Turning off the DRIP ('Data-rich, information-poor') - rationalising monitoring with a focus on marine renewable energy developments and the benthos

Wilding, Thomas; Gill, Andrew B.; Boon, Arjen; Sheehan, Emma; Dauvin, Jean-claude; Pezy, Jean-philippe; O?beirn, Francis; Janas, Urszula; Rostin, Liis; De Mesel, Ilse

Published in: Renewable and Sustainable Energy Reviews
Publication date: 2017
Publisher rights: © 2017 SAMS. Published by Elsevier Ltd.
The re-use license for this item is: CC BY-NC-ND
The Document Version you have downloaded here is: Publisher's PDF, also known as Version of record

The final published version is available direct from the publisher website at: 10.1016/j.rser.2017.03.013

Link to author version on UHI Research Database

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the UHI Research Database are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights:

1) Users may download and print one copy of any publication from the UHI Research Database for the purpose of private study or research.
2) You may not further distribute the material or use it for any profit-making activity or commercial gain
3) You may freely distribute the URL identifying the publication in the UHI Research Database

Take down policy
If you believe that this document breaches copyright please contact us at RO@uhi.ac.uk providing details; we will remove access to the work immediately and investigate your claim.

Download date: 02. Aug. 2019
Turning off the DRIP (‘Data-rich, information-poor’) – rationalising monitoring with a focus on marine renewable energy developments and the benthos

Thomas A. Wilding, Andrew B. Gill, Arjen Boon, Emma Sheehan, Jean–Claude Dauvin, Jean-Philippe Pezy, Francis O’Beirn, Urszula Janas, Liis Rostin, Ilse De Mesel

ABSTRACT

Marine renewable energy developments (MREDs) are rapidly expanding in size and number as society strives to maintain electricity generation whilst simultaneously reducing climate-change linked CO2 emissions. MREDs are part of an ongoing large-scale modification of coastal waters that also includes activities such as commercial fishing, shipping, aggregate extraction, aquaculture, dredging, spoil-dumping and oil and gas exploitation. It is increasingly accepted that developments, of any kind, should only proceed if they are ecologically sustainable and will not reduce current or future delivery of ecosystem services. The benthos underpins crucial marine ecosystem services yet, in relation to MREDs, is currently poorly monitored: current monitoring programmes are part of an ongoing large-scale modification of our seas and seabed that are occurring at local to regional scales that include commercial fishing, shipping, transport, aggregate extraction, aquaculture, dredging, spoil-dumping and hydrocarbon exploitation. The goods and ecosystem services already derived from predicted 3.6 GW of capacity to be installed within Europe by 2030 and potentially 188 GW (~50,000 wind-turbines) by 2050 illustrates the scale of development (European Ocean Energy Association 2010). MRE Developments (MREDS) are part of existing and ongoing modification of our seas and seabed that are occurring at local to regional scales that include commercial fishing, transport, aggregate extraction, aquaculture, dredging, spoil-dumping and hydrocarbon exploitation. The goods and ecosystem services already derived from

1. Introduction

Harnessing marine renewable energy resources (e.g. wind, wave, tidal stream) at commercial scales, involves deployment of multiple devices and associated infrastructure within a defined area of the coastal or offshore environment. Worldwide the Marine Renewable Energy (MRE) sector is growing significantly in line with technological development and the political imperative to act on climate change. The
The objectives of this paper are to:

- Highlight the need for clear management objectives in deriving a logical monitoring programme
- Review benthic metric selection and the need to set thresholds around which change can be logically assessed
- Consider relevant scales over which to assess change
- Review how and why ‘DRIPy’ data are currently generated and promote an alternative to null hypothesis significance testing as the basis for monitoring
- Propose a ‘traffic-light’ method for decision making in relation to thresholds
- Review current practices in assessing change in benthic environments
- Propose a structured approach to data-gathering to curtail the generation of DRIPy data
- Promote the re-direction of currently wasted resources to generating data which is useful to assessing relevant ecosystem-level change

Through our analyses, we look to promote a step change in the basis for environmental monitoring of MREDs that is applicable to any receptor group but we focus on the benthos because of its key and under-considered ecosystem role [19].

2. Monitoring MRED-benthic interactions

The need to include the benthos within ecosystem-led monitoring is fundamental [20,21]. The current problems associated with existing MRED-benthic monitoring programmes stem from a number of interlinked issues including the lack of clear management objectives (undefined scales, thresholds and metrics), fundamental problems with data gathering and interpretation and the technical/logistical problems associated with data collection.

2.1. Management objectives

Monitoring programmes, of any kind, can only be properly designed and successful where the objectives are clearly specified [22] and, in respect of MREDs, these are frequently lacking [14]. Without clear objectives, monitoring becomes a cataloguing exercise and, whilst this may have use in terms of ad-hoc future research programmes, such cataloguing is not cost-effective in development-specific studies on which to base management decisions. Objectives need to be clear and targeted [23].

In most nations, maritime activities are controlled via a number of regulatory measures that are specified at a range of political levels [24]. Within the EU the main drivers of monitoring in relation to the

---

1 In this context, ‘fish’ refers to all fish, and shellfish, with a focus on those of commercial and conservation importance including migratory / diadromous species such as eels and salmonoids.

2 The term ‘meaningful’, used here to mean a change that any stakeholder cares about, is deliberately subjective. The alternative ‘significance’ is not used because it should always be clarified as to whether the reference is to ‘statistical’, ‘ecological’ or other significance.

3 In this context ‘DRIPy data’ are characterised by being ‘data-rich yet information poor’.

4 See Meygen (http://www.meygen.com/the-project/) for an example of a commercial-scale tidal array in construction.
ecological effects of MREDS centre around the Birds Directive (2009/147/EC), the Habitats Directive (92/43/EEC) and the Environmental Impact Assessment Directive (including cumulative impact assessment) [7,17,25]. These directives are variously transposed into Member States’ (MS) national legislative framework and, consequently, there is limited common (EU) basis for monitoring around MREDS. Within Europe, the monitoring-emphasis is on iconic taxa such as marine mammals and birds or specially protected habitats (e.g. reef-forming species such as Sabellaria spinulosa (Leuckart, 1849)). Whilst Directives such as the Water Framework Directive [12] and Marine Strategy Framework Directive outline ecosystem-level criteria (e.g. ‘good-ecological status’ and ‘good environmental status’ respectively [26,27] these are not applied in relation to specific developments. However, the WFD can be applied to specific marine features, of indeterminate scale, which could encompass a specific renewable development (see §2.3). Whilst cumulative impact assessments are required for MREDS (e.g. as part of EIA in the EU) these are often lacking [28] largely because the underlying principles and definitions are poorly specified or understood [7,25,29]. Within the EU, the ecosystem-service-maintaining goals of Directives such as the MSFD have to be balanced against the ‘Blue Growth’ agenda (e.g. under the Integrated Maritime Plan [24] and within the context of marine spatial planning (under the Maritime Spatial Planning Directive) [30]. Across the EU there is a lack of strategic direction, both within and between Member States, in terms of understanding benthic-ecosystem-industry interactions [31] and this contrasts, at least in part, with other receptor groups.

Of the four commonly assessed receptor groups (mammals, birds, fish and benthos) the benthic-MRED interaction has been the least studied. For marine mammals and birds, management objectives are better articulated when compared with the benthos and the most logically justified monitoring objectives focus on the long-term sustainability of populations. Whilst there remain numerous data-gaps [32], many marine mammal and bird populations are relatively well understood in terms of their basic life-trait such as fecundity, age-to-maturity, natural mortality rates, spatial distribution and population mixing and boundaries [32,33]. For many species there exists extensive time-series data [34] and recently these have been augmented using data from tracking devices attached to individual animals [35]. Population sustainability in the face of developments, such as MREDS, has been assessed using models (e.g. Potential Biological Removal/Potential Consequences of Disturbance models and/or Population Viability Analyses) that are beginning to capture some of the complexity inherent in natural populations [6,14,32,33,35–37]. The issue of uncertainty, in assessing extinction risk, has also been addressed in relation to some marine mammals and birds, identified under the auspices of the International Union for Conservation of Nature [38].

Fisheries management objectives typically include maintaining stocks near their maximum sustainable yield and, again, many fisheries species are relatively well understood in terms of their fecundity and natural mortality [39] and likely non-linear relationships between stocks and some environmental drivers [40]. Historical fisheries data are available and give an indication of long-term trends in various populations (as a function of fishing intensity and environmental parameters) and modern genomics techniques have indicated the location and nature of some stock boundaries [39]. In addition, within Europe, many fish-stocks are statutorily monitored and afforded protection under the Common Fisheries Policy [24] which explicitly acknowledges the importance of defining spatial boundaries in respect of fishery management. In contrast, knowledge of benthic-ecosystem or benthic-population-scale processes is often lacking and/or not used in informing and justifying rational approaches to assessing MRED-benthic interactions. This is at odds with the ecosystem-approach to assessing change in relation to man’s activities in the oceans [8] and, within Europe, Directives such as the MSFD and the cumulative impact requirement under the EIA Directive.

MRED-benthos monitoring programmes around existing developments have tended to focus on three aspects: (i) macrobenthic species richness, abundance/biomass (via grab-sampling in sedimentary environments [14,17,41,42]), (ii) epibenthic cover/diversity (on hard substrata; e.g. [42–45] or (iii) epibenthic diversity as determined from trawls [42]. These three aspects are usually assessed at arbitrary spatial scales, are descriptive [42] and not linked to ecosystem-service provision. We believe that the most relevant benthic-related ecosystem services to be affected by changes in the benthos around MREDS will be related to fisheries [46], trophic networks (potentially affecting top-predators such as fish, birds and marine mammals) [32,35], sediment-based nutrient processing (biogeochemical reactor) and biodiversity [4].

2.2. Benthic metric selection and thresholds

The choice of metric (response variable) in benthic monitoring programmes is a major consideration and will determine the methodology employed, the appropriate spatial- and temporal-domain (§2.3) and the necessary confidence required in any assessment (§2.4.2). In general terms, whatever metric is chosen needs to be easily/cheaply measurable, sensitive to change and specific to the source of impact, predictable and anticipatory (able to signify impending change) [47]. In addition, the chosen metric should be relevant to assessing change that can be averted by management-actions, integrative (i.e. can indicate change over key gradients), have low variability (§2.5), be transferable (to various systems) and be clear and understandable [47]. We would add that metrics should be societally/stakeholder- and ecosystem-service-related. Hattam et al. [47] identified a wide-range of ecosystems services delivered by the Dogger Bank (North Sea), and proposed a number of metrics for the assessment of these services. These metrics including food provision (annual carbon production and monetary value), waste-treatment and assimilation (as measured by a healthy benthos in terms of biodiversity, §2.3.3), migratory and nursery habitat provision (as measured by the area of suitable habitat, quality of habitat, number of target species and occupancy rates) among many others.

Monitoring programmes that are based around a question such as ‘is there any change in X as a consequence of the development’ are limited (§2.4); monitoring should occur in relation to thresholds with a threshold being defined as a “target level or state based on the avoidance of unacceptable outcomes or an ecologically defined shift in system status” [48]. Identifying suitable metrics and negotiating threshold values, which meet the above criteria is difficult, potentially requiring multi-stakeholder consultation [49], and may constitute a significant part (in time and resources) of any monitoring programme [50,51]. The requirement for quantitative regulatory thresholds in respect of the benthos is, however, already part of the EU MSFD and WFD and EU member states have developed numerous benthic-diversity-based metrics [52–54] to classify the ecological status of sediments. The thresholds developed under the WFD offer some potential, at least conceptually, to MRED monitoring, but have been...
2.3. Relevant scales in relation to MRED-benthic interactions

Designating appropriate scales for monitoring programmes is a challenge particularly when assessing change at lower trophic levels where population boundaries are often poorly defined/understood or highly variable and where species are migratory and/or exhibit considerable ontogenetic changes in habitat utilisation (e.g. have a planktonic stage). What constitutes an appropriate scale to which inference should be made will, in part, be determined by the relevant regulatory framework: these may include locally agreed scales e.g. bays, inlets or specific features1 to regional, sea-scale multinational agreements (e.g. EU MSFD). From an ecosystem perspective geopolitical boundaries may be useful (e.g. MSFD Baltic and Celtic Sea sub-regions) but only where they coincide with eco/hydrologically defined boundaries that are relevant to the distribution of the species under investigation. In the UK, sea-areas have been defined by temperature, depth and current [64] and these with, potentially, any of the delineations as defined under the Marine Spatial Planning Initiative (a global initiative with ~20 participating nations [65]) may also be logical in terms of defining spatial limits to benthic monitoring.

2.3.1. Structure colonisation and scale of change

Many MREDs offer novel, surface piercing, hard and, at some scales, structurally complex, surfaces that may make a substantial relative contribution to this type of habitat over national or regional scales [66]. Where the new surface enables the local growth of large populations, particularly of ecosystem engineers (e.g. suspension feeders such as mussels [28]) or non-indigenous species then ecosystem-scale changes should be assessed. Mussels (in culture) have been shown to deplete overall plankton biomass [67], change planktonic communities [68,69] and enhance the reproductive success of their predators with potentially large-scale consequences [70]. Joschko et al. [71] predicted that a biomass of 265 kg (ash-free dry weight) of the blue mussel (Mytilus edulis L., 1758) would develop per wind turbine in the German Bight. When extrapolated across an 80-turbine array in the area, it was predicted that there would be a 10% increase in mussel biomass that would consume 1.4% of the annual primary production in that region [71]. Whether this, or similar cumulative effects from future planned developments, would constitute an ‘ecosystem-relevant’ change, within this spatial domain, depends on perspective and requires further research. MRED structures have also been associated with changes in jellyfish populations in the Baltic [72] by providing a hard-substratum for the assexual benthic phase (an effect also seen in respect of aquaculture structures in Thailand [73]) whilst artificial reefs have facilitated the spread of a harmful dinoflagellate, via its benthic stage, in large parts of the Gulf of Mexico [74]. Relevant scales for monitoring in these cases would be the individual structures (e.g. devices) but the results would need to be relevant within a region-scale context necessitating a profound understanding of the ecology of the species concerned and the receiving environment [56,58,75]. At certain scales changes in plankton (e.g. increased jelly-fish populations) may be a nuisance in relation to amenity use and/or damaging to other stakeholders (e.g. the aquaculture sector [76,77]) so any monitoring programme should link potential effects to other activities occurring within relevant spatial domains.

2.3.2. Non-indigenous species

MREDs may facilitate movement of non-indigenous and indigenous species into new habitats [2]. Movement of organisms between devices will occur in a number of ways including, for those with a benthopelagic life-history, the use of MRED devices as ‘stepping stones’ as

---

1 In any ecological assessment, change has to be compared with a baseline. When monitoring is initiated the observed baseline condition is often misunderstood to be the ‘natural’ state rather than the ‘current’ state (‘Shifting Baseline Syndrome’ [131]). Shifting, or unknown, baselines are particularly pertinent to coastal zones which have frequently been exposed to intense disturbance [132,133] and issues around baseline identification have caused problems in the implementation of the WFD [53,134]. Care should be taken to ensure that development objectives, of showing no ‘significant’ local-scale change, over baseline conditions, meet the longer-term, larger-scale objectives as part of ecosystem-scale assessment.

---

For example, St. Catherines’s Deep, an unusual channel of ~10 x 2 km km located on the south side of the Isle of Wight, UK and a potential tidal-development site. This locally unique feature might justifiably form its own spatial boundary within which monitoring should be focused.
seen in relation to oil and gas platforms [78] and which is dependent on the connectivity between the structures [45,79]. Other species may use construction and maintenance vessels on which to 'hitch-a-ride' [2]. This means that different deployment patterns /spatial distributions of devices, over regional scales, will have different impacts even whilst the overall number of devices remains constant. Location can play an important part in determining the prevalence of non-indigenous species (NIS), for example, the intertidal foundations of offshore windfarms have been associated with a much greater incidence of NIS compared with sub-tidal foundations in Belgium [45,80]. Assessing connectivity of colonising organisms between structures will require understanding planktonic larval ecology such as dispersive phase duration and larval behaviour, post-dispersal to adult habitat use [46] and hydrodynamics [79]. In relation to NIS the relevant scale for study could be by national sector (e.g. the Dutch North Sea) or entire regions (e.g. entire North Sea under the MSFD) depending on perspectives and spatially-defined objectives.

2.3.3. Biodiversity

Proponents of MREDs frequently cite that they will be associated with increased biodiversity [4,81]. Biodiversity is a complex, multifaceted phenomenon [82] and high biodiversity is often associated with increased system resilience to stressors ranging from physical disturbance to challenge by non-indigenous species [13,82,83]. High levels of biodiversity are, consequently, considered beneficial and actions to conserve/ enhance biodiversity are required under various international conventions (e.g. Aichi Biodiversity targets 2011–2020 under the Convention on Biological Diversity [84]). It is crucial that any assessment of biodiversity is understood to be scale-dependent: when placed in sedimentary environments the addition of hard substrata will inevitably enhance localised beta diversity (rate of change in diversity) because the structures invariably host different biological communities compared with the receiving environment. Macrobenthic alpha diversity (number of species per sample) may differ in proximity to hard-substrata mediated via structure-derived enhanced organic flux [85,86], changes in sediment stability [4] or changes in sediment attributable to colonising organisms such as mussels [87]. However, whether (how or why) increased alpha or beta diversity necessarily constitutes an enhancement to the system, on the scale of one or more ‘non-natural’ structures, has not been sufficiently explored and there is growing concern about ‘ocean sprawl’ [88,89]. There are currently no guidelines as to what levels (i.e. thresholds) of biodiversity are desirable, which assemblages (e.g. micro-, meio-, macro- and/or megabenthos) should be considered and over what scale (in space and time) it should be assessed [82]. In terms of regulatory compliance, the goal is usually to demonstrate an increasing biodiversity trend though the scales (time and space) over which this should be demonstrated and the underlying rationale are unknown [82]. Even a routine, explicit acknowledgement that appropriate scales for biodiversity-monitoring are ‘known-unknowns’ would constitute a step forward. The basis of metric choice, and thresholds, or even the acknowledgement that these cannot be determined/agreed, should be part of the environmental assessment in order to improve clarity and transparency of monitoring programmes (e.g. as required under the Marine (Scotland) Act, 2010).

2.3.4. The impact of fishing exclusion and assessing recovery

The extent to which MREDs exclude other activities (e.g. fishing) will, in part, determine their large-scale impacts on a range of receptors particularly fisheries and biodiversity [4,90]. Towed-gear fishing is banned around some operational wind-farms (e.g. in the Belgian and Dutch sectors) but not others (e.g. some in the UK sector) whilst tidal and wave devices are likely to be avoided by fishermen because of the entanglement risk. Where fishing effort exclusion occurs it may allow recovery of the benthos (with spill-over effects from increased local production extending an unknown distance beyond the MRED’s boundary [91]). This ‘protective’ effect [81] may allow benthic recovery of reef-forming ecosystem-engineers, such as reef-forming polychaetes [92] which will have implications for local biodiversity [4] and potentially regional-scale effects if acting as a nursery ground (e.g. for fish including commercial species) or a substantial source of propagules. The appropriate spatial- and temporal-scale for monitoring the degree of benthic recovery following fishery-exclusion will depend on the objectives of the monitoring programme and the sampling effort employed. Surveys of > 10 years might be inadequate to quantify recovery of some species, particularly where they are depleted because of overfishing (as by-catch) [93]. Appropriate timescales for assessing change/recovery in relation to fishery-effort cessation will vary between sites: > 10 years exclusion have resulted in a measurable difference in the North Sea benthos [94] but in other cases (e.g. a closure of 5 years) they have not [95]. There is considerable overlap between the monitoring objectives in relation to assessing the ‘efficacy’ of marine protected areas [96] and those to assess the ‘impacts’ of MREDs [6,97].

The identification of spatially/temporally delimited metrics and thresholds, in line with the overall management objectives, are the critical components to a logically based monitoring programme. However, many current monitoring programmes are deficient in this regard yet, based on null hypothesis significance testing, they still proceed [17,25][82.4].

2.4. Fundamental issues around data gathering and interpretation

Currently, most benthic monitoring effort focusses on detecting whether local impacts (changes) around MREDs are statistically significant or not at arbitrary spatial and temporal scales [17] and this results in DRIPy data. This approach, centred around null hypothesis significance testing (NHST), is not appropriate within the context of adaptive management as such null hypotheses are inevitably false and, in this context, failure to reject the null hypothesis merely reflects the lack of statistical power of the sampling design employed [98–100]. NHST as the mainstay of statistical analysis for the ecological/biological (and other) sciences stems from the unfortunate conflation of two different statistical approaches (Fisher v. Neyman and Pearson [101]) and has resulted in a ‘success’ being associated with a rejection of the null hypotheses. The deficiencies of NHST were acknowledged by Fisher, Neyman and Pearson in the 1950s in relation to their own fields [101] but may have persisted in scientific endeavour because it enables an attractively simple ‘statistical ritual’ to be adopted by non-statisticians (i.e. most scientists and regulators) [100]. In the context of monitoring the null-hypothesis is usually one of there being no-change in relation to a development [100]. This approach is particularly dangerous where ‘no evidence of impact’ (i.e. null hypothesis acceptance) is interpreted as ‘evidence of no impact’ [101]. There is a good argument that the obligation of proof should be reversed where environmental monitoring is concerned or, at least, that every environmental monitoring programme should fully investigate the power of the observational programme to detect threshold exceedance.

2.4.1. Null hypothesis significance testing and Type I and II errors

A common misunderstanding in relation to NHST is that an apparently simple question ‘is there an impact?’ can be unequivocally answered based around a P-value [10] threshold (usually 0.05) and that,
consequently, a threshold does not need to be specified. This is incorrect; justified thresholds are required (§2.2) around which to base effective management decision-making.

Accepting the logical requirement for a threshold raises two questions: first the required confidence around any assessment of threshold-exceedance that is made and, second, what should happen if the threshold is exceeded (this involves an analysis of risk, cost and liability). In environmental monitoring the meaning of the term ‘cost’ depends on perspective – a developer generally wishes to minimise the observational effort (e.g. the cost associated with taking samples) but other stakeholders will be more concerned with costs occurring as a consequence of the activity being monitored (e.g. loss of spawning habitat). Following the selection of a metric and threshold (§2.2), there needs to be agreement, between stakeholders, on the cost of either determining that the threshold is exceeded when, in reality, it is not or failing to determine threshold-exceedance when, in reality, it is occurring (known as Type I and II errors respectively). Stakeholders also need to agree how to approach decision-making in the face of varying degrees of uncertainty. In traditional monitoring programmes, the Type I error rate is set (generally, and arbitrarily, at 0.05) whilst the Type II error, which is arguably more important for conservation-risk management, is not specified and is frequently ignored [50]. A ‘total error rate’, split between Type I and II errors, has been suggested to accommodate respective costs [50,51,102]. This type of approach focuses attention on asking relevant questions (e.g. which metric to monitor) and to consider how much change is acceptable whilst balancing the stakeholder-specific costs of making both error types. Where the probabilities and costs of several outcomes are known, or can be reliably estimated, ‘decision theory’ can be used to maximise the positive outcomes [48].

2.4.2. An alternative approach to NHST

Balancing ‘precautionary’ and ‘risk-based’ decision making is core to the EIA process [103] within which monitoring is specified. Given the numerous limitations of NHST (§2.4.1) and the aspiration to manage ecosystem relevant changes, we suggest decision making be based on an upper confidence interval [6] in relation to the evidence provided by data, as illustrated in Fig. 1.

The scenario in Fig. 1 assumes that an initial threshold (defined via expert elicitation [49]) of 20 Units of Impact (UOI, Fig. 1, solid blue line) in the metric of interest has been identified beyond which change is considered unacceptable. In this example, expert opinion about the assessed impact indicates considerable uncertainty about the likely UOI (10 to 30 UOI, Fig. 1 ‘Expert opinion’) but that it is plausible that the threshold might be exceeded (Fig. 1, ‘Risk’) and monitoring is deemed necessary. Further consultation would be required to agree the degree of confidence that should be applied to any point estimate of the relevant parameter; here that confidence is arbitrarily set at 95/99% and relates to the upper confidence boundary only (UCB, Fig. 1, horizontal solid black lines extending horizontally from means A – E). The hypothesis under test is that the threshold (20 UOI) is not exceeded; this is not a null-hypothesis of no-impact. Provided the monitoring programme is designed appropriately (not biased or subject to confounding), it will generate mean-estimates that are centred on the true UOI-parameter. The likely spread of these mean estimates, should the same sampling strategy be repeated many times, is indicated by the lower and upper confidence intervals (under the frequentist paradigm; [104]). If the actual UOI was 10, the most likely outcomes shown in Fig. 1 are A, A*, B and D, which all result in the same point estimate (10) but with differing levels of confidence in that estimate which may result in differing management decisions. The evidence at A suggests that > 20 UOI is unlikely to be found and that monitoring could be reduced (Green). At B, the upper 95% CI indicates that 20 UOI might be exceeded and that monitoring should at least be continued (Amber) whilst the uncertainty is greatest for situation D which might, depending on circumstance, result in enhanced monitor-

![Fig. 1](image-url) – ‘Traffic light’ evidence for action (reduce, maintain/ enhance monitoring or enter into mitigation as a function of the point estimate (usually the mean, indicated here by the letter) and upper 95% and 99% confidence boundary (UCB) as indicated by the solid (all) and dotted (A* only) lines. The threshold of 20 units of impact (UOI)) has been agreed (by expert elicitation). Traditionally, a null hypotheses of no-impact is tested (zero UOI, double line) and this is accepted if the lower CI (Scenario G) includes it. We recommend that decisions be made on based around an agreed upper confidence interval. See main text §2.4.2 for further details. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

The commonly adopted NHST approach would be to test the null hypothesis of no-change (i.e. that the UOI is zero, double blue line on Fig. 1) and make decisions based on the lower confidence interval. Fig. 1G illustrates the results from a very weak sampling programme (large CI) and, assuming the correct value is 25 UOI (the mean of G which exceeds the threshold), one liable to abuse if used to justify inaction. The Type II error rate is proportional to sampling effort - additional sampling could result in Fig. 1:F and the correct rejection of the null even based on the lower CI. Type II errors are routinely made using a 95% CI (A) but not if a 99% CI has been employed (A*) but this would not, as is typically the case, be as a consequence of under-sampling (see below).

The scope for Type II errors are reduced by increasing the UCB and this process can be extended indefinitely but this commensurately increases the rate of Type I errors. In the case of Fig. 1: A, had the actual impact metric been 20 UOI, a Type II error would have occurred using a 95% CI (A) but not if a 99% CI has been employed (A*) but this would not, as is typically the case, be as a consequence of under-sampling (see below).

The widths of the confidence interval (precision) are synonymous with ‘statistical power’ (narrower intervals indicating greater statistical power). In a situation of limited sampling effort (budget), a balance has to be struck between precision (proportional to sampling density) against the population to which the observations can be inferred (proportional to spatial domain). How this is achieved is influenced by the characteristics of the response variable and scale of assessment.
and this should be dictated by the spatio-temporally delimited management objectives (§2.5).

Agreeing the balance between the relative risks (Type I and II errors) and confidence around parameter estimates poses major challenges in agreeing a monitoring protocol. However, justified thresholds e.g. around non-linear effects (§2.2) and pre-data-gathering agreements about how to accommodate uncertainty enable logically-based, defendable and auditable decision-making and a degree of flexibility, adaptability and compromise that is considered a part of successful EIA programmes [48,106–108].

2.5. Assessing change in time and space

With the necessary metrics, thresholds and scales identified the sampling programme must be designed. Accurately and precisely assessing change, in time and space, in highly naturally variable environments is not simple. Whilst new technologies are making the monitoring/mapping of some receptors realistic at large-scales, detecting change/threshold exceedance will always be challenging in the marine environment particularly where temporal/spatial variability is high and/or poorly quantified. Field-based monitoring programmes constitute observational studies and these, inevitably, do not lend themselves to inference (assigning cause to effect) because ‘treatments’ (e.g. wind-farm monopiles) are not randomly assigned to the region of interest. In these circumstances an assessment of change over both space and time requires an understanding of pre-development spatio-temporal variability in the metric under investigation in order to conduct ‘before-after-control-impact’-(BACI)-type comparisons [109–112] and this is routinely done in relation to MRED monitoring. BACI designs are useful where the effect is manifested as an acute and/or long-lasting change in the mean of the response variable but they perform less well where the impact is gradual or manifested as a change in variability [113] as is typical of the changes likely to occur around MREDs [86]. Minimising undesirable variability (i.e. that not attributable to the impact source) in monitoring programmes will maximise their cost-effectiveness [114]. Variability can be controlled by

---

Fig. 2. – Summary of a plausible route to a rationalised monitoring programme, focussed around identified thresholds within agreed spatial and temporal domains. We recommend that, if the estimated effort/cost of a monitoring programme required to assess a change, to agreed precision, is not logistically feasible and/or cost-effective then the suggestion is not to conduct the monitoring programme. The green, amber and red colours (X, Y, Z respectively) should be cross-referenced to Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).
conducting sampling at the same time of year [85], time of day, tidal state (if relevant) and with appropriate stratification (e.g. across different substratum types) and by including environmental and meteorological field-details in the statistical models used to make parameter predictions. Even with careful timing, naturally occurring temporal variability (e.g. as a consequence of unpredictable storm-induced resuspension events) suggests that change should be monitored over large-spatial scales [115].

Many MRED consenting authorities (e.g. in the UK, US (some States) and France) require pre-development site characterisation data to be collected, typically over a period of 1–3 years prior to development. Non-existent or poor-parameterisation of the temporal variability characterising benthic systems (i.e. as will happen when using only 1–3 years’ data) means that these approaches are only likely to detect the most extreme changes [116]. However, such comparisons have been used to support conclusions of no ‘significant’ impact with a ‘high degree of confidence’, presumably based on expert judgement of temporal variability (e.g. tidal device; Strangford Lough, UK [44], 1 offshore windfarm: East Anglia, UK [117]13). Overall, Franco et al. [14] showed that current North Sea offshore windfarm-statutory monitoring studies (UK Sector) could only reliably detect diversity/abundance changes of greater than 50%. Given the absence of a regulatory threshold or clear objectives associated with the monitoring programmes assessed by Franco et al. [14] it is impossible to say whether they were, and continue to be, fit for purpose.

Technologies, such as multibeam- and side-scan-sonar, are increasingly allowing cost-effective mapping of entire MRED sites (pre- and/or post-development) and, with appropriate ground-truthing, this approach reduces the scope for sampling error because whole populations can be mapped [92,118,119]. In the absence of such technology, mapping usually consists of taking samples across the area of interest and interpolating between them. Bijleveld et al. [120] found the optimal approach to benthic mapping was to take samples from a regularly-spaced grid complimented with some randomly located samples whilst Van der Meer [121], in relation to monitoring change in benthic communities over time, found that repeat sampling of randomly located stations was optimal. Data with inherent sampling dependencies (e.g. from BACI designs) pose particular statistical problems and require mixed- or generalised estimated equation (GEE) -models (see the ‘R’ package ‘MRSea’). Many monitoring programmes assess multivariate patterns using dissimilarity-matrix-based analyses [122] (e.g. Primer™ and various routines in the ‘Vegan’ library in R). These approaches are invariably adopted to test null-hypotheses of no-change (usually based on a P-value threshold of 0.05) with all the inherent problems of interpretation (§2.4). An alternative to dissimilarity-based tests is to use a multivariate model-based approach such as provided in the ‘MRVabund’ R package [123] which, unlike the former, enables an assessment of which taxa are changing in relation to multiple drivers.

Understanding long-term trends (e.g. occurring over the lifetime of MREDs) requires post-development comparisons with long-term time-series data [112,123,124] and such data are relatively rare,13 which causes problems for benthic monitoring in relation to any marine activity. Whilst knowledge of natural variability in the populations under investigation is essential to quantify MRED-related change there is currently little guidance on what constitutes an appropriate temporal or spatial scale in relation to MRED monitoring.

3. Conclusions

The coastal zone is facing a period of unprecedented human development resulting in both new pressures and an increase in existing pressures. These pressures arise from diverse activities including the expansion of MREDs which, within the EU, are part of the ‘blue-growth’ agenda [125]. To ensure that blue-growth only occurs in harmony with the environment [6103] decision makers need relevant information [126]. The benthos is a core component of the marine ecosystem and, despite considerable monitoring effort, its ecosystem-level interaction with MREDs is poorly understood. This is because most benthic monitoring is ‘DRIPy’. In the case of MREDs, we argue that benthic monitoring programmes should consider the development within the ecosystem and in the context of ecosystem-service provision [8,19,31,127]. In this respect, the relevant scale for monitoring is likely to be relatively large, extending well outside the boundary of any particular development, and will include numerous activities and developments including multiple MREDs [31].

In order to ‘turn-off the DRIP’ we recommend the approach summarised in Fig. 2. Monitoring studies should start with clear management objectives (§2.1) in terms of relevant metrics and spatial/temporal domains (§2.2 and §2.9) and, where possible, pre-defined thresholds of unacceptable change (§2.2). The metrics, thresholds and necessary confidence in any assessment should be agreed by relevant stakeholders and consenting authorities (§2.4, Fig. 2: Phase I) prior to the sampling-design phase (§2.5, Fig. 2: Phase II). We

13 International Bottom Trawl Survey (1970 - present; http://ocean.ices.dk/Project/IBTS/), the Norderney (German sector, North Sea) macrobenthic time-series data (1978 - present; [136]), the continuous plankton recorder data (1931 - present; http://www.sahibs.ac.uk/), the Shetland Oil Terminal macrobenthic data (1978 – present; http://www.soteag.org.uk), Bay of Morlaix, western English Channel (1977–present; [137]) and from the Northumberland coast (UK, North Sea) a macrobenthic/plankton record collected from 1972 - present [138].
acknowledge that it might be difficult, given the lack of scientific understanding or precedents, to identify/agree thresholds amongst diverse stakeholder groups. Under such circumstances, we urge regulators and developers to consider the value of the data that are being generated in terms of parameterising ecosystem models (Fig. 2: Phase III) [128–130]. Where thresholds can be identified, particularly where the response metric is likely to be strongly non-linearly related to the environmental pressure (§2.2), then consideration should be given as to whether any realistic sampling programme can assess, with the necessary precision, the likelihood of the threshold being exceeded (§2.4.2, Fig. 2: Phase II). In the event that the sampling programme is not considered likely to deliver useful data then the programme should not proceed in its proposed form (i.e. ‘DRIPy’ data should be avoided) and the resources used elsewhere (Fig. 2: Phase II). Where a logical monitoring programme can be designed, regulatory decisions, including the requirement to cease, reduce, maintain or enhance monitoring/enter into mitigation, can then be fully justified to all stakeholders reducing conflict, expenditure and risk (e.g. of litigation) (Fig. 2, Phase III).

In order to understand the ecological consequences of MREDs within a rapidly growing marine sector, with multiple, potentially competing, objectives there needs to be a strategic overview of the relevant spatial domain in which monitoring could occur (§2.3) and this can be challenging to manage [31]. This strategic overview could be provided by the regulators or other competent authority and, in Europe, this regional scale assessment could be initiated under the Strategic Environmental Assessment Directive (SEA Directive 2001/42/EC). Within Scotland (UK), the Scottish Offshore Renewables Research Framework (SpORRAn),[14] established under the auspices of the Scottish Regulator (Marine Scotland) includes a benthic sub-group and is adopting a more strategic, ‘joined-up’ ecosystem-level approach to understanding MRED-benthos interactions. Funding such non-development-specific monitoring/research is challenging and we recommend the Belgian approach where developers contribute to a central funding pot from which strategic research can be funded. Supporting basic, multi-sector, ecosystem-scale research is the best way to understand, and expedite, the sustainable future of the marine renewable energy industry. We urge all concerned (developers, regulators and their advisors) to question the basis of current benthic monitoring programmes and to ‘turn-off the DRIP’.

**Acknowledgements**

This manuscript was developed as part of the Terms of Reference of the ICES Working Group on Marine Benthal and Renewable Energy Developments (WGMBRED) finalised during the 2015 meeting at SAMS, Oban with support from SAMS and the Marine Alliance for Science and Technology, Scotland (MASTS). TW acknowledges support from the FP7 –ENERGY project ‘TIDES’ (Grant no. 322428).

###Appendix A

```r
###R CODE TO DEMONSTRATE ‘TRUTH INFLATION’.
#set parameters
effectsize=5
sample.size=10
nos.iterations=20
SDev=10
Mean=50
Pvalues=NULL; EffectSizeEstimate=NULL; StandardError=NULL
TestData=function(){
d=data.frame(Treatment = gl(2, k=sample.size, length=2*sample.size, labels = c('Low', 'High')),
    Test = c(rnorm(n=sample.size, mean=Mean, sd=SDev),
             rnorm(n=sample.size, mean=Mean-effectsize, sd=SDev)))
    return(d))

lm <- function (modelobject) {
    if (class(modelobject)! = "lm") stop("Not an object of class 'lm'")
    f <- summary(modelobject)$fstatistic
    p <- pf(f[1],f[2],f[3], lower.tail=F)
    attributes(p) <- NULL
    return(p)
}

for(i in 1: nos.iterations){
    d=TestData()
    M1=lm(Test~Treatment, data=d);
    Pvalues[i]=lm(M1))
    EffectSizeEstimate[i]= mean(d$Test[d$Treatment=="Low"])-mean(d$Test[d$Treatment=="High"])
    StandardError[i]=summary(M1)$coefficients[2,2]
}
TestDataSet=data.frame(cbind(Pvalues, EffectSizeEstimate, StandardError))#
require(ggplot2)
Pvalueplot=ggplot(TestDataSet, aes(x = Pvalues, y = EffectSizeEstimate))+
    geom_point(size = 4) +
    geom_errorbar(aes(ymin=EffectSizeEstimate-qt(0.975, sample.size)*StandardError,
                      ymax=EffectSizeEstimate+qt(0.975, sample.size)*StandardError),
                      yintercept=effectsize, col="red", lwd=1.5)+
```

References


