Fisheries Research 213 (2019) 151–159

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

A zigzag survey design for continuous transect sampling with guaranteed equal coverage probability

Alf Harbitz

Institute of Marine Research, Framsenteret, P.O. Box 6606 Langnes, NO-9296 Tromsø, Norway

ARTICLE INFO

Handled by George A. Rose

Keywords:
Parallel survey design
zigzag survey design
Equal coverage probability
Curved transect sampling

ABSTRACT

Marine resource surveys in large areas have high cost, and to find an optimal survey design with regard to efficiency and scientific outcome is an important issue. A randomized zigzag design for straight line and curved transects is developed that guarantees equal coverage probability, i.e., each point in the study area has the same probability of being covered by the transect. The basic idea is to fit automatically either the smallest rectangle, or the smallest circular sector enclosing the actual area. Then a recipe for the location of zigzag legs that provide equal coverage probability everywhere in the rectangle or circular sector is outlined, and thereby also at any location within the study area, which simplifies unbiased abundance estimation. The cost of this approach is the unwanted distance to be traveled from the point where a transect leg leaves the study area to the point where the next leg enters. A comparison of a randomized parallel, straight line zigzag, and curved zigzag approach applied to 7 sandeel areas with great variety revealed an average off-effort traveling distance of 28%, 9% and 6%, respectively. Thus, it appears that the developed zigzag design is far more efficient than the parallel design.

1. Introduction

Survey design in terms of a recipe for a transect route for continuous sampling is an important issue related to, for example, high cost abundance estimation of marine resources in large areas at sea. Often there is a sparse knowledge of the spatial structure of the target species, such as the location of possible aggregations. In these cases, it can be appropriate to apply a randomized, contrary to a fixed, survey design in order to reduce estimation bias. Sampling related to a randomized survey design is often denoted design-based sampling, and this is the approach considered in this paper.

Equal coverage probability is an essential term related to design-based sampling that will be used throughout the paper and therefore needs a clear definition. The paper restricts its attention to horizontal transects with a constant sampler width, $2w$, perpendicular to the transect direction, within a study area with a predefined border (Fig. 1). The randomized survey design provides different independent transects so that each point in the study area has a positive probability of being located inside the area covered by the transect. If any point in the study area, independent of location, has the same probability of being covered by a random transect, the survey design provides what we define equal coverage probability. Such a design has the advantage that it may provide very simple unbiased abundance estimators.

Parallel designs are far more frequently applied than zigzag designs. A major reason for this is its simplicity with regard to obtain equal coverage probabilities as well as a generally simpler analysis. Another advantage is its ability to map the actual study area by proceeding along each leg until no more abundance is observed (Petitgas, 1993). A major drawback is that the transportation between the parallel legs normally is not included in the data analysis, which in practice easily might result in a large waste of effort. The latter drawback with parallel designs is a major argument to also consider a zigzag alternative.

An advantage with a zigzag design that to the author's knowledge has not been emphasized, is the possibility to examine the mobility of the population under study by applying a return design: First a zigzag is run from one side of the area to the other, then the complementary zigzag is applied in the opposite direction. In this way there will be several cross points with different time lags. Say, for example, that a variogram (Rivoirard et al., 2000; Chilès and Delfiner, 1999) based on the along-track observations (close in time) reveals strong positive correlations at small scale. If these correlations disappear or are strongly weakened for close observations in space, but with substantial time lags, this is a strong indication of mobility. In a worst case scenario, the same animal is observed several times during the same survey, with a likely overestimation of the abundance as a result.

A general and easily applied recipe for a randomized zigzag design based on a design axis (straight line) through the area is outlined in Strindberg and Buckland (2004), SB henceforth. The paper strongly

E-mail address: alf.harbitz@hi.no.

https://doi.org/10.1016/j.fishres.2019.01.015
Received 16 August 2018; Received in revised form 10 December 2018; Accepted 19 January 2019
Available online 30 January 2019

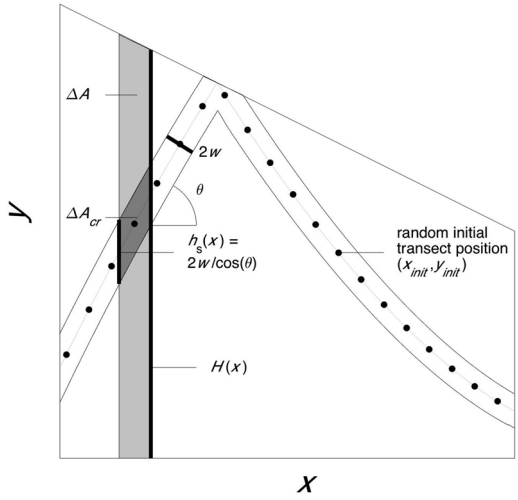

Fig. 1. Illustration of the concept of equal average coverage probability, independent of x, by Strindberg and Buckland (2004). For each vertical strip with width Δx, the ratio of the area ΔA_{cr} covered by the sampler, and the area, ΔA, of the strip (or $h_s(x)/H(x)$), must be a constant independent of x. $2w$ is the width of the sampler perpendicular to the transect direction.

recommends that the turning points at the boundary of the study area should be at the intersection between equidistant lines perpendicular to the design axis (x-axis), and the boundary, in order to obtain approximately equal coverage probabilities. They also demonstrate how one generally can construct a transect sampler with equal (average) coverage probabilities perpendicular to the design axis independent of x-value (Fig. 1). However, their approach does not guarantee equal coverage probabilities along the perpendicular y-axis, and we will demonstrate that an irregular border easily causes rapid fluctuations in coverage probabilities.

Randomized designs have the advantage that design-unbiased abundance estimators can be constructed independent of the spatial structure of the animal locations, based on, e.g., the coverage probability, π_i, of any location in the study area to be included. A natural choice here is the well-documented Horvitz-Thompson (HT, henceforth) estimator (Horvitz and Thompson, 1952; Thompson, 2002). With equal coverage probabilities, the HT estimator for the total abundance (biomass or numbers) is reduced to the simple average of the density observations, e.g., abundance per square nautical mile, multiplied by the study area. A big challenge, however, is to find an appropriate variance estimator, and this topic is treated further in the Horvitz-Thomspon Section 2.5.

The main objective of this contribution is to outline a concept for a randomized zigzag design that guarantees equal coverage probabilities per unit area at any location in the study area, including return zigzags. The basic idea is to develop a randomized zigzag sampler with equal coverage probabilities for any circular or rectangular sector. The next step is then to enclose the actual study area by the sector with the smallest area, and to develop algorithms that automatically find the needed parameters for this optimum enclosing area. The cost of this approach is the transport from the point where a transect leg leaves the study area to the point where the next leg enters the study area again. Applied to an actual curved-shaped minke whale study area in the Antarctic, this unwanted transport posed just a minor part of the total transect length, contrary to a parallel design alternative with the same coverage of the study area. Similarly, efficiency was improved by application of the enclosing rectangular and enclosing circular approaches to several sandeel areas in the North Sea, compared to the parallel design.

2. Methods

2.1. The Strindberg-Buckland (SB) even coverage concept

The SB concept for a randomized zigzag sampler with even coverage is illustrated in Fig. 1 for a trapezoidal study area, where $2w$ is the width of the sampler perpendicular to the travel (transect) direction. First a random transect point, $(x_{\text{init}}, y_{\text{init}})$, within the area is generated (see Fig. 1). Then the transect points in each direction from the initial point are calculated by an iterative procedure. Let $\Delta A_{cr} = h_s \bullet dx$ denote the area of the cross section between the transect and a vertical strip with height, H, width, dx, and area $\Delta A = H \cdot dx$. To obtain equal coverage probability independent of x, the ratio $\frac{\Delta A_{cr}}{\Delta A} = h_s/H$ must be constant independent of x. This constant must also be equal to the ratio between the total area covered by the sampler, $L \bullet 2w$, and the study region area, A, where L is the total transect length and $2w$ is the width of the sampler. This leads to the following expression for the angle, θ, between the transect direction and the x-axis:

$$\theta(x) = cos^{-1}\left(\frac{A}{LH(x)}\right), \quad L \geq A/\min\,(H(x)) \tag{1}$$

where H is the height of the study region in the y-direction at position x. The concept depends on the assumption that $L \geq A/\min(H(x))$, i.e. that the effort (transect length), L, needs to be sufficiently large. Note also that there are two different solutions for θ, because $\cos\theta = \cos(-\theta)$. In the SB concept either of these two values at the initial (x_{init}, y_{init}) point is chosen with equal probability.

Though the equal coverage probability concept above ascertains equal average probabilities along the y-axis independent of x, it does not guarantee equal coverage probabilities along a vertical strip for different y-values. In fact, the region borders might have a rather dramatic effect on the coverage probabilities in the y-direction, as illustrated in Fig. 2 for the Antarctic region (documented in Fig. 7 in SB) and described further in the result Section 3.1.

2.2. The equal coverage probability sampler for a circular sector

We created a new randomized design such that a random transect has the same probability of covering any small unit area domain within a circular sector independent of location (x,y). An example is shown in Fig. 3 by the curved zigzag transect with width $2w$, along with a definition of terms to be used in the following outline. The transects are cyclic with a repeated shape and cycle α_c. A random transect cycle can be generated by first choosing a starting point at $r = r_1$ (the inner circle) and a random angle α_{rnd} from a uniform distribution $U(\alpha_1, \alpha_1 + \alpha_c)$. How the rest of the transect is generated is further described below. First, we outline the concept of the equal coverage probability approach.

Consider a circularly shaped strip at radius r with an infinitesimally small width dr and area $r(\alpha_2 - \alpha_1)dr$, and a survey transect with width $2w$ much smaller than r. Further, let the u-axis denote an axis along the circular strip at any point, i.e. perpendicular to the radial through that point (Fig. 3). The cross section between the strip and the transect comprises a parallelogram with area dA_{cr} that depends on the angle, θ, between the u-axis and the transect direction. Due to symmetry dA_{cr}, and thus the value of θ, must be equal for each cross section at the same value of r (but different α's) to obtain equal coverage probability within the strip. Further, we see that the expected number of cross sections is independent of r, while the strip area increases with r. Thus θ has to change with r in order to obtain equal coverage probability. The remaining mathematics is given in Appendix A. It turns out that

$$r \cdot \sin\,\theta = k = \text{constant} \tag{2}$$

Thus, as r increases θ must decrease. Note that k cannot be greater than r_1, which puts a limit to the maximum number of legs before the approach breaks down. k is a crucial parameter, and it is outlined further

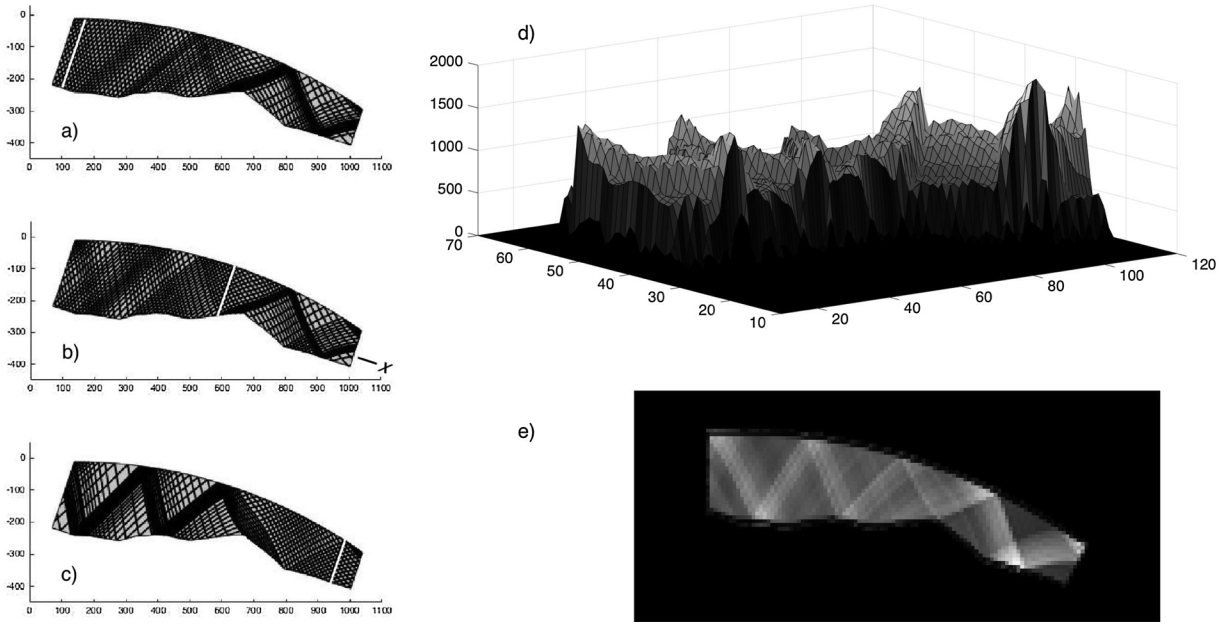

Fig. 2. Figure panels a)–c) illustrate the caustic-like effect an irregular border (the bottom one in this case) may have on transects that start with equidistant positions (see white lines) perpendicular to the design axis defined by the *x*-axis shown in b). All transects are constructed by the adjusted-angle zigzag sampler (SB), providing equal probabilities on average for any *x*-value. In the right panels the result of 10 000 simulations and random start positions within the study area is shown in terms of coverage probability figures shown as a 3D plot in d) and a 2D contour plot in e), confirming the impression of the left panel figures.

below how the *k*-value can be determined.

As shown in Appendix A, the (r,α) points on a random transect leg between r_1 and r_2 in a counterclockwise direction, starting at (r_1, α_{rnd}), can be calculated by the equation

$$\alpha(r) = \alpha_{rnd} + \sin^{-1}(k/r) + \sqrt{(r/k)^2 - 1} - \sin^{-1}(k/r_1) - \sqrt{(r_1/k)^2 - 1} \tag{3}$$

where $\sin^{-1}$ is the arcsine function, and $\alpha_{rnd} = \alpha_1 + R \bullet \alpha_c$ with R being a random variable from the uniform $U(0, \alpha_c)$ distribution. The corresponding Cartesian coordinates are $(x, y) = (r\cos\alpha, r\sin\alpha)$. The cycle, α_c, corresponding to the α-span of two legs, can be found as follows:

$$\alpha_c = 2(\alpha(r_2) - \alpha(r_1)) = 2\cdot\left(\sin^{-1}\frac{k}{r_2} + \sqrt{\left(\frac{r_2}{k}\right)^2 - 1} - \sin^{-1}\frac{k}{r_1} - \sqrt{\left(\frac{r_1}{k}\right)^2 - 1}\right) \tag{4}$$

A set of increasing *r*-values from r_1 to r_2 are chosen, and the corresponding α-values are calculated from Eq. (3). Once the first leg is calculated, the next leg starts at $(r_2, \alpha_{rnd} + \alpha_c/2)$ and stops at $(r_1, \alpha_{rnd} + \alpha_c)$ with successive *r*-values in decreasing order from r_2 to r_1, corresponding to mirroring the first leg around the radial that passes through $(r_2, \alpha_{rnd} + \alpha_c/2)$. Now one cycle is created and this is repeated at both sides till the whole sector span (α_1,α_2) is covered. To keep the points in an appropriate order, just sort all points by increasing (decreasing) α-values if a counterclockwise (clockwise) order is wanted.

We now focus on how an appropriate *k*-value is found. Let *L* denote the total transect distance available, and note that for a given *k*, the total distance will change somewhat from one realization to another. One should therefore run a sufficient number of simulations to calculate the statistics of the resulting distances and adjust the *k*-value until a satisfactory result is obtained. A simple approach is to start with the maximum value $k_{init} = r_1$, and then adjust the *k*-values after n_{sim} simulations. An efficient approach in the simulations is to replace the

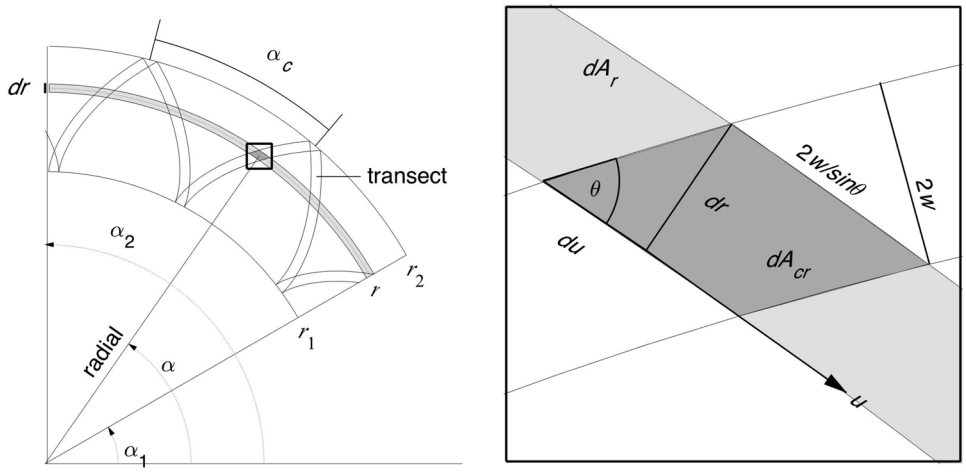

Fig. 3. Illustration of the concept for a randomized transect within a circular sector with equal coverage probability everywhere. Now it is the ratio between the area, dA_{cr}, covered by the sampler within a thin circular sector strip with area $r(\alpha_2 - \alpha_1)dr$ that must be a constant independent of *r*. See text for further details.

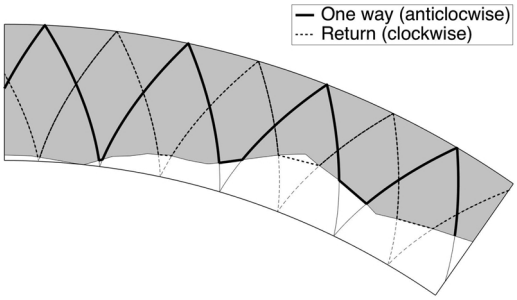

Fig. 4. An example of a return zigzag transect, from a sampler with guaranteed equal coverage probability everywhere (the thick lines), with the unwanted transport indicated by the thin lines. In this example the clockwise return zigzag (dashed curve) is translated a half α-cycle compared to the anticlockwise zigzag (solid curve).

random α_{rnd} values with equidistant values, where the difference between succeeding values is α_c/n_{sim}. The k-value is adjusted until e.g., the maximum distance, or an appropriate quantile, does not exceed L.

Instead of choosing $k_{init} = r_1$, an appropriate initial value can be found numerically as described in Appendix A.

A return zigzag design with equal coverage probability is easily implemented during each simulation by letting the return zigzag be the same as the one described before with a deliberately chosen α-difference, $\Delta\alpha_{ret}$, e.g., $\Delta\alpha_{ret} = \alpha_c/2$, but with the opposite order of the transect points (Fig. 4). There will be an "unwanted" transport along one of the sector edges to join the forward and backward transects that must be excluded in the analysis. An interesting feature of a return design is the possibility to examine possible population dynamics during the survey by e.g., comparing the variograms (Petitgas, 1993) at small scale and short time intervals along the transect with the variograms at small scale but large time intervals at the crossing points between the forward and backward transects. If, for example, small scale – small time intervals reveal strong spatial correlations that disappear at small scale with large time intervals, this is a strong indication that the population has moved during the survey. To the contrary, if the variograms look the same they might be applied to get reasonable model-based estimates for the variance of the abundance estimator.

2.3. Implementation of the equal coverage probability sampler to realistic domains

In the previous section it was outlined how one can construct random zigzag transects in an idealized circular sector that guarantees equal coverage probability everywhere. If we follow this recipe for any circular sector enclosing a realistic area, equal coverage probability is obviously obtained also for the enclosed area. However, the original parts of the transect outside the realistic area represent a waste of time and can be shortened, often considerably, by taking the fastest straight-line transport from each point where a leg leaves the area to where the next enters.

The same considerations as above are valid for the equal-spaced zigzag sampler proposed in SB applied to a perfect rectangle enclosing a real study area. For simplicity, let the rectangle have horizontal and vertical edges with the x-axis being horizontal, and let x_{min} and x_{max} denote the minimum and maximum x-values for the rectangle. A simple approach that guarantees equal coverage probability is then to let the initial transect value be $(x_{min} + x_{rnd}, y_{min})$ with x_{rnd} being uniformly distributed $U(0,2\Delta x)$ where Δx is the x-spacing between two succeeding legs and $2\Delta x$ thus is the zigzag x-period. In this case our rectangle approach and the SB equal-spaced approach coincide.

Note, however, that the SB paper does not specify clearly the randomization part for the equal-spaced zigzag sampler, and it is easy to believe that the initial x-values can be taken randomly between $x_{\min}$ and $x_{\max}$ following a uniform distribution. If this is done, however, equal coverage probability is not obtained, unless $x_{\max} - x_{\min}$ is exactly a multiple number of the x-distance between two successive legs. This point is not obvious but is easy to verify by simulations.

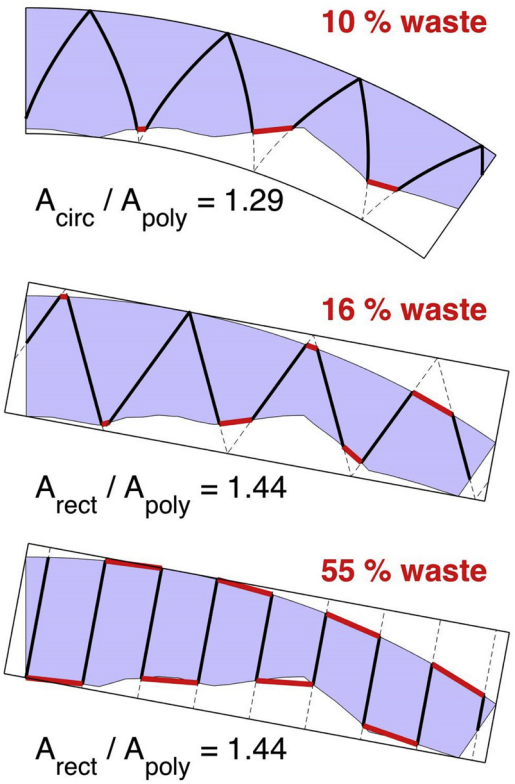

Fig. 5. An example of comparison between the parallel, rectangular and circular equal coverage probability approaches applied to the Antarctic example. The figure illustrates a general trend that the zigzag designs with equal coverage probabilities are far more efficient than a parallel design, and in many cases the circular approach is also considerably more efficient than the rectangular approach.

A common challenge for the circular as well as the rectangular enclosing approaches is to find the optimal circle sector or rectangle. A natural approach is to minimize the enclosing area, which is equivalent to minimize the ratio between the area of the enclosing and the real domain. Some results are seen in Fig. 5.

For the rectangle area, there is only one unknown parameter, which is the rectangle orientation (rotation). For a given orientation it is straight forward to find the rectangle that at each of the four edges touch the enclosed area at least at one of the polygon points without intersecting the area. Then one can search for the rotation angle giving the smallest "touching" area. The results in this paper are found efficiently by an algorithm that applies the Matlab m-file fminsearch.m (m is the extension of file names for command scripts in Matlab).

For the optimal enclosing circular sector, the circle origin is the unknown, i.e. there are two unknown parameters in this case, x_0 and y_0. For each choice it is easy to find the two (straight line) radial parts of the sector touching the realistic area without intersecting it, and in the same way the circular part with maximum radius. To find the optimum circular part with minimum radius, however, the circle with radius equal to the smallest distance from the origin to the border polygon points may still intersect the area. Thus, also the shortest distances from the origin to the straight lines between two succeeding polygon points have to be checked out as the possible smallest radius for a circle that still hits the polygon domain without intersecting it.

Also, for the circular sector results in this paper, the optimal origin is found automatically by applying the fminsearch m-file in Matlab. To succeed, however, the choice of a reasonable initial origin might be

necessary. One way to do this is to manually choose 3 points in the area that appear to lie on an appropriate circle reflecting the gross shape of the area and use the origin for the circle going through these 3 points as an initial value.

2.4. Simulation of coverage probability

First the border polygon values, (x,y), and the available distance to travel, L, are scaled by a common factor so that the area consists of an appropriate number of square pixels (cells) with size one. This is a balance between high resolution and extensive computer time on one side, and fast analysis and coarse resolution on the other. Then the distance between succeeding transect points is chosen to be smaller than 1, e.g., 0.99. For each simulation each of the pixels hit get a weight equal to the length of the transect intersecting the pixel. Then n_{sim} simulations are run, and the accumulated weights are successively calculated in each simulation. Note that this is different from just counting the number of times each pixel is hit, because the probability of hitting a pixel is strongly dependent on the transect direction with regard to the pixel orientation.

An alternative and simpler option (SB), which is more appropriate for straight line transects, is to model the transect legs as "rectangles" with width $2w$, grid the area and accumulate the number of times each grid cell node ("point") is covered by the transect over the simulations. For an equal coverage probability design the number of scores should converge to the same number as the number of simulations increase. This approach, however, is more time demanding for curved legs, because these need to be approximated by far more points than the few points needed to define the straight lines.

2.5. The Horvitz-Thompson estimator applied to acoustic abundance data

The HT estimator applied to minke whale counting surveys in the Antarctic is well described in SB. Here we focus on the HT estimator applied to a classical acoustic fish survey with a downward looking echo sounder mounted on the bottom of a ship hull. We assume that the echo beam has a constant geometry with a known efficient beam width, $2w$, that will increase with depth. Thus, strictly speaking, this is in conflict with the assumption of a constant transect width. We can imagine, however, that the water is partitioned in successive depth layers where the beam width is approximately constant in each layer. If we have an unbiased estimator for the abundance in each layer, we will have an unbiased estimator for the whole water column as well.

Let the study area be gridded in N quadratic cells so small that we can ignore the border effect that only a random portion of the cells hitting the transect border will be included by the transect. Further, let n denote the number of cells included in the transect, and let ρ_i denote the density of a target species in cell i in terms of biomass per square nautical miles, kg/n.mi.^{2}. The HT estimator, B_{HT}, for the total biomass, B, is now

$$B_{HT} = \left(\frac{1}{N}\sum_{i=1}^{n}\frac{\hat{\rho}_i}{\pi_i}\right)\cdot A \tag{5}$$

where $\hat{\rho}_i$ is the estimated density in cell i from the acoustic records and π_i is the probability of cell i being covered by the transect and A is the area of the study area. For an equal coverage probability design the HT estimator is simplified to

$$B_{HT}^{eq} = \left(\frac{1}{m}\sum_{j=1}^{m}\hat{\rho}_j\right)\cdot A = \bar{\hat{\rho}}_j\cdot A \tag{6}$$

i.e., the average density estimate, over m successive parts of the transect of equal length, e.g., 1 n.mi., multiplied by the study area. If each local density estimate is unbiased, this simple HT estimator will provide a design-unbiased estimator for the true abundance, independent of the spatial structure and possible hot spots in the spatial distribution of the target animals. Design-unbiasedness means that the average abundance estimate over several simultaneous random surveys would converge towards the true abundance.

To illuminate the assumption of unbiased local density estimates, let $\hat{\rho}_{tr} = (\hat{\rho}_1,\ldots,\hat{\rho}_m)$ be a random transect observation vector where the whole transect is partitioned in m successive parts of equal length (and area), each represented by a local estimate of biomass density. Assume that we can ignore the uncertainty of this local estimate at the time of observation. Further, let $f_{\boldsymbol{x},t}(\hat{\rho}_{tr})$ denote the distribution of observed densities as a function of space, $\boldsymbol{x} = (x,y)$, and time, t. Suppose local stationary conditions in the sense that for a random survey in the same study area during the same survey time, T, the corresponding $\hat{\rho}_{tr}$ can be considered a random sample from the $f_{\boldsymbol{x},t}$ - distribution. Different samples will then be independent with regard to the $f_{\boldsymbol{x},t}$ - distribution, and unbiased local estimates follow, whatever spatial correlations that are involved. The assumption will be violated in case of a sufficiently large drift of the population during the survey, which is a challenge for a parallel design as well. In this case, however, a return zigzag is an appropriate design to reveal such a drift, as pointed out in the main text.

Though it is straight forward to construct an exact analytic expression for the design-based variance of the HT-estimator (Thompson, 2002), it is hard to find appropriate estimators (Murthy, 1957; Brewer and Hanif, 1983, pp. 90–91). In practice other options are often applied, as described in SB with references (Borchers et al., 1998; Marques and Buckland, 2003). One appealing approach is to treat each transect leg as a primary sampling unit, and bootstrap among the legs. A challenge by zig-zag transects is that positive correlations are often present at the turning point between two successive legs, thus causing neighbor leg observations to be dependent. If, on the other hand, the observations reveal, e.g., stationary conditions in a geostatistical sense (Chilès and Delfiner, 1999), the zigzag provides good directional information at a range of different scales, which is useful to get good estimates of variogram parameters. In such a case it is straight forward to estimate the estimation variance of the HT-estimator based on the estimated variogram (Chilès and Delfiner, 1999).

There is a range of bias issues related to the local density estimates, such as how fish reacts to the echosounder and the uncertainty of discriminating between different species. This represents challenges for any survey design, and is outside the scope of this paper.

2.6. Computer efficiency

For the curved zigzag, the most time-consuming design, the number of points for L_{tot} was about 1000 for each simulation. To run 1000 simulations, the computing time was c. 1 min with a Macbook Pro with a 2.2 GHz Intel Core i7 processor and 16 GB MHz DDR memory. All scripts are written in Matlab and were run by Matlab R2015b.

3. Results

3.1. The Antarctic example with the SB adjusted-angle zigzag sampler

In this case the adjusted-angle zigzag sampler introduced by SB is applied, which ascertains that the average coverage probability in strips perpendicular to the design axis indicated in Fig. 2b is independent of position along this x-axis. In Fig. 2 each of the 3 different left panels (a–c) show 20 transects, 10 at each of the two possible θ-values, that are initiated with equidistant starting points along the axis perpendicular to the design axis, marked by white lines. In all cases we see that as the transects move away from their starting points, they develop a clustering behavior. We also see that the patterns are quite different in the 3 cases.

In the right panels (d and e) in Fig. 2, a 3D coverage probability plot based on 1000 simulations are shown in the upper panel (d), where for

each simulation a random point in the region is chosen as an initial value. We clearly see a non-constant probability surface, which is further illustrated by the 2D contour plot in the lower panel. The major reason for this deviation is in the author's opinion the irregular shape of the southern border. This is in a sense analogous to the physical phenomenon called[1] caustics, which can be seen, for example, as concentrated bright curves on the bottom in a swimming pool (see Fig. 6) caused by the irregular water surface that cause incoming light rays to cluster.

Obviously, the substantial deviation from equal coverage probability described above is not a wanted feature of a design with the intention of providing equal coverage probabilities and may lead to severely biased abundance estimates if equal coverage probabilities are assumed. To reduce such unwanted effects, one of 3 options suggested in SB is to apply a convex hull surrounding non-convex areas. This would probably cause much less deviation from equal coverage probability for many different convex hulls. In fact, an enclosing rectangle, as applied in the present paper, is the only option among all enclosing convex geometries that can guarantee perfect equal coverage probability.

3.2. The Antarctic example with enclosed circular sector and the proposed equal coverage probability zigzag sampler

Fig. 4 shows the original Antarctica region with an enclosed circular sector and an example of a return zigzag that guarantees equal coverage probability in the study area. In this case it is natural to let the circular upper border of the sector be equal to the original upper border, though the sector does not necessarily become the one with smallest area. Note also that the sector end edges between the inner and outer radii need to be perpendicular to the sector circles at the intersection points between the edges and the circles. Thus, some deviation from the original borders will in general appear also at the sector edges, though not in this case because longitudes are perpendicular to latitudes.

The result of 10,000 simulations following the recipe for the equal coverage probability sampler is illustrated in Fig. 7, except that in this case the initial angle α_{rnd} is changed by exactly the angle span of one cycle, $\Delta\alpha$, divided by 10,000 from one simulation to the next. In the upper panels the results of the robust simulation method (W) are shown, and in the lower panels are shown the results by the naïve counting approach (N). The nice results by the robust simulation method confirm that the sampler provides equal coverage probabilities as expected, and also that the robust simulation method works. The results by the naïve simulation approach clearly illustrate that this is not the appropriate way to simulate coverage probabilities in this case.

A typical example of applying the random parallel, rectangular zigzag and circular zigzag approaches are shown in Fig. 5. Over 1000 simulations, the average relative waste, i.e., ratio of distance traveled off-effort and on-effort, was 9.5%, 15.7% and 57.5%, for the curved zigzag, straight line zigzag and parallel zigzag approach, respectively. Note that the straight-line zigzag approach is synonymous with the equal-spaced zigzag sampler in SB with a rectangle as an enclosing non-convex hull, while the more efficient curved zigzag has the convex circular sector as the enclosing hull.

In the simulation runs the area is gridded so that the total study area contains 2759 pixels. To obtain real scale values we can use the fact that the upper border of the study area is along the 64.5 degrees latitude south. Set the earth radius to 6400 km. The radius of the circle defining the 64.5 degrees latitude is then 6400 km·cos(64.5°) = 2755 km. On our gridded scale, the corresponding radius is r_2 = 202.51 pix, so we get a factor 2755 km/202.51 pix = 13.60 km/pix to transform from our pixel scale to real km's. So, as an example, an average L_{eff} = 178 pix corresponds to 178pix·13.60 km/pix = 2421 km.

[1] "Caustics are natural phenomena in which nature concentrates the energy of waves.", ref: https://arxiv.org/abs/1706.01589.

Fig. 6. Illustration of the physical caustic phenomenon. The irregular water surface border cause light rays to cluster in the bright irregular curves seen at the bottom of a swimming pool in this case.

3.3. The North Sea sandeel fields

Fig. 8 shows some sandeel fields where the SB equal-spaced zigzag has been applied at many of the fields for several years. We also see examples of the return zigzag design and the parallel design. Based on the same effort as applied in the abundance surveys, the efficiency of the parallel (P), rectangular zigzag (R) and curved zigzag (C) equal coverage probability alternatives are compared by simulations. The rectangular zigzags are constructed based on the equal-spaced zigzag described in SB for the enclosing rectangle. Based on 10,000 simulations, the average distance travel results in terms of wasted divided by inside area travel in percent are shown in Table 1. In 7 of the cases, all approaches worked with an average of 28% waste for P, 9% for R and 6% for C. As we see, the zigzag designs appear to be considerably more efficient than the parallel design, and most efficient for the circular enclosing approach, though the latter could not be applied in all cases due to a too "fat" minimal circular enclosing sector.

4. Conclusion

A zigzag sampler with guaranteed equal coverage probabilities is constructed for a pure circular sector, with an exact analytical expression for the curved transects so that no iterative procedure is needed to calculate the transect points. An equal coverage probability is therefore obtained for any domain enclosed by a circle sector by applying the circle sector sampler. This is also obtained by the equal-spaced zigzag sampler (SB) for a rectangle and thus any area enclosed by the rectangle.

The cost of the approach is the unwanted straight-line transport between the points where a circle sector transect leg leaves the study area to the point where it re-enters the study area. In a practical example from a whale counting area in the Antarctic, this cost was on average about 9.5% of the transect effort inside the area, as compared to 57.5% when using a randomized parallel design.

The paper outlines how one, for a given study area, can construct algorithms that automatically finds the enclosing rectangle as well as the enclosing circular sector with minimum area.

The computer efficiency of the constructed zigzag sampler and the coverage probability simulation method is high, where 1000 simulations for realistic examples require in the order of 1 min on a modern laptop.

5. Discussion

The improved efficiency obtained with the zigzag approach compared to the parallel approach depends on the shape of the study area as well as the available survey effort. The greater effort (more legs in the zigzag survey design) the minor drawback will be the unwanted travel distance relative to the transect length, for both parallel and zigzag designs. When the unwanted transport becomes a challenge, this will be due to deviation in shape from the rectangle or circular sector, and few legs will easily provide a substantial variety in needed effort from one

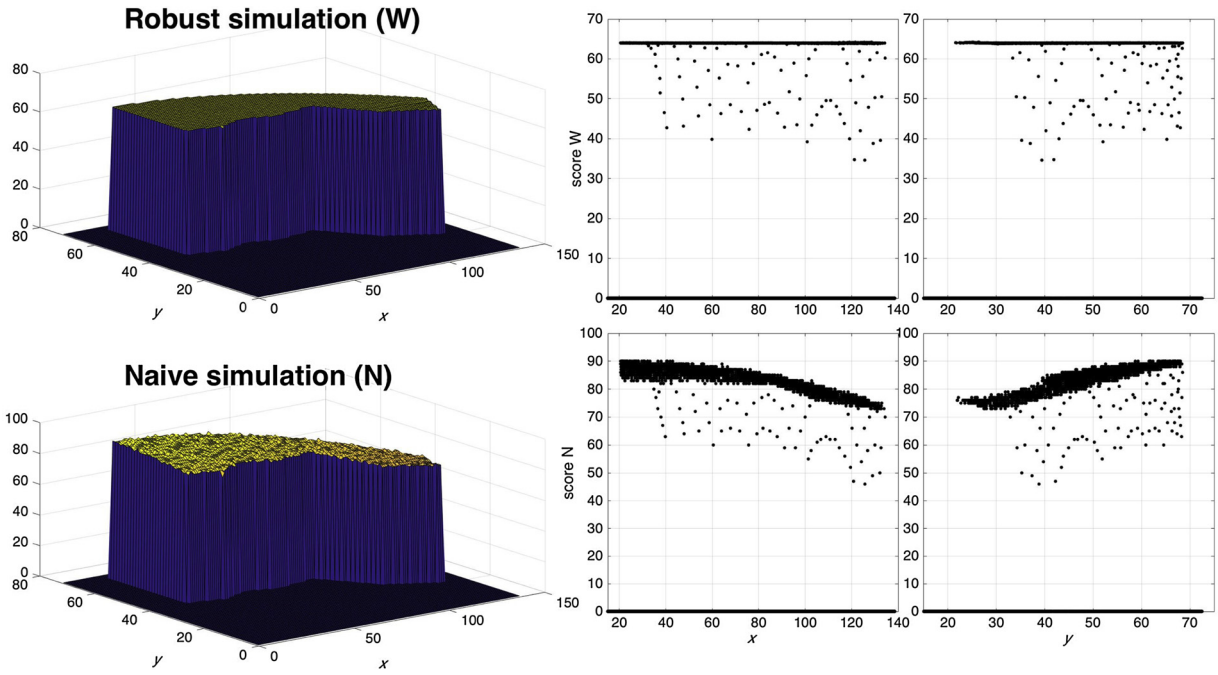

Fig. 7. The results of n_{sim} = 10,000 simulations of the exact equal coverage probability applied to the Antarctic study area with constant $\Delta\alpha_{sim} = \Delta\alpha/n_{sim}$ between succeeding initial azimuth angles. From the upper panels with the robust simulation approach (W), hardly any deviation from a constant value is seen. The sporadic values between 0 and maximum y-value are just border effects. The lower panels clearly show a deviation from equal coverage probabilities by applying the naïve approach of counting the number of pixels hit (N).

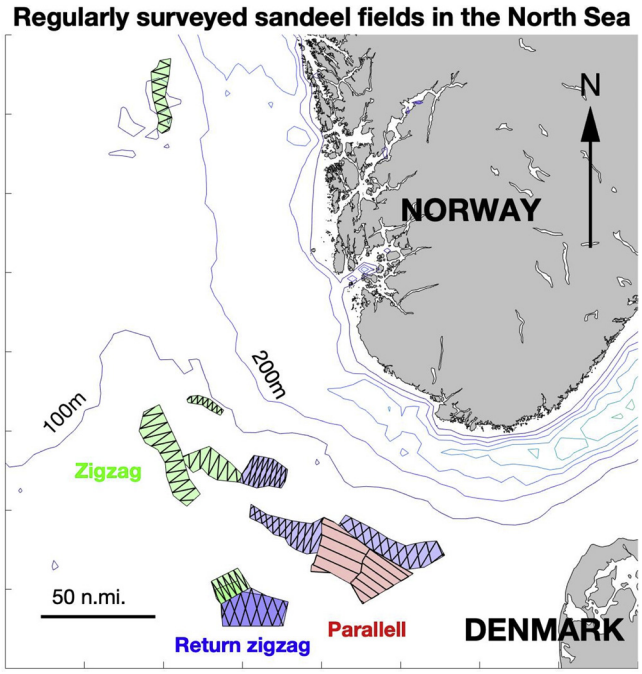

Fig. 8. Actual sandeel fields and real acoustic transects in the North Sea with the SB equal-spaced zigzag design in several fields. Note that parallel and return zigzag are conducted as well.

random transect to another. This and other features can efficiently be studied by simulations in a concrete case at hand.

Another challenge with the circular sector sampler is that it breaks down for a sufficiently large effort and a sufficiently "thick" sector in terms of the ratio between the difference of the sector radii and the smallest sector radius. This can be overcome by moving the circle origin further away from the study area, thereby making the enclosing circle more "rectangle-like". This will, however, increase the off-effort transport.

Table 1
Comparison of unwanted off-effort travel distance as percentage of the on-effort travel distance within 11 different sandeel fields in the North Sea, for the parallel (P), rectangle enclosing zigzag (R) and circular sector enclosing zigzag (C) equal coverage probability designs.

Field no:	P%	R%	C%
1	30	7	8
2	27	10	–
3	23	8	8
4	27	3	2
5	11	1	2
6	34	8	4
7	29	10	–
8	33	14	–
9	52	25	16
10	20	4	5
11	26	13	–
Average:	28%	9%	6%

The paper has focused on equal coverage probability. For abundance estimation purposes, equal coverage probability is not needed if the coverage probabilities are known, and, for example, a Horwitz-Thompson estimator can be applied as demonstrated in SB. Even with deviation from equal probability coverage, however, it is an advantage to have a probability that varies smoothly over the study area. If fitting of a parametric surface to the simulated probabilities is wanted it should be noted that the variance of the transect length intersecting a pixel also depends on the transect directions with regard to the pixel orientation. Though not shown here, it is quite easy to calculate these variances which might be convenient to include in the estimation of the surface parameters.

The results in this paper indicate rather strongly that a randomized design with equal coverage probability is more efficient for zigzag than for parallel transects when it comes to minimizing unwanted off-effort travel distance. This is, however, only one of many aspects to be considered in the comparison of zigzag versus parallel transects. If, for

example, the border of the study area has to be decided based on the sample observations during the survey, the parallel designs appear to be superior by just continuing the transects until the abundance disappears, without destroying the even coverage probability (Petitgas, 1993). In addition, if the population under study is strongly correlated in space, it is a general statistical result that evenly located observations minimize the variance of the abundance estimator. This might also be in favor of a parallel design, because a zigzag design will have more close and correlated observations in vicinity of the turning points. On the other hand, if modelling of e.g., parametric variograms are wanted, zigzag designs may be more informative for model selection and parameter estimation.

In general, parallel and zigzag designs each have advantages and disadvantages. Both options should be continuously examined, as new experiences and data may influence the basis on which a survey design was decided. In general, parallel designs have dominated, probably due to its convenience and easy interpretation and easily obtainable equal probability coverage. One can hope that the results of this paper contribute to look more thoroughly at zigzag designs as an alternative option.

Acknowledgement

The work was supported by the IMR Strategic project “Reduce Uncertainty in the Stock Assessment (REDUS)”. The reviewers have contributed substantially to improve the original version of the article. Jofrid Skardhamar at IMR has helped making the paper more readable.

Appendix A. The construction of zigzag transects with equal coverage probabilities everywhere within a circle sector

Note from Fig. 3 that for any narrow circular strip like the bright grey one, the strip is crossed by a zigzag transect the same number of times (on average), independent of r. From the right panel, we see that the cross-section area, dA_{cr}, is

$$dA_{cr} = (2w/\sin\theta)\cdot dr \tag{A1}$$

The total area, dA_r, within the strip is

$$dA_r = r\cdot(\alpha_2 - \alpha_1)\cdot dr \tag{A2}$$

To obtain equal coverage probability the ratio dA_{cr}/dA_r must be a constant independent of r and α. This means that

$$r\cdot \sin\theta = k = const \tag{A3}$$

Thus, as r increases, θ must decrease, and k cannot exceed r_1. It will turn out that k will increase with the travel distance L available to survey the area. A simple approach to find an appropriate k-value is to start with $k = r_1$ as described in the main text. At the end of this appendix it is described how an appropriate “guestimate” can be found automatically, but simulations should anyhow be run to be sure that an appropriate k-value is found, e.g., one that not causes a transect to exceed L.

For the time being, it is assumed that k is known. We imagine that we start constructing the transect at the inner circular boundary at a random point (r_1, α_{rnd}) with α_{rnd} being a random variate from the uniform distribution $U(\alpha_1,\alpha_1 + \alpha_c)$ where α_c is the α-cycle (Eq. (A9)) corresponding to two successive transect legs in the anticlockwise direction (one upwards and the next downward). Imagine that we move anticlockwise. Let u be a Cartesian coordinate along the axis perpendicular to the radial with positive direction clockwise. Then (see Fig. 3)

$$\tan\ \theta = dr/du \Rightarrow du = dr/\tan\theta \tag{A4}$$

By utilizing the expression for θ in Eq. (A3) the last expression above can be expressed as:

$$du = dr/\tan\ \theta = dr/\tan(\sin^{-1}(k/r)) = \sqrt{(r/k)^2 - 1}\ dr \tag{A5}$$

which leads to the following differential ds along the transect:

$$ds = \sqrt{du^2 + dr^2} = (r/k)dr \tag{A6}$$

Integrating the above expression from $r = r_1$ to $r = r_2$ gives the following expression for travelled distance, D, along one leg:

$$D = (r_2^2 - r_1^2)/(2k) = (\bar{r}\cdot\Delta r)/k \tag{A7}$$

where $\bar{r} = (r_1 + r_2)/2$ and $\Delta r = r_2 - r_1$. To find the correspondence between r and α, we utilize the relationship $du = \pm r\cdot d\alpha$ where the plus sign corresponds to an anticlockwise, the minus sign corresponds to a clockwise propagation, and dα is the span of α-values covered by du. We apply the expression for du in Eq. (A5), and integrate from $\alpha = \alpha_{rnd}$ to α and from $r = r_1$ to r to get

$$\alpha(r) = \alpha_{rnd} \pm \sin^{-1}(k/r) \pm \sqrt{(r/k)^2 - 1} \mp \sin^{-1}(k/r_1) \mp \sqrt{(r_1/k)^2 - 1} \tag{A8}$$

where the plus/minus sign corresponds to an anticlockwise/clockwise propagation, respectively.

By now choosing a set of increasing r-values from r_1 to r_2, and calculating the corresponding α-values, the resulting points (r,α) will be succeeding points on the first leg from (r_1, α_{rnd}) to $(r_2, \alpha_{rnd} + \alpha_c/2)$, the latter point being the starting point for the next “downward” leg. By now choosing decreasing r-values from r_2 to r_1 and calculating the corresponding α-values, the next leg and thus a whole cycle of the transect is determined. This cycle is now repeated by translating the α-values by $\pm n\cdot\alpha_c$, with $n = 1,2,\ldots$ until the whole span (α_1,α_2) is covered. When all the transect points within the circular sector are found, they can be ordered according to increasing (decreasing) α-values if an anticlockwise (clockwise) direction is wanted.

The Cartesian coordinates corresponding to (r,α) are $(x,y) = (r\cdot\cos\alpha, r\cdot\sin\alpha)$.

The transect cycle, α_c, can be calculated from Eq. (A8) as $2(\alpha(r_2) - \alpha(r_1))$:

$$\alpha_c = 2\cdot\left(\sin^{-1}\left(\frac{k}{r_2}\right) + \sqrt{\left(\frac{r_2}{k}\right)^2 - 1} - \sin^{-1}\left(\frac{k}{r_1}\right) - \sqrt{\left(\frac{r_1}{k}\right)^2 - 1}\right) \tag{A9}$$

With a distance L available, this corresponds to $n_{leg} \approx L/D$ legs, where the leg length D is given by Eq. (A7). This is also approximately equal to the total α-span $\alpha_2 - \alpha_1$ divided by half the cyclic period, $\alpha_c/2$ (see Eq. (A9)). Solving the equation

$$L/D - 2\cdot(\alpha_2 - \alpha_1)/\alpha_c(k) = 0 \quad \text{(A10)}$$

with regards to k, will provide an appropriate initial value for k.

References

Borchers, D.L., Buckland, S.T., Goedhart, P.W., Clarke, E.D., Hedley, S.L., 1998. Horvitz-thompson estimators for double-platform line transect surveys. Biometrics 54, 1221–1237.

Brewer, K.R.W., Hanif, M., 1983. Sampling with Unequal Probabilities. Springer-Verlag, New York.

Chilès, J.P., Delfiner, P., 1999. Geostatistics. Modeling Spatial Uncertainty. John Wiley & Sons, Inc., New York.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 47, 663–685.

Marques, F.F.C., Buckland, S.T., 2003. Incorporating covariates into standard line transect analyses. Biometrics 59, 924–935.

Murthy, M.N., 1957. Ordered and unordered estimators in sampling without replacement. Sankhyā 18, 379–390.

Petitgas, P., 1993. Geostatistics for fish stock assessments: a review and an acoustic application. ICES J. Mar. Sci. 50, 285–298.

Rivoirard, J., Simmonds, J., Foote, K.G., Bez, N., 2000. Geostatistics for Estimating Fish Abundance. Blackwell Sciencehttps://doi.org/10.1002/9780470757123.

Strindberg, S., Buckland, S.T., 2004. Zigzag survey designs in line transect sampling. J. Agric. Biol. Environ. Stat. 9 (4), 443–461. https://doi.org/10.1198/108571104X15601.

Thompson, S.K., 2002. Sampling, 2nd edn. John Wiley and Sons, New York.