error of 5% of the mean (hence a precision of 95%) would be considered satisfactory, although values between 10-25% (precision between 75-90%) have also been indicated as acceptable in ecological research (Southwood & Henderson, 2000). Precision seldom reached 95% in the analysed OWF studies (only in some cases with log-transformed abundance data or square root transformed species richness), but the use of 3 replicates as applied in most OWF case studies produced mean estimates with acceptable levels of precision between 75% and 90% for all community descriptors except for biomass. This agrees with the higher variability of biomass data, suggesting increased replication is needed to estimate this community parameter with a precision similar to the others.

4.3. Data analysis, power & sensitivity

Traditional hypothesis-testing statistical techniques (e.g. ANOVA) are widely used in marine studies. In the context of impact assessment at OWFs they could be employed to test differences in community or environmental attributes over space (between stations or areas) and time (e.g. pre- and post-construction). Despite this, rigorous statistical testing apparently has not been widely described within existing wind farm benthic monitoring reports, perhaps due to the lack of clarity in survey rationale and testing hypotheses. Instead, most studies applied hypothesis-generating

procedures in conjunction with expert judgement to summarise benthic community attributes and assess potential impacts. Where hypothesis-testing analysis was applied (e.g. Emu, 2007), statistical testing was largely confined to the Analysis of Similarities (ANOSIM), a non-parametric technique available within the PRIMER package. Such a technique is sufficiently powerful to detect relatively subtle differences between communities (Somerfield et al., 2002) and may analyse nested hierarchical designs, but, in contrast to other distance-based multivariate approaches (e.g. PERMANOVA), it does not allow testing interactions between factors (e.g. temporal (before/after) and spatial (control/impact)), this being essential to detect impacts within BACI-type designs (Underwood & Chapman, 2003). The 2-level nested ANOVA applied in this study also does not fulfil this requirement. Consequently, while our results inform OWF benthic monitoring as currently practiced, one should not assume that they are transferable to studies with more appropriate designs (e.g., BACI) for detecting impacts at OWFs.

Power analysis can be used to determine the adequate level of sampling effort that allows detection of a real effect (of a set size) with a required power and significance (Cohen, 1988; Green, 1989). However, despite this potential, power analysis remains a much under-used technique in ecology (Andrew and Mapstone, 1987; Green, 1989), possibly because it requires the knowledge of several parameters, including the variance estimate for the sample population, the statistical significance (α) and power ($1-\beta$), and the effect size to be detected. While variance can be estimated using the actual sample variance in existing studies (Thomas, 1997), and setting α at 0.05 is usually considered adequate (Underwood & Chapman, 2003), often there is no clear a priori indication of what is the adequate level of power. Furthermore, regulators are particularly imprecise regarding the level of change to be detected - it was particularly notable and of concern that surveys were designed without indicating the level of change to be detected.

The five-eighty convention, whereby acceptable power is set at 80% and statistical significance at 5% (Cohen, 1988), is widely applied in ecology, hence attributing a higher importance to the cost of making Type I error than Type II error. This approach, however, is not exempt from criticism (Di Stefano, 2003), as within EIA both errors might lead to undesired consequences (for the environment and the development such as marine renewable energy) and in some cases a higher cost of making a Type II error has been suggested (Peterman, 1990). An even higher uncertainty is associated with defining a priori what is the minimum change (e.g. in species richness) that is biologically relevant. Setting scientific and social thresholds of significance relative to the endangering of the marine environment is generally difficult (Köller et al., 2006). In OWF impact assessment, this may be further hampered by the lack of knowledge on the likely size and scale of

impacts on benthos following construction and operation activities. Guidance for sewage sludge disposal at sea is given whereby changes due to organic enrichment should not exceed 200% for benthic abundance and 50% for species richness and biomass (Jones et al., 1994), but it is uncertain if these levels could be applied to OWFs where the main effects are associated with structural habitat changes (Boehlert & Gill, 2010; Wilson et al., 2010). Rogers et al. (2008) considered a 10% change when assessing the power of benthic surveys on offshore soft sediments, although the basis for this more conservative threshold was not explained.

The average sampling effort applied in the OWF case studies (4 stations per impact type area, with 3 replicates per station) allowed detecting a change between areas of at least 50% of benthic species richness (5 species). In turn, a higher level of sampling effort is required to detect a similar relative change in benthic abundance (5 stations) and biomass (10 stations). In this study, as in Quintino et al. (2012), data transformations had a positive effect on sensitivity and precision. If Rogers et al.'s (2008) smaller effect threshold is considered, a much higher sampling effort would be required with 10 to 15 stations per area required to assess 10% changes in species richness, 15 to 20 stations for abundance and 20 to 50 stations for biomass, depending on data transformation. This reflects the high level of sampling effort highlighted by Rogers et al. (2008) for assessing offshore benthic communities (particularly abundance and biomass). Although the changes as assessed here cannot be explicitly associated with the effect of an impact, as the analysis was focused only on spatial variability within OWF areas and the full BACI design could not be assessed, it is emphasised that an impact removing 50% of the species in an area would be regarded as catastrophic and unacceptable by regulators and conservationists.

The analysis applied in this study provides a conceptual framework for standardised tests that can be applied to benthic monitoring data. Due to limitations in data availability, the habitat variability and temporal (including impact-induced) changes could not be factored into the analysis, but their inclusion in future monitoring designs is recommended to allow a more powerful impact assessment of OWFs on the marine environment. Wilson et al. (2010) also highlighted the need for more post-construction environmental auditing to allow checking the accuracy of predictions made during EIA and determining the nature and amount of monitoring required. Notwithstanding the above limitations, this study demonstrates a valuable tool for researchers, developers and regulators to understand the implications of sampling design and its modifications (often unavoidable due to logistic and financial constraints) for the sensitivity of the assessment of benthic communities at OWFs, so that confidence in the results can be attributed and the most appropriate response applied. However, there is the need for standardised protocols and clearer guidance on the

assumptions and requirements of monitoring programmes (e.g. the level of change to be detected), and this needs to be agreed between the scientific community and regulatory bodies. In particular, regulators and nature conservation managers have to be much clearer regarding the required degree of change required to be detected by the monitoring. This is particularly important when considering that larger scale developments are planned within UK Round 3, where various potential impacts ranging from point source to diffuse of unknown extent or severity may be compounded by other activities and in-combination effects. Further research is also needed to investigate the power of other community analysis statistical techniques especially, as shown by Gray & Elliott (2009), each of a large suite of methods gives further information. In particular, multivariate analyses are often considered more sensitive (hence powerful) in community changes detection than their univariate counterparts (Clarke & Warwick 2001; Somerfield et al., 2002; Sampaio et al., 2011; Quintino et al., 2012), but to date no theoretical statistical framework is available to formally quantify and compare the power of these tests.

Acknowledgments

The research was funded by the former Department for Business, Enterprise and Regulatory Reform (now Department of Energy and Climate Change). We acknowledge the previous input of Dr Jim Allen for his contribution to the study and especially the detailed comments of 2 anonymous referees.

References

- Andrew, N.L. & Mapstone, B.D. (1987) Sampling and the description of spatial pattern in marine ecology. *Oceanography and Marine Biology: an Annual Review* (eds H. Barnes & G. Barnes), Vol. 25. pp. 26-69. Aberdeen University Press.
- Boehlert, G.W. & Gill, A.B. (2010) Environmental and Ecological Effects of Ocean Renewable Energy Development: A Current Synthesis. *Oceanography*, 23, 68-81.
- Boyd, S.E. (2002) Guidelines for the conduct of benthic studies at aggregate dredging sites. CEFAS, Lowestoft.
- CEFAS (2004) Offshore Wind Farms: Guidance Note for Environmental Impact Assessment in Respect of FEPA and CPA Requirements. Prepared by the Centre for Environment, Fisheries and Aquaculture Science(CEFAS) on behalf of the Marine Consents and Environment Unit..

453 CEFAS, FERA & SMRU (2010) Strategic Review of Offshore Wind Farm Monitoring Data Associated
 454 with FEPA Licence Conditions. Report to Defra and MMO, Project Code ME1117.

455 Clarke, K.R. & Warwick, R.M. (2001) Changes in Marine Communities: an Approach to Statistical
 456 Analysis and Interpretation, 2nd edn. Primer-e, Plymouth, UK.

457 Cohen, J. (1988) Statistical power analysis for the behavioral sciences, 2nd edn. LEA, Hillsdale, New
 458 Jersey.

459 Di Stefano, J. (2003) How much power is enough? Against the development of an arbitrary
 460 convention for statistical power calculations. *Functional Ecology*, 17, 707-709.

461 Elliott, M. & O'Reilly, M.G. (1991) The variability and prediction of marine benthic community
 462 parameters. *Estuaries & Coasts: spatial and temporal intercomparisons* (eds M. Elliott & J-P.
 463 Ducrotoy), pp. 231-238. Olsen & Olsen, Fredensborg.

464 Ellis, J. & Schneider, D. (1997) Evaluation of a gradient sampling design for environmental impact
 465 assessment. *Environmental Monitoring and Assessment*, 48, 157-172.

466 Ferraro, S.P. & Cole, F.A. (1990) Taxonomic level and sample size sufficient for assessing pollution
 467 impacts on the Southern California Bight macrobenthos. *Marine Ecology Progress Series*, 67, 251-
 468 262.

469 Garel, E., Rey, C.C., Ferreira, Ó & van Koningsveld, M. (2014) Applicability of the “Frame of
 470 Reference” approach for environmental monitoring of offshore renewable energy projects. *Journal*
 471 *of Environmental Management*, 141, 16-28.

472 Gill, A.B. (2005) Offshore renewable energy: ecological implications of generating electricity in the
 473 coastal zone. *Journal of Applied Ecology*, 42, 605-615.

474 Gray, J.S. & Elliott, M. (2009) Ecology of marine sediments. From science to management, 2nd edn.
 475 Oxford University Press, Oxford.

476 Green, R.H. (1979) Sampling Design and Statistical Methods for Environmental Biologists. Wiley,
 477 Chichester.

478 Green, R.H. (1989) Power analysis and practical strategies for environmental monitoring.
 479 *Environmental Research*, 50, 195-205.

480 IEEM (2010) Guidelines for Ecological Impact Assessment in Britain and Ireland: Marine and Coastal.
 481 Institute of Ecology and Environmental Management.

482 Inger, R., Attrill, M.J., Bearhop, S., Broderick, A.C., James Grecian, W., Hodgson, D.J., Mills, C.,
 483 Sheehan, E., Votier, S.C., Witt, M.J. & Godley, B.J. (2009) Marine renewable energy: potential
 484 benefits to biodiversity? An urgent call for research. *Journal of Applied Ecology*, 46, 1145-1153.

485 Jones, J., Rowlett, S.M., Rees, H.L. & Portmann, J.E. (1994) Marine Pollution Monitoring
 486 Management Group. Sixth Report of the Group Co-ordinating Sea Disposal Monitoring. Aquatic
 487 Environment Monitoring Report, MAFF Direct. Fish. Res., Lowestoft.

488 Judd, A. (2012) Guidelines for data acquisition to support marine environmental assessments of
 489 offshore renewable energy projects. CEFAS, Lowestoft.

490 Köller, J., Köppel, J. & Peters, W. (2006) Offshore Wind Energy? Research on Environmental Impacts.
 491 Springer-Verlag Berlin and Heidelberg GmbH & Co. K., Berlin.

492 Ling, E.N. & Cotter, D. (2003). Statistical power in comparative aquaculture studies. *Aquaculture*,
 493 224, 159-168.

494 Morrissey, D.J., Howitt, L., Underwood, A.J. & Stark, J.S. (1992) Spatial variation in soft-sediment
 495 benthos. *Marine Ecology Progress Series*, 81, 197-204.

496 Parsons, T.R., Takahashi, M. & Hargrave, B. (1984) Biological oceanographic processes, 3rd edn.
 497 Pergamon Press, London.

498 Peterman, R.M. (1990) Statistical power analysis can improve fisheries research and management.
 499 *Canadian Journal of Fisheries and Aquatic Science*, 47, 2-15.

500 Quintino, V., Azevedo, A., Magalhães, L., Sampaio, L., Freitas, R., Rodrigues, A.M. & Elliott, M. (2012)
 501 Indices, multispecies and synthesis descriptors in benthic assessments: Intertidal organic enrichment
 502 from oyster farming. *Estuarine Coastal and Shelf Studies*, 110, 190-201.

503 RenewableUK (2014) Offshore Wind. <http://www.renewableuk.com/>, last accessed 14/05/2014.

504 Rogers, S.I., Somerfield, P.J., Schratzberger, M., Warwick, R., Maxwell, T.A.D. & Ellis, J.R. (2008)
 505 Sampling strategies to evaluate the status of offshore soft sediment assemblages. *Marine Pollution*
 506 *Bulletin*, 56, 880-894.

507 Sampaio, L., Rodrigues, A.M. & Quintino, V. (2011) Can biotic indices detect mild organic enrichment
 508 of the seafloor? *Ecological Indicators*, 11, 1235-1244.

509 Sokal, R.R. & Rohlf, F.J. (1995) Biometry: the principles and practice of statistics in biological
 510 research. W.H. Freeman, New York.

511 Somerfield, P.J., Clarke, K.R. & Olsgard, F. (2002) A comparison of the power of categorical test
 512 applied to community ecology data from gradient studies. *Journal of Animal Ecology*, 71, 581-593.

513 Southwood, T.R.E. & Henderson, P.A. (2000) *Ecological methods*, 3rd edn. Blackwood, Oxford.

514 Thomas, L. (1997) Retrospective Power Analysis. *Conservation Biology*, 11, 276-280.

515 Tokeshi, M. (1993) Species abundance patterns and community structure. *Advances in ecological*
 516 *research* (eds M. Begon & A.H. Fitter), Volume 24, pp. 112-186. Academic Press, London.

517 Underwood, A.J. (1994) On Beyond BACI: Sampling Designs that Might Reliably Detect
 518 Environmental Disturbances. *Ecological Applications*, 4, 4-15.

519 Underwood, A.J. (1997) *Experiments in ecology. The logical design and interpretation using analysis*
 520 *of variance*. Cambridge University Press, Cambridge.

521 Underwood, A.J. & Chapman, M.G. (2003) Power, precaution, Type II error and sampling design in
 522 assessment of environmental impacts. *Journal of Experimental Marine Biology and Ecology*, 296, 49-
 523 70.

524 Vaissière, A.-C., Levrel, H., Pioch, S. & Carlier, A. (2014) Biodiversity offsets for offshore wind farm
 525 projects: The current situation in Europe. *Marine Policy*, 48, 172-183.

526 Ware, S.J. & Kenny, A.J. (2011) *Guidelines for the Conduct of Benthic Studies at Marine Aggregate*
 527 *Extraction Sites*, 2nd edn. Marine Aggregate Levy Sustainability Fund (MALSF).

528 Wilson, J.C., Elliott, M., Cutts, N.D., Mander, L., Mendão, V., Perez-Dominguez, R. & Phelps, A.
 529 (2010). Coastal and Offshore Wind Energy Generation: Is It Environmentally Benign? *Energies*, 3,
 530 1383-1422.

531 Zar, J.H. (1999) *Biostatistical analysis*, 4th edn. Prentice-Hall, Upper Saddle River, NJ.

532 Zuur, A.F., Ieno, E.N. & Elphick, C.S. (2009) A protocol for data exploration to avoid common
 533 statistical problems. *Methods in Ecology and Evolution*, 1, 3-14.