



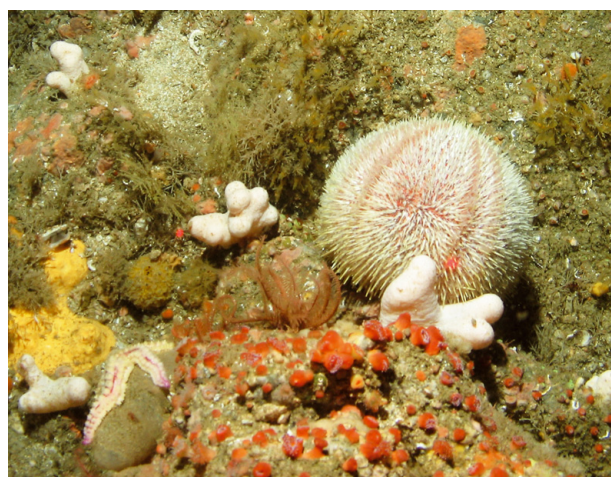
FEATURE ARTICLE

The visual fast count method: critical examination and development for underwater video sampling

Jon Barry*, Roger Coggan

Centre for Environment, Fisheries and Aquaculture Science (Cefas), Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

ABSTRACT: The visual fast count (VFC) estimator was originally developed for censusing reef-fish communities by diver observation, but has more recently been used for estimating abundance of the benthic taxa observed in underwater video transects. It is effectively a sub-sampling method that helps estimate the abundance of visible taxa by segmenting the observation period (e.g. a dive or video record) into equal aliquots and recording the frequency of each taxon only within the first segment in which it appears. Multiplication factors are applied to these counts to estimate the total abundance for the whole observation period. We were interested in the bias that this method may have, as not all species are evenly distributed throughout the observation period. Consequently, we examined the statistical characteristics of the VFC method through simulation studies, with a view to understanding this bias and informing decisions about the use of the VFC in certain applications. We derived the theoretical expected value of the abundance estimator for a random uniform spatial pattern of individuals and also for a clustered spatial process where the number of individuals per unit area follows a negative binomial distribution. We used these results to create new VFC estimators, all of which reduced the bias inherent in the original estimator. We show that this bias is greatest when the mean number of animals is small. We then carried out a number of simulation studies representing random uniform, clustered and regular patterns and compare the standard VFC estimator to a number of other estimators, including the new ones we have developed. In some circumstances, VFC estimators can provide good estimates of species abundance. However, they are also vulnerable to bad performances—particularly if the sampling transect is not homogeneous. If VFC is to be used, we recommend a VFC method based on a method-of-moments estimator, assuming a negative binomial distribution where the order of looking at video segments is randomised.



Seabed image of rock habitat, illustrating variability in abundance and spatial distribution of benthic taxa.

Photo: Roger Coggan, Cefas

KEY WORDS: Transect surveys · Negative binomial distribution

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Estimating the abundance of organisms is fundamental to the monitoring and management of the marine environment. In many cases, the population density of benthic or demersal taxa is estimated from abundance counts in samples taken from a known area of seabed, as with grab, core, dredge or trawl sampling techniques. However, these techniques are not universally applicable, performing poorly on hard substrates (e.g. rocky reefs), and their destructive nature

*Email: jon.barry@cefas.co.uk

is frequently undesirable, particularly in studies concerned with conservation or monitoring the status of a habitat over time; clearly one would wish to avoid the potential of the monitoring study itself degrading the habitat. Consequently, an increasing reliance is placed on non-destructive, observational sampling techniques using divers or underwater imagery (photos and video). These techniques present new challenges to estimating abundance, as the area of seabed sampled is less precisely known than when using grabs or cores and the observer may not detect all objects of interest (Buckland et al. 2001). These issues have been addressed in the past by censusing taxa along a transect line of known length, and counting only those taxa within a limited distance of the transect line, with the assumption that all taxa are detected (Brock 1954, Sale & Sharp 1983). This 'strip transect' is a special case of 'distance sampling', which typically records the number of objects detected and the perpendicular distance to each object from a transect line and does not assume complete detection. Instead, it uses the distance data to determine a detection function, the probability of detecting an object at distance y from the transect line, and ultimately estimates numbers per unit area of seabed (see Buckland et al. 2001 for discussion).

Practical difficulties of laying transect lines underwater led to the development of species–time census techniques, which relate the frequency of observations to a unit of time (e.g. number of species per minute of observation) rather than a unit of area (number of species per square metre of seabed). Jones & Thompson (1978) developed the rapid visual technique (RVT) when working on reef fish assemblages. They described the assemblage in terms of the rank order of encounter of the various species; that is, they divided the total dive time into segments of equal duration and ranked those species encountered in the first segment as more abundant than those encountered in the second segment, and so on. As divers did not have to attempt to count fish, but only record when new species were encountered, they could rapidly assess relatively large areas of reef. Multiple dives provided sample replication and results were tallied to express relative abundance for a species as a sum of ranks. However, DeMartini & Roberts (1982) found the RVT to be inaccurate and biased compared to a full census technique (identifying and counting every fish encountered), as it overemphasised widespread but 'rare' species and underemphasised schooling species, whose behaviour caused them to be patchily distributed, though abundant.

A second species-time method was developed by Kimmel (1985) and was also applied to a census of reef fish assemblages. It is a hybrid between the full census and the RVT methods and is known as the visual fast

count (VFC). The dive was divided into 5 equal time segments (of 10 min duration) and the abundance of each species was counted during the time segment in which it was first encountered. Instead of applying a ranking to the abundance score, Kimmel applied a weighting factor based on expected frequencies, following the logic that if a species was encountered in the first time segment, it would be expected to be present in all 5 time segments, so the count was multiplied by 5; if it was first encountered in the second time segment it would be expected to occur half of the time and so was weighted by a factor of 2.5, and so on. The final results were expressed as relative abundance, the abundance for each species divided by the total abundance for all species. In testing this new method against the strip transect and RVT methods, Kimmel found that the VFC significantly outperformed the RVT method for estimating relative abundance (assuming the strip transect method was accurate) and that 6 replicates captured 98% of the species encountered.

More recently, the VFC method has been applied to analysing video imagery of the seabed in Stangford Lough, Northern Ireland (Magorrian & Service 1998, Roberts et al. 2004, Strong et al. 2006), where there is concern about the effects of demersal fishing operations on seabed communities. Here, the target of the counts were epibenthic invertebrate macrofauna rather than demersal fish. Magorrian & Service (1998) appear to have misinterpreted Kimmel's methodology, in that they counted the number of each species in each of 5 time segments of a video tow, but then still applied the weighting factors as described in the previous paragraph. However, Strong et al. (2006) followed Kimmel correctly, and assessed the performance of the VFC for assessing the relative abundance of epibenthic taxa compared to the direct count method. They concluded that the VFC was a reliable estimator of relative abundance, that it offered substantial reduction in post-survey processing time, and that it could be applied to video where the visual field was not of constant area. Such conclusions make the VFC a potentially attractive method for analysing seabed video material, a practice that will likely increase in future in response to national and international initiatives (e.g. the European Marine Strategy Framework Directive; European Parliament and Council 2008) that require the status of the seabed and the efficacy of conservation zones to be assessed and monitored.

Before beginning our theoretical and simulation work, we need to be clear as to what the VFC, or any other estimator, is estimating. We assume that the transect is itself a sample of a larger population (e.g. a lough or an area of seabed) and that the purpose of the transect is to estimate the density of selected taxa within this larger population. This is subtly different to

the approach of Strong et al. (2006) who evaluated the performance of the VFC estimator by comparing it against the total count for the transect. It does not particularly matter which of these approaches is used, but given that the ultimate purpose of any sampling regime is usually to estimate levels in the wider population, then we prefer our approach.

It is also notable that the VFC has been used to derive measures of both relative and absolute abundance (Roberts et al. 2004, Strong et al. 2006), the former being relevant to studies investigating change in community composition, while the latter investigates changes in population density. These applications do not, in general, change the implications of our theoretical and simulation work. However, there are special situations when this might be the case, and these are considered towards the end of our simulation study.

We also derive an expression for the bias of the VFC estimator assuming a random uniform and a clustered pattern of individuals along the transect. By 'random uniform' we mean that the number of individuals along the transect has a Poisson distribution and their locations have a uniform distribution. We use this result to derive new, potentially less-biased, estimators. We then evaluate the performance of our estimators using random uniform, clustered and regular patterns. We also simulate individuals on a transect that has a density gradient—a situation that is likely to lead to reduced performance by the VFC estimators. Finally, we report on the time taken to count species in some real video footage using the VFC method, and discuss the implications of time saved against performance of the estimators. All computing is done using the free statistical package R (R Development Core Team 2008).

MATERIALS AND METHODS

Video. For application to underwater video sampling, we simulate material collected under conditions of good visibility from a downward-looking camera towed (or drifted) at a constant altitude above the seabed providing a fixed field of view, typically about 1 to 2 m in width. Counting is limited to those taxa large enough to be detected at the chosen speed of tow and distance from the seabed, and includes taxa that form distinct colonies, such as erect sponges. The section of video to be analysed should ideally represent a single habitat/substrate type (e.g. mud, sand, gravel, cobble, rock) because benthic communities vary significantly when these change.

VFC method and estimator. We define the VFC method based on Kimmel (1985). The video sample is divided into n equal segments representing equal aliquots (or sub-samples) of the whole sample. The advantage of the VFC method lies in the significant

reduction in time spent counting abundant species, useful in cases where it is the order of magnitude that is more important than the precise number of individuals. Equally, it ensures that the whole sample is analysed, not just a 2 min sub-sample, so rarer taxa are not ignored. Thus, the VFC is essentially a method of sub-sampling a video record of the seabed; as with all sub-sampling methods, the driver is to reduce the time and effort that would have to be spent in processing material from the entire sample.

Formally, we can define the VFC estimator as:

$$\text{VFC1} = \frac{nc_i}{i} \quad (1)$$

where i is the video segment ($1 \leq i \leq n$) in which the species is first found and c_i is the count of the species in segment i . Most of the applications we have seen have used $n = 5$ segments. Thus, in these applications, we obtain

$$\text{VFC1} = \frac{5c_i}{i} \quad (2)$$

While the choice of $n = 5$ is clearly arbitrary, we will retain this for all examples used in the present study. The nomenclature VFC1 is used to denote Kimmel's original form of the VFC calculation.

Bias of the VFC estimator. Here we derive the expectation, and hence the bias, of the VFC1 estimator as defined in Eq. (2). Note that the bias of an estimator is the expected value of the parameter being estimated (i.e. the mean value of the estimates if a sampling procedure was repeated a large number of times) minus the true value of the parameter. We define the expectation as a function of the mean of the underlying process and the probability, $p(0)$, that no points will be observed in a segment. We then give explicit formulations of this bias for situations where the number of points per segment follow Poisson and negative binomial distributions. The advantages of adopting this approach are that: the negative binomial distribution has a parameter k which can be used to represent clustering in the underlying spatial distribution of points (representing individual organisms); and the Poisson distribution arises where points have a random uniform pattern. We now derive the bias of the VFC1 estimator.

Define $c_i; i = 1, \dots, 5$ as the number of points falling in the i th segment, as in Eq. (2). The i th interval is defined as values in $[i, i + 1)$. The 6 possible scenarios for the standard VFC method are:

- a: $c_1 \geq 1$ (species found in 1st segment)
- b: $c_1 = 0, c_2 \geq 1$ (species found in 2nd segment)
- c: $c_1 = 0, c_2 = 0, c_3 \geq 1$ (species found in 3rd segment)
- d: $c_1 = 0, c_2 = 0, c_3 = 0, c_4 \geq 1$ (species found in 4th segment)
- e: $c_1 = 0, c_2 = 0, c_3 = 0, c_4 = 0, c_5 \geq 1$ (species found in 5th segment)
- f: $c_1 = 0, c_2 = 0, c_3 = 0, c_4 = 0, c_5 = 0$ (species not found)

Defining $p(0)$ as the probability that c_i is zero ($i = 1 \dots 5$), the corresponding probabilities, Pr , of each of these scenarios happening are:

$$\begin{aligned} \text{Pr}(a) &= 1 - p(0) \\ \text{Pr}(b) &= p(0)[1 - p(0)] \\ \text{Pr}(c) &= p(0)^2[1 - p(0)] \\ \text{Pr}(d) &= p(0)^3[1 - p(0)] \\ \text{Pr}(e) &= p(0)^4[1 - p(0)] \\ \text{Pr}(f) &= p(0)^5 \end{aligned} \quad (3)$$

Then the expected value, $E[\text{VFC1}]$, of the VFC1 estimator is given by

$$E[\text{VFC1}] = \frac{\lambda[\text{Pr}(a) + \text{Pr}(b)/2 + \text{Pr}(c)/3 + \text{Pr}(d)/4 + \text{Pr}(e)/5]}{1 - p(0)} \quad (4)$$

where λ is the mean number of individuals on the transect (i.e. all 5 segments). The denominator is an adjustment to allow for the fact that the density estimate in the segment that is counted is conditional on there being at least 1 individual in that segment. This is the key reason why the VFC method is biased—because it stops when it finds a non-empty segment rather than choosing a random number of segments to sample (all of which may be empty). Thus, the bias B as a function of λ is

$$B(\lambda) = E[\text{VFC1}] - \lambda \quad (5)$$

It is a well known result that if the pattern of points on the transect is random uniform with mean λ then the number falling within any segment will have a Poisson distribution with mean $\lambda/5$. Thus, for the Poisson case, $p(0) = \exp(-\lambda/5)$. If we want to consider the situation where the points have some clustered pattern, then the negative binomial distribution (Elliott 1971, He & Gaston 2000) can be used to represent the distribution of the number of points in a segment. The probability density function of the negative binomial distribution can be written as

$$p(x) = \frac{(k+x-1)!}{x!(k-1)!} \left(1 + \frac{\mu}{k}\right)^{-k} \left(\frac{\mu}{\mu+k}\right)^x \quad (6)$$

where μ is the mean and the parameter k that can be interpreted as representing the degree of clustering (low k implies high clustering). The Poisson distribution is the limiting case of the negative binomial distribution as k becomes large. For the negative binomial distribution applied to the number of points per segment we have $\mu = \lambda/5$ and so $p(0) = \left(1 + \frac{\lambda}{5k}\right)^{-k}$. Thus, we can calculate the bias in Eq. (5) for either the Poisson or the negative binomial distributions simply by inserting the relevant value for $p(0)$ into Eqs. (3) & (4).

Fig. 1 shows the percentage bias relative to the size of λ (i.e. $100B(\lambda)/\lambda$) as a function of λ ; lines are plotted for when the abundance of points per segment is Pois-

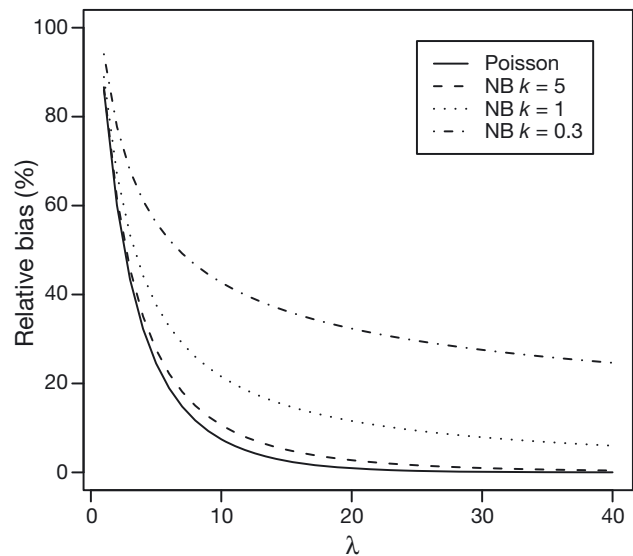


Fig. 1. Percentage relative bias ($100 \times \text{bias}/\lambda$) of the visual fast count method as a function of the mean λ for a Poisson distribution of individuals per segment (random uniform pattern) and for 3 negative binomial (NB) distributions

son (random uniform pattern) and when the abundance has 3 different ($k = 0.3, 1, 5$) negative binomial distributions. As the values of k become smaller, representing more clustered patterns, the bias decays more slowly with increasing mean λ . In Fig. 1, the relative bias for the random uniform pattern and the least clustered of the negative binomial distributions ($k = 5$) is virtually zero for λ greater than about 20. Clearly, the bias is most important for small values of λ (i.e. for rare taxa).

Alternative estimators. One benefit of knowing the bias is that it allows us to derive alternative VFC estimators. For example, if we knew the value of λ then we could simply adjust our estimator by the bias $B(\lambda)$. This would give us an estimator $\text{VFC1}^+ = \text{VFC1} - B(\lambda)$, with expectation λ (from Eq. 5) and hence zero bias. However, because we do not know λ , a pragmatic alternative is to replace the value of λ in the bias term $B(\lambda)$ by its estimator VFC1 . This gives the new estimator

$$\text{VFC2} = \text{VFC1} - B(\text{VFC1}) \text{ if } \text{VFC2} \geq 0 \quad (7)$$

Clearly, a negative count is impossible, so if VFC2 is negative then we set it to zero. Intuitively, it seems likely that this estimator will have reduced bias compared to the standard VFC1 estimator.

Another potential improvement of the VFC estimator can be derived using the standard method-of-moments approach. Here, we simply equate the expected value of VFC1 from Eq. (4) with its observed value from Eq. (2) and then solve for λ (if we are assuming a clustered pattern then we need to provide a value for k). Thus, our estimator is the value of λ that solves the equation

$$\text{VFC1} = \frac{\lambda[\text{Pr}(a) + \text{Pr}(b)/2 + \text{Pr}(c)/3 + \text{Pr}(d)/4 + \text{Pr}(e)/5]}{1 - p(0)} \quad (8)$$

We denote by VFC3^r the estimator obtained with $p(0)$ from a Poisson number of individuals per segment (i.e. r denotes a random uniform pattern), and VFC3^c the estimator obtained with $p(0)$ obtained when the number of individuals per segment has a negative binomial distribution with $k = 1$ (i.e. c denotes a clustered pattern). Note that any value of k could have been chosen but we used $k = 1$ as it we felt it was a 'medium' level of clustering. We used the R function 'optimize' to find VFC3^r and VFC3^c. We now consider some additional estimators that might be used as alternatives to the VFC estimators.

Random sampling estimator (RS). One of the main drivers for use of the VFC method is that it saves time because not all individuals within a transect need to be counted; indeed, given that they are present, individuals of the same species are counted in only 1 segment. However, there are conventional statistical sampling approaches where we could also achieve reductions in sampling effort. For example, we could simply select one of the segments at random, count the number of each taxon and multiply the counts by 5. Similarly, 2 or more segments could be randomly sampled. However, for the purposes of the present study where we want to compare random sampling with the VFC techniques, a reasonably fair comparison is to randomly sample the same number of segments as used in the VFC. Thus, for example, if the VFC method finds a particular taxon is absent in the first 2 segments observed, but present in the third, then the equivalent random sampling method would sample 3 segments. Thus, if the species is first found in the i th segment using VFC and c_j is the count of the species in a randomly chosen segment j , then we define the RS estimator as

$$\text{RS} = 5 \frac{\sum_{j=1}^i c_j}{i} \quad (9)$$

This estimator has the attractive property that it is an unbiased estimation of λ .

Whole-transect estimator (TRA). In habitats where faunal abundance is low (i.e. all epibenthic macrofauna are 'rare'), such as on mobile sands, it might be that using VFC estimators saves little effort. In such instances, it might be appropriate to count all of the individuals within a transect (we denote this estimator by TRA).

To measure the efficiency of the VFC estimators in terms of time spent counting individuals, we can compare the time for VFC against the time taken to count the whole transect. We define this efficiency (EFF) as

$$\text{EFF} = 5t_1 / \sum_{i=1}^5 t_i \quad (10)$$

where t_i is the time taken to count segment i ($i = 1, \dots, 5$). Thus, for example, if all taxa are very common then counting will take place only in the first segment and so the efficiency will be close to its maximum, 5. As species becomes rarer, the efficiency of the VFC methods decreases because the times taken to process Segments 2 to 5 become closer to that taken to process Segment 1. We will report the actual efficiencies calculated from 2 small empirical studies in the section 'Results: Timing of VC'. Note that if we can assume that all species have the same pattern and density (and hence each species will, on average, take the same time to count) we can calculate the expected efficiency assuming a Poisson or negative binomial number of individuals per segment as

$$E[\text{EFF}] = 5 \text{Pr}(a) / [\text{Pr}(a) + \text{Pr}(b) + \text{Pr}(c) + \text{Pr}(d) + \text{Pr}(e)] \quad (11)$$

where $\text{Pr}(a)$ to $\text{Pr}(e)$ are given in Eq. (3). This is because $\text{Pr}(a)$ to $\text{Pr}(e)$ represent the proportions of the species that are counted in each segment. We use this formula for our simulation studies described in the next section.

Comparisons of the estimators. In this section, we report on a simulation study to illustrate and explore the performance of the estimators defined above. As well as the bias, we also measure the relative mean absolute error (RMAE) of the estimator. Over m repeat simulations, where each time an estimate $\hat{\lambda}$ of λ is calculated, we define the RMAE as

$$\text{RMAE} = 100 \sum_{j=1}^m |\hat{\lambda}_j - \lambda| / (m\lambda) \quad (12)$$

We prefer the use of absolute differences to squared differences because the absolute differences are more readily interpretable in terms of the average error of estimation. We use the relative mean absolute error because this allows comparison relative to the actual value of λ , thus enabling us to compare performances across different values of λ .

The RMAE is generally a better measure of an estimator's performance than the bias because it is possible for an estimator to be biased but still to be closer to the true value, on average, than an unbiased estimator. In the first part of our simulation study we compare both the bias and the RMAE of the 4 VFC estimators (VFC1, VFC2, VFC3^r and VFC3^c), the RS estimator, and the TRA estimator from the full transect.

We carry out comparisons for the 3 main classes of spatial pattern: (1) random uniform, (2) clustered and (3) regular. We cover a range of different values of λ in our simulations. These are chosen so that they cover relatively rare species that may not be seen until later

intervals (if at all) and also common species, which are likely to be seen in the first interval.

The simulation studies are carried out by simulating counts on a transect between 0 and 5; hence it has 5 segments of 1 unit each. The estimators are then calculated for this simulation and the process repeated 100 000 times. The bias and RMAE are calculated empirically from these 100 000 simulations (except for where the theoretical bias is known). If any single iteration produces a zero count, then the estimator is assigned a value of 0 for that iteration.

Timed trials. To place the simulation study in context and demonstrate the potential time-efficiency saving offered by the VFC, we conducted 2 timed trials using the VFC protocol outlined above. Trial 1 used video from a rocky reef habitat and Trial 2 from a gravel habitat. In both cases, counts were made of epibenthic macrofauna.

RESULTS

Random uniform pattern

If observations have a random uniform pattern then the number of individuals falling on the transect between 0 and 5 will have a Poisson distribution with mean λ (the number per segment will have mean $\lambda/5$). Table 1 shows values of the bias and RMAE of our 6 estimators, for λ between 0.5 and 50.

The bias of the VFC estimator is largest in percentage terms when the species is rare. Thus, when $\lambda = 0.5$ the bias is 104% and when $\lambda = 1$ it is 86%. The biases for the VFC2, VFC3^f and VFC3^c estimators are much less than that for the VFC1 estimator up to $\lambda = 10$ (up to $\lambda = 5$ for the VFC3^c estimator). The RS and the TRA estimators both have the advantage of being unbiased. We can see that the efficiency of the VFC estimators is low for small λ ; it is 1.2 when $\lambda = 0.5$ and 1.4 when $\lambda = 1$. For these rare species, not much more effort is required to sample the whole transect.

The results in terms of RMAE need careful examination as they demonstrate well the strengths and weaknesses of the VFC method. For values of $\lambda \leq 2$, the RS estimator outperforms the VFC1 estimator and performs equally with the VFC2 estimator. The VFC3^c estimator has lower RMAE than the other estimators (except TRA) for the low values of λ , from 0.5 to 3. However, for λ between 3 and 20, the VFC1 has a marginally lower RMAE than the RS estimator and, for λ between 5 and 20, a marginally lower MAE than VFC2, VFC3^f and VFC3^c. Thus, for these intermediate values of λ , the VFC1 seems to be the best, albeit biased, estimator out of these 4. We can see also from Table 1 that the whole-transect estimator TRA has lower RMAE

Table 1. Random uniform pattern. Efficiency (EFF), bias and relative mean absolute error (RMAE) for the 6 estimators VFC1, VFC2, VFC3^f, VFC3^c, TRA and RS (see 'Material and methods'). Except where the theoretical value is known, all numbers are generated from 100 000 simulations of a random uniform pattern

λ	EFF	Bias (% of λ in parentheses) and RMAE (2nd row)					
		VFC1	VFC2	VFC3 ^f	VFC3 ^c	TRA	RS
0.5	1.2	0.5 (104)	0.1 (17)	0.2 (32)	0.1 (24)	0	0
		227	153	154	146	121	153
1	1.4	0.9 (86)	0.1 (14)	0.3 (26)	0.2 (17)	0	0
		160	125	116	107	74	121
2	1.9	1.2 (60)	0.2 (12)	0.3 (15)	0.1 (6)	0	0
		102	96	88	79	54	98
3	2.4	1.3 (43)	0.2 (8)	0.3 (10)	-0.0 (-1)	0	0
		79	76	71	61	45	85
5	3.2	1.2 (27)	0.2 (4)	0.2 (4)	-0.4 (-8)	0	0
		53	66	66	60	35	64
10	4.3	0.7 (7)	0.0 (0)	-0.1 (-1)	-1.2 (-12)	0	0
		47	53	53	52	25	51
20	4.9	0.2 (1)	-0.2 (-1)	-0.2 (-1)	-2.0 (-10)	0	0
		38	40	40	40	18	39
50	5.0	0.0 (0)	0.0 (0)	-0.1 (0)	-2.4 (-5)	0	0
		25	25	25	26	11	25

than any of the other estimators (approximately half the level of VFC1). However, we have to balance this against the extra efficiency of the other estimators; the VFC estimators are nearly 5 times more efficient than TRA when $\lambda > 20$.

In summary, apart from TRA, VFC3^c is the best estimator for rarer species and VFC1 is the best for species of intermediate density. However, the magnitude of VFC1's improvement over VFC3^c is less at intermediate levels of λ than VFC3^c's improvement over VFC1 at lower levels of λ . The good performance of VFC3^c is surprising given that it was designed to work for a clustered pattern.

Clustered pattern

The number of individuals per segment is assumed to have a negative binomial distribution. Values of λ used are 1, 3 and 10; values of the clustering parameter k are 0.3, 1 and 5.

The results of the simulation study are shown in Table 2. We can see for the VFC1 estimator that the bias is worse for more highly clustered patterns (i.e. those with lower k). The biases for the 2 estimators derived from random uniform patterns, VFC2 and VFC3^f, also get worse as the clustering increases (i.e. as k gets smaller). If, for some reason, the density of a species changes along the length of the transect,

Table 2. Clustered pattern. Efficiency (EFF), bias and relative mean absolute error (RMAE) for the 6 estimators VFC1, VFC2, VFC3^f, VFC3^c, TRA and RS. Except where the theoretical value is known, all numbers are generated from 100 000 simulations of independently drawn negative binomial realisations in each segment, with parameters λ and k

λ, k	EFF	Bias (% of λ in parentheses) and RMAE (2nd row)				TRA	RS
		VFC1	VFC2	VFC3 ^f	VFC3 ^c		
1, 5	1.4	0.9 (87)	1.2 (8)	0.3 (27)	0.2 (16)	0	0
		161	126	119	109	75	122
1, 1	1.4	0.9 (89)	1.2 (11)	0.3 (32)	0.2 (21)	0	0
		169	135	128	117	81	126
1, 0.3	1.3	0.9 (94)	1.4 (18)	0.4 (44)	0.3 (31)	0	0
		188	156	150	137	93	134
3, 5	2.3	1.4 (46)	3.3 (16)	0.4 (13)	0.0 (1)	0	0
		84	81	76	65	47	88
3, 1	2.1	1.6 (53)	3.7 (33)	0.7 (24)	0.4 (11)	0	0
		100	98	93	81	57	97
3, 0.3	1.7	2.0 (68)	4.2 (61)	1.3 (43)	0.9 (28)	0	0
		129	129	125	112	75	112
10, 5	4.1	1.1 (11)	10.2 (12)	0.2 (2)	-0.9 (-9)	0	0
		54	60	61	59	30	60
10, 1	3.3	2.2 (22)	11.4 (68)	1.3 (13)	0.3 (3)	0	0
		75	81	81	77	43	80
10, 0.3	2.4	4.3 (43)	13.6 (180)	3.4 (34)	2.4 (24)	0	0
		110	116	113	108	66	105
20, 5	4.7	0.5 (3)	20.1 (5)	0.1 (0)	-1.7 (-8)	0	0
		50	52	52	52	24	51
20, 1	4.0	2.3 (12)	21.8 (89)	1.5 (8)	0.0 (0)	0	0
		74	76	75	75	39	77
20, 0.3	2.8	6.5 (32)	25.7 (285)	3.2 (16)	0.0 (10)	0	0
		108	111	98	98	64	106

$k = 1$ for $\lambda = 10$ and 20 ; however, when $\lambda = 1$ and 3 , its bias is least when $k = 5$. It seems that, for these smaller values of λ , the tendency for the bias to increase with degree of clustering outweighs the fact that VFC3^c was purposefully designed using $k = 1$, which represents a medium level of clustering.

In terms of the RMAE, VFC3^c clearly does better than the other estimators (except for TRA, which is the best of all the estimators) for $\lambda = 1$ and 3 . For higher values of λ , VFC3^c generally does the same or better than the other estimators (excluding TRA). The overall good performance of VFC3^c covers the range of the $3 k$ values used, and is not restricted to $k = 1$ as might have been expected.

In terms of efficiency, this decreases as the degree of clustering increases. In Table 2, this effect is more pronounced for the larger values of λ . The reason for the reduced efficiency with clustering is that species are less likely to be found in early segments when the pattern is clustered.

Regular pattern

Regular patterns are characterised by points that are less clustered than would be expected from a random uniform pattern. This might be due to, for example, the organisms exhibiting some form of territoriality. Diggle (2003) describes some point processes that can be used to model these regular patterns. However, we consider a simple process based on randomly perturbing a completely regular pattern. We describe it below.

(1) Initially, locations of organisms are generated in a completely regular pattern along a transect. Mathematically the interval between individuals (INT) is $\text{INT} = 5/\lambda$ and they occur on a transect that starts at -4INT and finishes at $(5 + 4\text{INT})$. The extension by 4INT units at each end of the transect is to eliminate any edge effects in the simulations. Thus, the perturbations explained below can push individuals off or on to the transect.

(2) We then perturb the locations by adding a value U to each location, where U is a uniform random variable within the limits of $-R$ to $+R$ (see next paragraph for explanation of R). Any perturbed locations that are <0 or >5 are removed.

R is a measure of the process' departure from regularity. Thus, as R gets larger, the process becomes less regular and closer to random. For the simulations we have used $R = 0.1, 0.6$ and 1 , and $\lambda = 1, 3, 10$ and 20 .

The results of the simulations for the regular processes are summarised in Table 3. The VFC1 estimator is highly biased for $\lambda = 1$; out of the VFC estimators, VFC2 has the lowest bias for this value of λ . VFC3^c has the highest bias when λ is ≥ 3 . In terms of RMAE, VFC3^c performs the best of the VFC estimators when $\lambda = 1$ and 3 , but less well for higher values of λ . However, RS is the worst estimator when $\lambda = 3$. As one might expect, estimators generally perform better when perturbation is least ($R = 0.1$). Overall, it is noticeable that RMAE values are less than for the equivalent values of λ for the random uniform and clustered patterns.

The TRA estimator does very well because the regular pattern means that the number over the whole length of the transect will be close to its true value. Thus, if one is looking predominantly for rare species, then it is better to sample the whole transect as there are little gains in efficiency from using VFC, and the RMAE for TRA is very low.

Non-homogeneous transects

There is potentially a major problem for the VFC estimators because they sample more species in segments at the beginning of a transect than at the end. If, for some reason, the density of the species is different

Table 3. Regular pattern. Efficiency (EFF), bias and relative mean absolute error (RMAE) for the 6 estimators VFC1, VFC2, VFC3^f, VFC3^c, TRA and RS. Except where the theoretical value is known, all numbers are generated from 100 000 simulations of a perturbed regular process (*R* is a measure of the process' departure from regularity)

λ, R	EFF	Bias (% of λ in parentheses) and RMAE (2nd row)				RS	TRA
		VFC1	VFC2	VFC3 ^f	VFC3 ^c		
1, 0.1	1.0	1.3 (128)	0.2 (20)	0.4 (43)	0.3 (34)	0	0
		130	101	84	76	81	1
1, 0.6	1.0	1.2 (125)	0.2 (19)	0.4 (41)	0.3 (32)	0	0
		133	102	85	77	87	8
1, 1	1.1	1.2 (120)	0.2 (17)	0.4 (38)	0.3 (30)	0	0
		133	101	85	78	89	13
3, 0.1	3.0	1.0 (33)	-0.2 (-8)	-0.2 (-7)	-0.4 (-15)	0	0
		47	39	35	28	64	1
3, 0.6	3.0	1.0 (34)	-0.2 (-7)	-0.2 (-5)	-0.4 (-14)	0	0
		50	42	37	30	65	8
3, 1	2.7	1.2 (41)	0.0 (1)	0.1 (3)	-0.2 (-7)	0	0
		58	51	46	38	70	13
10, 0.1	5.0	0.0 (0)	-0.8 (-8)	-0.8 (-8)	-2.1 (-21)	0	0
		7	14	14	24	7	1
10, 0.6	4.8	0.3 (3)	-0.5 (-5)	-0.5 (-5)	-1.7 (-17)	0	0
		29	35	35	37	32	6
10, 1	4.6	0.4 (4)	-0.4 (-4)	-0.4 (-4)	-1.6 (-16)	0	0
		35	41	41	41	39	9
20, 0.1	5.0	0.0 (0)	-0.2 (-1)	-0.2 (-1)	-2.3 (-11)	0	0
		7	8	8	15	7	1
20, 0.6	5.0	-0.2 (-1)	-0.5 (-2)	-0.5 (-3)	-2.5 (-12)	0	0
		24	25	25	27	24	5
20, 1	5.0	-1.0 (-5)	-1.3 (-7)	-1.4 (-7)	-3.2 (-16)	0	0
		29	30	30	33	29	6

this will cause further bias in its VFG estimator. This might happen, for example, if transects were conducted over a gravel bed that had been sorted by tidal currents, so that the proportion of larger particles (i.e. cobbles) increased with distance along the transect.

To illustrate the effect of non-homogeneous transects numerically, we have carried out further simulations where we introduce a negative percentage trend *T* into the mean λ along the transect, but preserve the overall mean of the transect as λ . The values of λ in each of the 5 segments are:

- Segment 1: $\lambda + 2T\lambda/100$
- Segment 2: $\lambda + T\lambda/100$
- Segment 3: λ
- Segment 4: $\lambda - T\lambda/100$
- Segment 5: $\lambda - 2T\lambda/100$

We have chosen a negative trend to illustrate the situation where the VFC1 estimator will work less well, i.e. VFC1 should overestimate λ even further in this sit-

uation than with no trend. We have carried out the simulations with trends of *T* = 10% and *T* = 40% and a range of values for λ . The pattern of points along the transects is random uniform.

Table 4 illustrates the results for the VFC1 and RS estimators. However, we have also calculated 3 modified VFC estimators. The first of these, VFC1*, is the same as VFC1 except that the order in which the transects are sampled is random. This should remove the bias due to the trend. VFC3^{f*} and VFC3^{c*} have similar random choice of segments but are based on the method-of-moments estimators VFC3^f and VFC3^c.

For the 10% trend and when $\lambda = 1$, VFC1 has a similar level of bias to that for the homogeneous transects described in Table 1 (86% here, 102% in Table 1). However, for larger λ , the bias never drops to zero and reaches an asymptote of 20% of λ . This is because all counts will be taken from the first segment, which is 20% above the mean level (as defined in the simulation study). A similar, though more extreme, version of this effect occurs for the 40% trend. For the 10% trend, the RMAEs of the VFC1 and the RS estimators are similar except for $\lambda = 1$ when RS is much lower (VFC1 = 176, RS = 126). However, for the 40% trend, the

Table 4. Non-homogeneous transects. Bias and relative mean absolute error (RMAE) for the VFC1, VFC1*, VFC3^{f*}, VFC3^{c*} and RS estimators for 10% and 40% decreasing trends *T* along the transect. Except where the theoretical value is known, all numbers are generated from 100 000 simulations of a random uniform spatial pattern with mean λ

λ	<i>T</i> (%)	Bias (% of λ in parentheses) and RMAE (2nd row)				RS
		VFC1	VFC1*	VFC3 ^{f*}	VFC3 ^{c*}	
1	10	1.0 (102)	0.9 (85)	0.2 (25)	0.5 (17)	0
		176	159	116	107	126
5	10	2.1 (42)	1.3 (25)	0.2 (5)	-0.4 (-8)	0
		63	54	66	61	67
10	10	2.5 (25)	0.8 (8)	0.0 (0)	-1.2 (-12)	0
		55	48	54	53	54
20	10	4.1 (20)	0.6 (1)	0.0 (1)	-1.9 (-10)	0
		44	39	41	42	41
50	10	10.0 (20)	0.0 (0)	0.0 (0)	-2.4 (-5)	0
		31	28	27	28	28
1	40	1.5 (153)	0.9 (89)	0.3 (30)	0.2 (19)	0
		227	163	120	110	141
5	40	4.7 (93)	1.6 (33)	0.6 (12)	0.0 (0)	0
		104	62	73	67	82
10	40	8.2 (82)	1.6 (16)	0.8 (8)	-0.4 (-4)	0
		94	58	64	61	71
20	40	16.1 (81)	1.4 (7)	1.0 (5)	-0.8 (-4)	0
		86	55	57	56	61
50	40	39.9 (80)	0.6 (1)	-0.2 (0)	-0.3 (-5)	0
		80	53	51	51	54

RMAEs of VFC1 remain higher, for all λ , than those of the RS estimator. The levels of bias for VFC1*, VFC3* and VFC3c* are, as we would expect, similar to those for the random uniform pattern shown in Table 1. In terms of RMAE, VFC3c* out-performs all of the other estimators for both trends when $\lambda = 1$. For higher values of λ , performance amongst the estimators is more equal, though VFC1* performs best for $\lambda = 5$ and 10.

Overall, the more different the first segment is to the others (i.e. the greater the trend), the worse VFC1 will perform compared to the other estimators.

Relative abundance

We noted in the 'Introduction' that many practitioners have used the VFC approach to estimate the relative abundances of species in a population. Formally, we define the percentage relative abundance (RA) for species j^* as

$$RA_{j^*} = 100\lambda_{j^*} / \sum_{j=1}^N \lambda_j \quad (13)$$

where λ_{j^*} is the abundance of species j^* , N is the number of species and $\sum_{j=1}^N \lambda_j$ is the sum of abundances

over all species. To estimate RA we simply plug in our estimates for λ_{j^*} and the λ_j . Estimating relative rather than actual abundance can complicate matters because RA for species j^* is influenced by the abundance of the other species. This is particularly important when examining the VFC methods because these estimators of abundance are more biased for low abundances than for high ones. We have carried out a stylised simulation study to illustrate how this can be important.

We simulated transect data from a population where there were 100 species, all with a random uniform spatial pattern. The densities we allocated to the 100 species were according to 2 scenarios:

Scenario 1: Each species has a mean $\lambda = 1$. This represents an impoverished community where all epifauna are present at similar low numbers, for example on sand (i.e. there is a high degree of 'evenness' in the community; Pielou 1966).

Scenario 2: Species 1 has a mean $\lambda = 1$. The remaining 99 species have means $\lambda = 10$. This represents an abundant epifaunal community, such as that found on rocky reefs, with numerous taxa at the lower trophic levels and a single higher-level species (maybe a top predator, or something that is rare or is using the habitat as a shelter or refuge) in low abundance.

For both scenarios, we then estimated RA (Eq. 13) of the first species (i.e. $j^* = 1$) using the VFC1 and the RS methods. As previously, we repeated this process

100 000 times and calculated the RMAE from the following modified version of Eq. (12):

$$RMAE = 100 \sum_{j=1}^m |R\hat{A}_1 - RA_1| / (m\lambda) \quad (14)$$

where RA_1 is as defined in Eq. (12) with $j^* = 1$ and $R\hat{A}_1$ is its estimate.

Table 1 has shown that VFC1 is highly biased for $\lambda = 1$ and that it is outperformed by the RS estimator in terms of RMAE. However, in our first simulation scenario (1 above) we might expect VFC1 to do quite well because although the actual estimate for the first species will be biased, the estimates for all of the other species will have the same bias. Thus, when calculating RA for Species 1, we would expect all of the biases for its component estimates to cancel out. We can show this algebraically. If the bias for each of the 100 estimators is B , then RA based on Eq. (13) for any species j^* is

$$RA_{j^*} = 100(\lambda_{j^*} + B) / \sum_{j=1}^N (\lambda_j + B) = 100\lambda_{j^*} / \sum_{j=1}^N \lambda_j \quad (15)$$

Indeed, this cancellation is what seemed to happen in our simulations. In the first of our simulation scenarios, the RMAEs for VFC1 and RS were 88 and 122 respectively. Thus, VFC1 had outperformed the RS estimator.

For the set of simulations defined by Scenario 2 above, where Species 2 to 100 each had a mean of $\lambda = 10$, we would not expect the relative VFC1 estimator for Species 1 to do well. This is because the bias for the first species with λ is 0.86, which is greater than the bias of 0.74 for the other 99 species, and hence the bias for the first species will not be cancelled out. This was reflected in the simulation study where the relative RMAE for VFC1 was 148 and for RS was 121. Hence, even though VFC1 was estimating the same value of λ , i.e. $\lambda = 1$, it performed much better under Scenario 1 than under Scenario 2.

Timing of VFC

Table 5 shows the time taken for processing each 2 min segment, the time saved by using VFC and the efficiency, using the VFC method compared to sampling the whole transect. In practice, the amount of time saved would be dependent on the characteristics of the area being sampled, the number of segments taken and the skill of the analyst in identifying species. Nevertheless, the results are indicative of considerable amounts of time being saved when using VFC methods. In Trial 1, most species are seen in the first segment and so the time taken to count new species in Segments 2 to 5 is much less. The same, though less pronounced, effect occurs in Trial 2. We can see that the VFC method is 3.2 (Trial 1) and 2.6 (Trial 2) times more efficient than sampling the whole transect.

Table 5. Timed trials. Results of 2 timed trials using the visual fast count methodology for estimating abundance of epibenthic macrofauna. Trial 1 on rocky reef, Trial 2 on gravel, each having a 10 min video divided into 5×2 min segments. Count: no of. individuals counted in each segment; time: how long each segment took to analyse. Time saved = (Segment 1 time \times 5) – total time. Efficiency = (Segment 1 time \times 5) / Total time

Segment	Trial 1		Trial 2	
	Count	Time (mm:ss)	Count	Time (mm:ss)
1	411	19:50	83	13:45
2	1	2:54	0	3:53
3	2	3:10	4	3:22
4	0	2:30	2	2:52
5	1	2:40	1	2:50
Total	415	31:04	90	26:42
Time saved		68:06		42:03
Efficiency		3.2		2.6

DISCUSSION

Much of the previous work on estimating bias in visual-based surveys has focused on diver-swum transects used to census reef fish populations (Sale & Sharp 1983, St. John et al. 1990, Cappo & Brown 1996, Cheal & Thompson 1997). Sale & Sharp (1983) found that the degree of bias was directly related to the width of the transect used, underestimating total abundance of reef fish by between 11.1 and 26.7% for a 1 m wide transect, and more for wider transects. These investigations examined the accuracy and bias of the visual-transect census method, in which data from the entire transect was recorded. This is different to our investigation of the bias associated with the VFC, which is a method for sub-sampling a video record.

For studies that need only relative measures of abundance, bias is not a problem so long as it remains constant for each species. In the special case of the VFC estimators, the problem of methodological bias lies in the fact that different organisms exhibit different spatial patterns and have different abundance levels within the same video sample; this induces different biases for different species. Andrew & Mapstone (1987) reviewed problems that spatial patterns impose on sampling marine organisms, but not in respect of the VFC technique.

The essential reason that the standard VFC estimator is biased is because sampling stops when individuals are found in a segment. By then assuming that the rest of the transect is the same as the part already sampled, we overestimate the abundance of the species. We have demonstrated this bias both theoretically and empirically and that this bias and the RMAE are worst for the rarer species. We have demonstrated this in

simulation studies with random uniform, clustered and regular patterns. However, we have also demonstrated that our new VFC estimators have much reduced bias and RMAE. These estimators also performed favourably when compared to the RS estimator. One reason for this, particularly at intermediate densities ($\lambda = 3$ to 10) is because the VFC estimators for a particular taxon are less likely to have a zero value than the RS estimator. The VFC estimators are zero only for species that are absent from the whole transect. However, estimators such as RS have a much higher probability of obtaining a zero value because they sample from a subset of the available segments. For small values of λ , this does not affect the performance of the RS estimator because the true value of λ is close to zero. However, for larger values of λ , this increased likelihood of a zero estimate increases the RMAE.

In general, the VFC estimators do badly when the species density or pattern changes (e.g. as part of some trend) along the whole of the transect. The VFC sampling method for a particular taxon is not random in that segments at the beginning of the transect will be more likely to be counted for that taxon than segments at the end. For common species, nearly all of the counting for the VFC estimators will be from the first segment. Thus, if the early segments are different to the later ones then the VFC estimators will perform badly and give biased estimates. In practice, this could be avoided by good sampling design and ensuring that transects are randomly chosen within the population examined. However, for small areas such as a gravel bed, edge effects on a transect from the boundaries of the bed might make this problem difficult to avoid. This has led to our suggested estimators VFC1*, VFC3^r* and VFC3^c*, which randomise the order in which the video segments are looked at and hence removes this bias problem.

Overall, the method-of-moments VFC3^r and VFC3^c estimators have done better than the VFC1 and VFC2 estimators, with VFC3^c performing the best. The slight drawback with VFC3^r and VFC3^c is that their estimation requires the solution of a non-linear equation. However, this is straightforward to do with standard optimising software that is widely available in many statistical and mathematical packages.

The RS estimator has the desirable property of being unbiased and has better RMAE for rarer species than VFC1. However, throughout our simulation studies RS was generally outperformed by VFC3^c. Because RS is based on a random sampling technique, it will give unbiased results even if the transect is non-homogeneous. However, one of its main drawbacks when considered against the VFC estimators is that it will detect fewer species. Thus, for studies in which identifying rarer species is important, RS would not be recommended.

In terms of time-saving for VFC analysis compared to sampling the whole transect, our 2 studies on 10 min video clips (summarised in Table 5) showed time savings of 68.7% and 61.2% (68 and 42 min respectively) using the VFC. Our trials show that after the highly abundant taxa have been counted in the first segment, counts for the successive segments proceed at around 1 to 2 times real time (e.g. 2 to 4 min per 2 min segment). Clearly, if some of the taxa are at the limit of the resolution of the image (i.e. they are small in the image), then these counts will take longer, as the smaller taxa are more difficult to spot.

Our trials indicate that the VFC method represents a significant reduction in processing time of around 65%. Such savings become highly desirable in studies with a lot of video material to process, for example 2 recent studies in the English Channel by James et al. (2007, 2010) that collected 65 and 100 video transect samples respectively. If it takes, say, 80 min to make a full count analysis on a 10 min video clip, then 100 such clips would take approximately 17 full working days (at 8 h work-time d⁻¹) to analyse. Using the VFC method would give a time saving of nearly 11 d of staff time.

In terms of RMAE, we have shown, not surprisingly, that the best thing to do is to sample the whole transect (i.e. use the TRA estimator). If the majority of the species were rare species, the difference in sampling effort needed to do this would be not much more than for the VFC estimators. However, if there is a significant number of common species, the effort is considerably more (i.e. VFC is much more efficient). In the timed trials, Trial 1 had 363 colonies of the soft coral *Alcyonium digitatum* (commonly known as 'dead men's fingers'), producing an estimated 1815 colonies in the full 10 min video. It is the magnitude of this count, rather than the precision, which is of most interest to ecologists studying community structure and population dynamics. Multivariate analysis of community structure, such as the approaches developed by Clarke & Warwick (2001), typically transforms abundance data to down-weight the dominance of highly abundant taxa, which are usually more robust to environmental stressors and change than the rarer, more specialist species. There is therefore little need to have precise counts of total abundance in highly numerous taxa; you just need to know that there are a lot of them compared to the other taxa present, as it is the rarer taxa that are likely to respond first to environmental change, so these become the focus of the ecologist's attention. When faced with analysing a set of video images, the operators need to make some form of cost-benefit judgement when selecting whether or not to use whole-transect counts or VFCs. Prior viewing of all the available material will help inform this decision.

We have used the negative binomial distribution to represent the abundance of points in a segment when the underlying process is clustered. As noted by Diggle (2003), for example, only a few special cases of cluster processes give rise to a negative binomial distribution and there are plenty of cluster processes that do not do so (e.g. the family of clustered processes introduced by Neyman & Scott 1958). However, we feel that using the negative binomial distribution here gives a good illustration of the results to be expected for VFC when the underlying process is clustered.

The standard VFC method has been used several times (Kimmel 1985, Magorrian & Service 1998, Roberts et al. 2004) and its use is being further proposed as a tool to analyse video footage (Strong et al. 2006). Strong et al. (2006) compared results from the VFC method with those from direct counts of the whole transect, and concluded that the VFC method is 'a reliable alternative to direct counts'. They arrived at this conclusion after performing a number of multivariate comparisons between the estimates and direct counts. They did, however, note that the VFC method may have reduced the variability of the estimates compared to those from the whole transect.

Given that we have spent much of the present paper evaluating the VFC technique, we need to answer the obvious question: Should the VFC method be used in practice? Answering this question is not simple, as it depends, to some extent, on the practical circumstances of the survey being conducted. For the purposes of these recommendations we are assuming that the survey is a video survey and that the species are counted during video playback, on a system that can control the speed of that playback. We hope that the following points will provide useful guidance:

- (1) The whole video sample should be viewed in a fast-forward mode (say 2× normal play speed) to assess if it is suitable for analysis; that is, it meets the assumptions of image quality and habitat homogeneity laid out earlier (see 'Materials and methods: Video'). This will also provide the analyst with an overview of the composition of the benthic community in terms of rare versus common species, which will inform their selection of an appropriate analysis methodology (as detailed in the following paragraphs).

- (2) If processing time is not an issue then the best possible action is to count all individuals of all species on the whole transect. Similarly, if there are only a few very common species and they are present along the transect then it may be reasonable to estimate the abundances of these species using a VFC method, but to count all occurrences of the other species. We have shown that VFC should perform well for the common species.

- (3) Given that with video sampling it is relatively easy to randomise the order of the segments, we would

generally recommend using the estimators based on doing this (our preference would be for VFC3^{c*}) so as to avoid any potential bias due to non-homogeneity of the transect.

(4) If it is not possible to randomise the order of sampling the segments then VFC3^c is the best of the VFC estimators that we investigated. However, further work is needed to generate the optimum value of *k* to use for this estimator.

(5) If practitioners are concerned about using VFC techniques, but still want to save processing time, then the RS provides a safe and unbiased alternative. However, we need to recognise that this will not identify as many species as the VFC methods.

(6) The same concerns and recommendations about using VFC1 in points (1) to (3) above also apply to calculations of relative abundance.

The VFC method offers significant efficiencies when processing video samples of the seabed, but those who use it should be aware of its properties and potential bias. We have provided a number of alternative versions of the basic VFC estimator and shown how these perform in a number of sampling situations. Before any attempt is made to apply the VFC or any other method, the video sample itself must meet certain suitability criteria. The image should be clear and have a relatively stable altitude from the seabed. It must also have a significant period representing a homogenous habitat. The VFC method is not applicable in some circumstances, for example in estimating densities of superabundant species, as found on mussel or brittle-star (ophiuroid) beds. In those circumstances, still photographs may be more appropriate, regarding each image as if it were a quadrat sample. The VFC would also not be appropriate for studies that derive biometric measurements of taxa (such as fish lengths) from video samples. Statistical robustness here requires a suitable sample size, maybe of 100 individuals, to build up a length–frequency curve.

Acknowledgements. We acknowledge the suggestion from Chris Ferro, Mathematics Research Institute, Exeter University that led to the method-of-moments estimators. We are also grateful to a referee whose suggestion to use the negative binomial distribution led to the development of the estimators for clustered populations. Work by J.B. was carried out using funding awarded to Cefas from the Defra Marine Environmental Indicators Project M4118.

LITERATURE CITED

- Andrew NL, Mapstone BD (1987) Sampling and the description of spatial pattern in marine ecology. *Oceanogr Mar Biol Annu Rev* 25:91–112
- Brock VE (1954) A preliminary report on a method of estimating reef fish populations. *J Wildl Manag* 18:297–308
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling. Oxford University Press, Oxford
- Cappo M, Brown IW (1996) Evaluation of sampling methods for reef fish populations of commercial and recreational interest. Tech Rep No. 6. CRC Reef Research Centre, Townsville
- Cheal JA, Thompson AA (1997) Comparing visual counts of coral reef fish: implications of transect width and species selection. *Mar Ecol Prog Ser* 158:241–248
- Clarke KR, Warwick RM (2001) Change in marine communities: an approach to statistical analysis and interpretation, 2nd edn. PRIMER, Plymouth
- DeMartini EE, Roberts D (1982) An empirical test of biases in the Rapid Visual Technique for species-time censuses of reef fish assemblages. *Mar Biol* 70:129–134
- Diggle PJ (2003) Statistical analysis of spatial point patterns, 2nd edn. Arnold, London
- Elliott JM (1971) Statistical analysis of samples of benthic invertebrates. Sci Publ 25. Freshwater Biological Association, Windermere
- European Parliament and Council (2008) Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Off J Eur Union 51(L164):19–40
- He F, Gaston KJ (2000) Estimating species abundance from occurrence. *Am Nat* 156:553–559
- James JWC, Coggan RA, Blyth-Skyme VJ, Morando A and others (2007) The eastern English Channel marine habitat map. Sci Ser Tech Rep 139. Cefas, Lowestoft. www.cefas.co.uk/publications/techrep/tech139.pdf
- James JWC, Pearce B, Coggan RA, Arnott SHL and others (2010) The south coast regional environmental characterisation. British Geological Survey Open Report OR/09/51. Marine Environment Protection Fund, Cefas, Lowestoft. www.alsf-mepf.org.uk/projects/2008/rec-0802/final-report.aspx
- Jones RS, Thompson MJ (1978) Comparison of Florida reef fish assemblages using a rapid visual technique. *Bull Mar Sci* 28:159–172
- Kimmel JJ (1985) A new species-time method for visual assessment of fishes and its comparison with established methods. *Environ Biol Fishes* 12:23–32
- Magorrian BH, Service M (1998) Analysis of underwater visual data to identify the impact of physical disturbance on horse mussel (*Modiolus modiolus*) beds. *Mar Pollut Bull* 36:354–359
- Neyman J, Scott EL (1958) Statistical approaches to problems of cosmology (with discussion). *J R Stat Soc B* 20:1–43
- Pielou EC (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* 13:131–144
- R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. www.R-project.org
- Roberts D, Davies C, Mitchell A, Moore H and others (2004) Strangford Lough ecological change investigation (SLECI). Report to Environment and Heritage Service. Queen's University, Belfast
- Sale PF, Sharp BJ (1983) Correction for bias in visual transect censuses of coral reef fishes. *Coral Reefs* 2:37–42
- St. John J, Russ GR, Gladstone W (1990) Accuracy and bias of visual estimates of numbers, size structure and biomass of a coral reef fish. *Mar Ecol Prog Ser* 64:253–262
- Strong JA, Service M, Michell AJ (2006) Application of the visual fast count for the quantification of temperate epibenthic communities from video footage. *J Mar Biol Assoc UK* 86:939–945