Scottish Environment Protection Agency
Buidheann Dìon Àrainneachd na h-Alba

Marine finfish farm regulation

Seabed mixing zone limit

Compliance assessment methodology

Scottish Environment Protection Agency

May 2022

Every day SEPA works to protect and enhance Scotland's environment, helping communities and businesses thrive within the resources of our planet.

We call this **One Planet Prosperity**

1 Introduction and permitted area

1.1 The environmental standards with which operators must comply are specified in the farm's permit.

1.2 All permits will include a requirement to meet two standards for the biological condition of the seabed. One of the standards concerns the size of the observed mixing zone, relative to its maximum allowed area. The other must be met within the mixing zone at the outer edges of the pens. It is the methods used to assess the area-based standard which is the focus of this document.

1.3 The maximum permitted area of the mixing zone is specified in the farm's permit. It is equivalent to size of the non-overlapping area lying within 100 metres of pens in all directions (see Figure 1(a)). The permitted area is therefore a function of the pen dimensions, the number of pens comprising the farm, and their respective positioning.

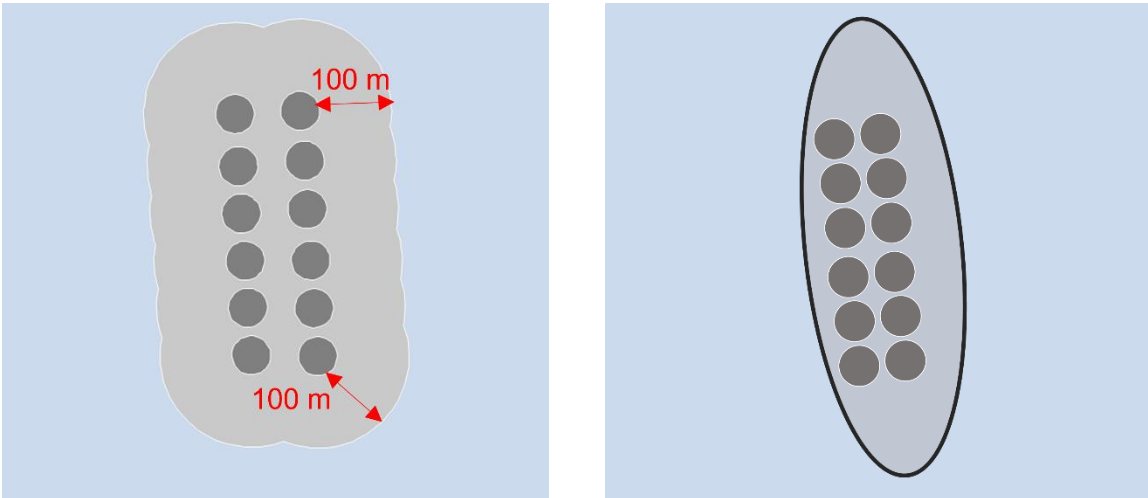
1.4 The shape of a mixing zone around a farm is affected by local characteristics (e.g. the strength and direction of the bed current) that determine how far deposited matter spreads across the seabed in different directions. To reflect this, a mixing zone may extend more than 100 metres in some directions provided its area does not exceed the maximum permitted area specified in the farm's permit (see Figure 1(b)).

1.5 The biological standard applicable to all soft sediments is 0.64 and above as measured on the infaunal quality index (IQI; set out in the 2014 Standards Directions[1]). An IQI of 0.64 represents the good/moderate quality boundary. The observed mixing zone is defined as the area of seabed degraded to an IQI of less than 0.64 (i.e. to less than good quality).

[1] https://www.gov.scot/publications/implementing-water-environment-water-services-scotland-act-2003-assessing-scotlands/

1.6 A statistical model has been developed by SEPA which uses IQI monitoring from the vicinity of a marine finfish farm to assess the size of the observed mixing zone area, whilst also quantifying the uncertainty around that estimate. This observed mixing zone is then compared against the permitted area. This document describes that model and methodology.

1.7 This statistical model has been tested successfully against a series of real-world datasets.



(a) **(b)**

Figure 1: Schematics showing: (a) a farm's permitted mixing zone, based on a 100 m buffer projected from pen edges; and (b) the observed mixing zone, based on IQI monitoring. Note that the observed mixing zone may be offset from the farm's centre.

2 Model overview & data inputs

2.1 The goal of the statistical model described here is to use the monitoring data to estimate the footprint area, allowing it to be compared against the permitted area.

2.2 Further information on how a monitoring survey should be conducted is detailed separately in the Environmental Monitoring Plan (EMP) guidance[2].

2.3 Each aspect of the model is described in greater detail in remainder of the document, although a high-level view of the main steps is provided here (see Figure 2).

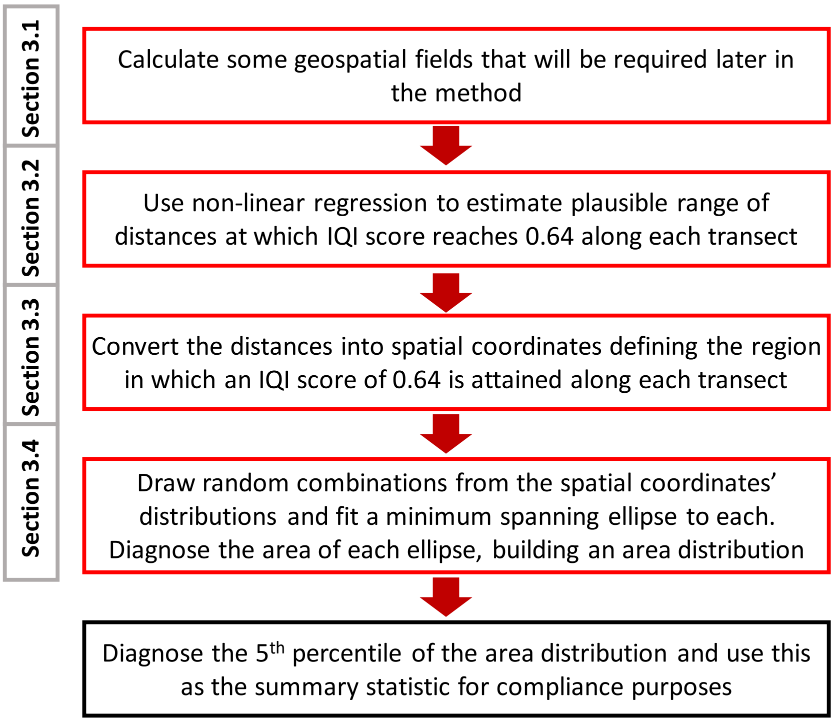


Figure 2: Overview of the modelling process

2.4 A farm's observed mixing zone area is calculated where possible by collecting IQI data along a minimum of 4 transects, each comprised of a minimum of 7 stations, radiating out from the fish farm in different directions (comprising a minimum of 28 stations). The Environmental Monitoring Plan Guidance provides information on how to calculate the area when it is not possible to collect IQI data from all transects.

[2] Scottish Environment Protection Agency (2022) *Marine Finfish Farms: Seabed environmental standards - Demonstrating compliance.* Accessible at: https://www.sepa.org.uk/regulations/water/aquaculture/pre-application/

2.5 The spatially structured monitoring points are appropriately oriented and scaled to capture the distance to the IQI for good quality along each axis of the farm's expected depositional footprint.

2.6 With increasing distance from the pen edge, it is expected that IQI will tend to increase (although every IQI result is also subject to random noise). The distance from the fish farm at which IQI reaches 0.64 is of primary interest. It is expected that the monitoring submitted will clearly demonstrate a return to good biological quality along each transect.

2.7 The limited number of IQI stations means that the estimate of this distance will be uncertain, to some extent. Non-linear regression is used to determine a plausible range of distances at which IQI attains a value of 0.64, and hence provide a measure of that uncertainty.

2.8 A method of using these distance estimates to infer a footprint area is required, so that compliance can be assessed against the permitted area. Such an approach requires some assumption as to the geometric shape of the footprint.

2.9 The simplest and most flexible assumption for this shape is an ellipse. It is also consistent with the view that a farm's discharge may be expected to spread further along the axis of the tidal flow than perpendicular to the flow.

2.10 Under this elliptical footprint assumption, the final step is to take the distances/spatial positions at which IQI attains 0.64 and use these to construct a series of ellipses which just enclose each set of positions. Such an ellipse is known as a minimum spanning ellipse. The area of each of these minimum spanning ellipses is diagnosed, yielding a distribution of footprint areas, commensurate with the uncertainty in the distance to a stable IQI condition of 0.64 or better along each transect.

2.11 When using this method of assessing compliance, SEPA will determine a farm to be non-compliant with the mixing zone standard when it is confident (95% certain) that the observed mixing zone area exceeds the maximum permitted mixing zone area. The means by which this is done is by comparing the 5^{th} percentile ($P_{5\%}$) of the distribution of estimated footprint areas against the maximum permitted mixing zone area.

2.12 The following sections will describe the above in greater detail.

2.13 In due course, SEPA is aiming to provide an online calculator, which allows the input of IQI data by finfish farm operators and then provides estimation of the observed mixing zone area.

3 Model description

The model uses several scripts, run sequentially, to move from input IQI data to an estimate of the size of a site's footprint. These scripts are detailed below.

3.1 Geospatial precursors

3.1.1 Later parts of the method rely on certain geospatial information. These are derived for each transect here.

3.1.2 The calculated fields include:

- The best-fit bearing for each transect
- Distances of each station from pen edge
- Spatial separation between stations
- Whether the scatter of stations around the best-fit bearing remains within acceptable constraints

3.1.3 A summary table describing this information is outputted by the model. In addition, some of these calculated fields are used in later parts of the model.

3.2 Probabilistic regression approach

3.2.1 The goal of this part of the model is to use the IQI monitoring to estimate the distance along each transect at which the IQI attains a value of 0.64. These distances are used later in the method to quantify a farm's mixing zone.

3.2.2 Although it may be reasonable to assume that true IQI increases monotonically with distance from the farm, this is not always true of real-world data since all measurements are subject to random variation.

3.2.3 The approach used here is to fit a non-linear model that adequately describes how the IQI increases with distance from the pen edge along each transect. Each transect is fitted separately.

3.2.4 This regression model can then be used to interpolate between observations and estimate the distance at which the mean IQI would equal 0.64.

3.2.5 The first step is to establish the distance to good quality, based on the reduced analysis approach (this is the distance to the first of two consecutive stations with an IQI of 0.64 or above; see the EMP guidance for further details[2]). This may be used in cases where this reduced analysis approach has been explicitly chosen by the operator to reduce analysis costs; it also acts as a useful fallback in cases where the data submitted are unsuitable for fitting regression models to.

3.2.6 An initial non-linear model is now fitted to the data. This is only a first attempt for reference purposes and may be superseded later in the process. The initial, and subsequent, models are fitted using the DRC package[3].

[3] Ritz C., Baty F., Streibig J.C., Gerhard D. (2015) *Dose-Response Analysis Using R.* PLoS ONE 10(12): e0146021. https://doi.org/10.1371/journal.pone.0146021

3.2.7 A simpler, linear regression is also fitted, acting as an informal lack-of-fit test. Experience has shown that, for well-planned transects of appropriate length, it is highly unusual for a linear regression to out-perform a non-linear one. If it does, a warning is returned in the output summary table.

3.2.8 Several alternative non-linear models are now used to compare against the initial one. The models which will be used are:

- A three-parameter Michaelis-Menten function
- A four-parameter logistic function
- A five-parameter logistic function

3.2.9 The forms of these models are detailed in the appendix.

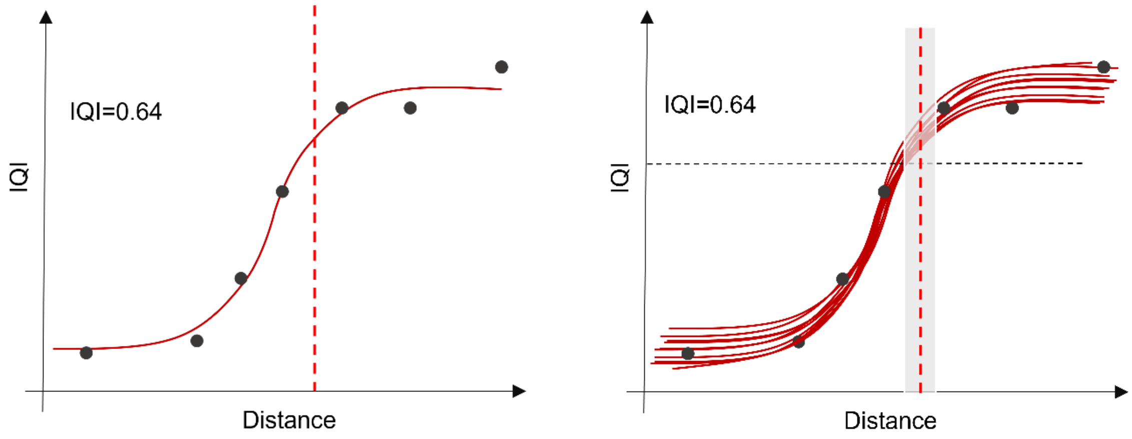
3.2.10 For the purposes of model comparison and final selection, these models are ordered using the Akaike Information Criterion[4] (AIC). This metric is used to identify the best performing model overall.

3.2.11 Since it is important to balance model complexity against parsimony, the best performing model is then compared against simpler models (with fewer parameters). If a simpler model does not perform significantly worse than the best fitting one ($\Delta AIC \leq 10$), this model is preferred.

3.2.12 When trying to estimate any quantity, it is important to adequately acknowledge uncertainty. It is accepted that different ways exist to quantify uncertainty around model fits. Resampling the residuals between the best-fit curve and monitoring data is the approach favoured here. This generates a series of new datasets to which the model can be refitted. This approach has the advantage of performing well on non-linear models fitted to a relatively small number of data points[5].

[4] Lambert, B. (2018). *A student's guide to Bayesian statistics* (p. 231). Sage.
[5] Ritz, C., Streibig, J. C., & Streibig, J. C. (2008). *Nonlinear regression with R* (Vol. 10; p. 96). New York: Springer.

3.2.13 Within the present context, this means that rather than relying on a single best-fit estimate for the distance at which IQI reaches 0.64, we perform each regression multiple times on resampled data. This generates a series of curves, around the best-fit estimate (see Figure 3).

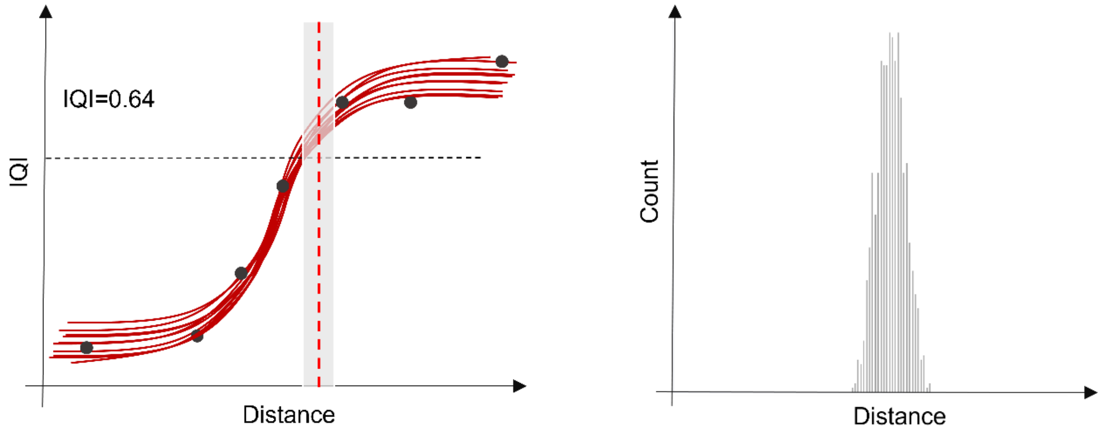


(a) **(b)**

Figure 3: Approaches to interpolating between data points to determine distance(s) at which IQI reaches 0.64. In (a), a single best-fit curve is fitted, which does not quantify the uncertainty around that fit. A single distance to IQI=0.64 is estimated (red, vertical line). In (b), the data are resampled to generate a series of curves (a spaghetti plot). These curves yield a region at which IQI is likely to reach 0.64 (grey, shaded area).

3.2.14 The data are resampled 500 times – this number being chosen since it was sufficient for the test datasets' compliance result to converge, whilst avoiding unnecessary computational expense.

3.2.15 The distance at which IQI on each of these curves reaches 0.64 is diagnosed. The result of this approach is a likelihood distribution for the quantity of real interest - the distance to good quality for each transect (see Figure 4).

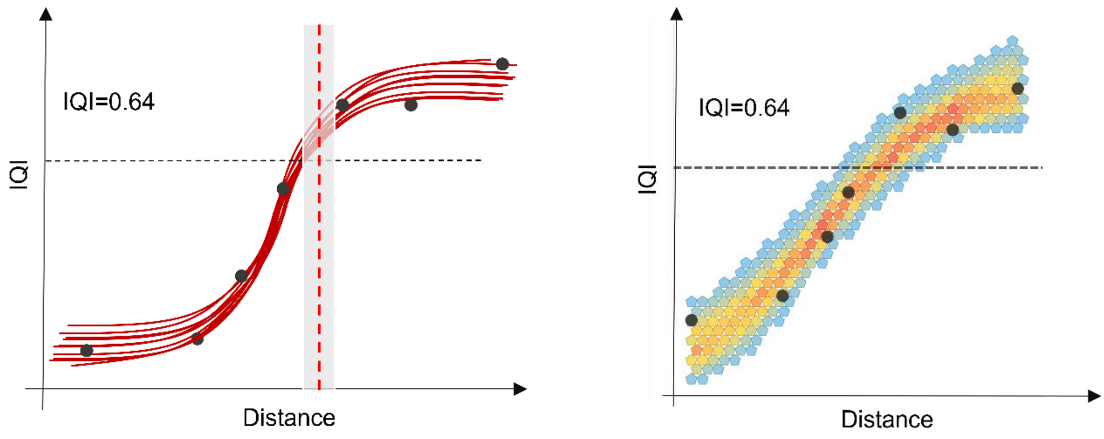


(a) **(b)**

Figure 4: Regression curves based on resampled data, and corresponding histogram of distances at which IQI reaches 0.64.

3.2.16 The percentage of these curves which reach 0.64 is also checked. In cases where this is less than 100%, ambiguity remains as to whether a good quality condition has been achieved. Consequently, the reduced monitoring approach is used in these situations – with the distance to the first of two consecutive stations with an IQI of 0.64 or above being used.

3.2.17 For ease of visualization & storage reasons, the resampled model fits are distilled down to a single heatmap summarising these model fits (see Figure 5), along with a distribution of estimated distances to good.



(a) **(b)**

Figure 5: Different ways of displaying results of probabilistic regression. In (a), each curve is displayed in a spaghetti plot. In (b), these curves are distilled down to a heatmap, where each hexagon is coloured by the number of curves which pass through it.

3.2.18 It is important to note that the number of data points used to construct these non-linear models is close to the lower limit of what is reasonable. This is a consequence of the challenges of acquiring and analysing IQI samples. The placement of the IQI stations must be carefully considered when the EMP is produced, as is stressed in the relevant guidance[2]. In an ideal situation, and assuming a transect consists of 7 IQI stations, 2-3 of these would fall outside the mixing zone with the remainder distributed closer to the pen edge.

3.2.19 Situations where the form of the IQI evolution with distance is not well constrained by the monitoring data may lead to issues with model fitting. Such scenarios, and the approach SEPA will use to handle them, are described in Figure 6.

3.2.20 Where the reduced monitoring rule, rather than a resampled model, is used to generate a distance-to-good distribution this is expected to be conservative/precautionary. It is used in cases where: 1) insufficient IQI stations have been analysed to fit a model to; or 2) sufficient IQI stations have been analysed, but the resulting regression model is of insufficient quality to use.

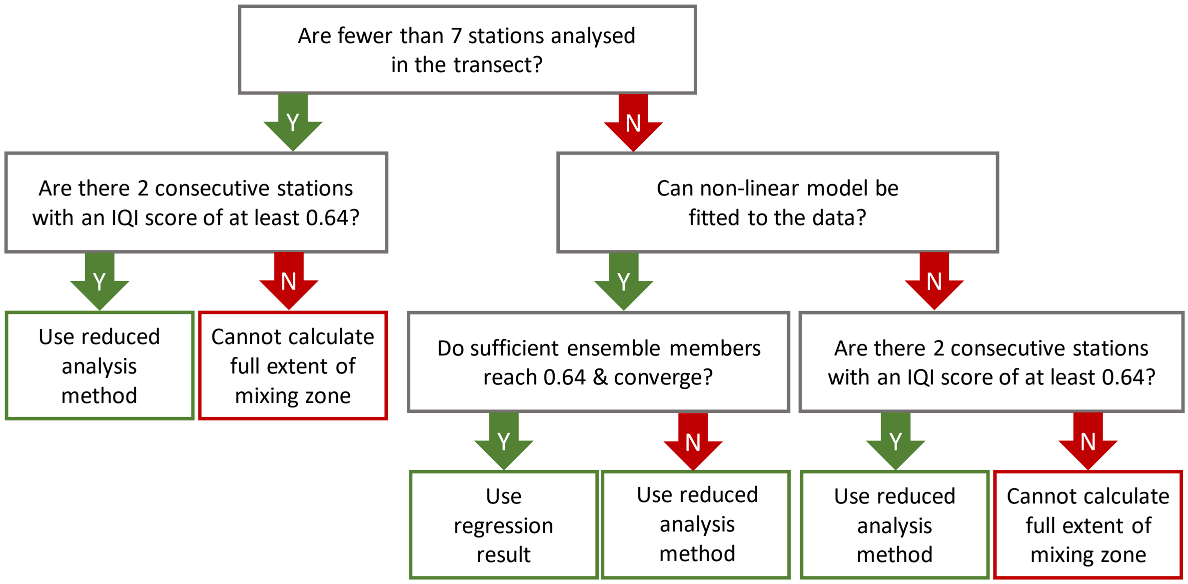


Figure 6: Approaches to situations where data is unsuitable for use with the probabilistic regression model

3.3 Estimating breach coordinates

3.3.1 This step takes the likelihood distributions, from the previous step, and transforms them into equivalent spatial coordinates.

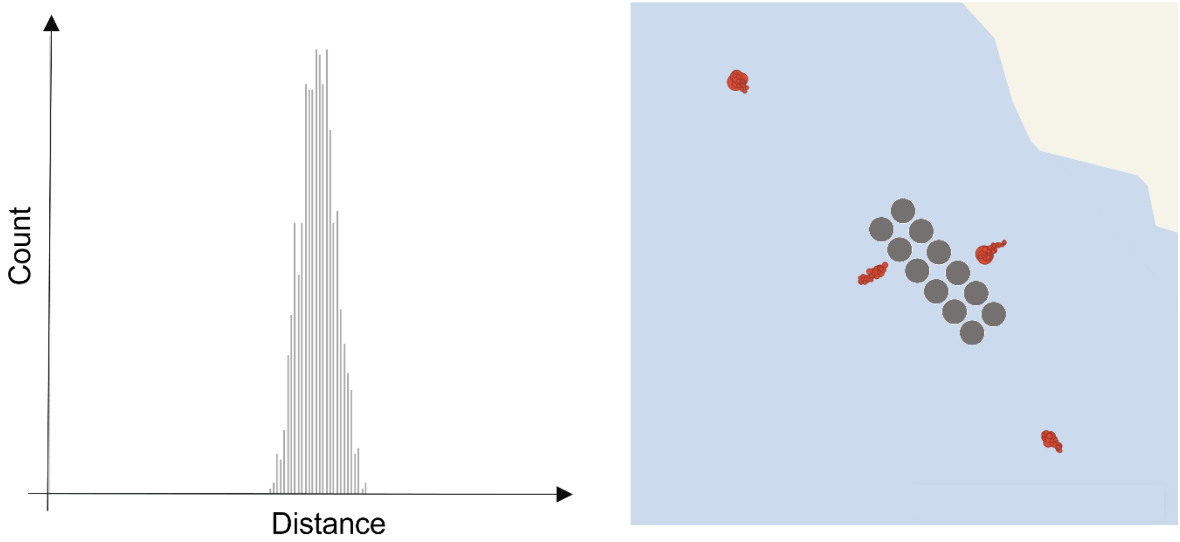
3.3.2 A single set of breach coordinates is attained by projecting from the pen edge station by one of the distances to good along the best-fit bearing for that transect.

3.3.3 This is repeated for every distance to good which forms the likelihood distribution, constructed in the previous script.

3.3.4 This process is repeated for every transect in the survey.

3.3.5 The result is a dataset of spatial coordinates, characterising the region in which an IQI of 0.64 (good) is expected to be achieved.

3.3.6 These positions are the spatial equivalents of the distance likelihood distributions calculated in the previous step (Figure 7).



(a) **(b)**

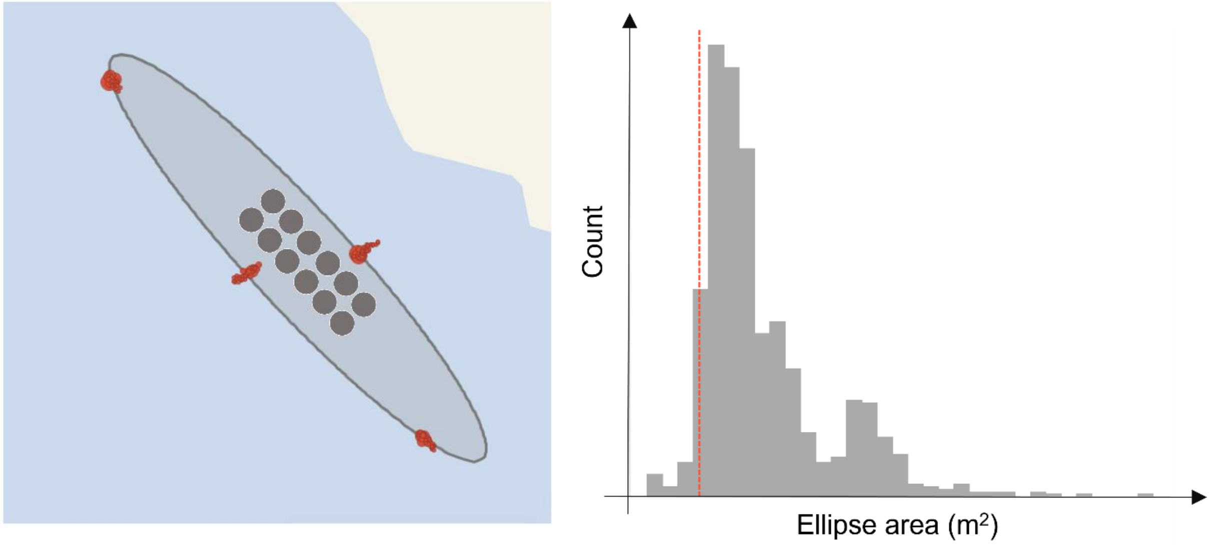
Figure 7: Conversion of (a) distances to good into (b) corresponding spatial coordinates. Breach positions are indicated by red dots, sized by frequency.

3.3.7 In addition to the above, the spatial coordinates corresponding to the best-fit breach distance (distance at which IQI=0.64) are also diagnosed.

3.4 Area approach

3.4.1 These spatial coordinates are now used to quantify the observed footprint area. This allows compliance to be assessed against the permitted area.

3.4.2 This is accomplished by taking these distributions of breach coordinates for each transect, randomly drawing one result from each transect, then drawing the minimum spanning ellipse which will enclose these locations. The area of the resulting ellipse is then diagnosed.



(a) (b)

Figure 8: Indicative ellipse and histogram of footprint areas. In (a), an indicative ellipse is shown, based on the best-fit regression results. In (b), the footprint area distribution is displayed. The red line indicates the 5th percentile, which is used for compliance assessment.

3.4.3 This is then repeated many times, until a distribution of areas has been created for the farm's footprint (Figure 8).

3.4.4 Summary statistics can be used to summarise any distribution. For this work, the $P_{5\%}$ footprint area is used for compliance assessment. This statistic is consistent with SEPA being 95% confident that a site has failed, before classing it as failing.

3.4.5 The $P_{5\%}$ footprint area is compared against the permitted area. If this observed $P_{5\%}$ footprint area is larger than what is permitted, SEPA can be sufficiently confident that the site has failed to meet this standard.

3.4.7 The model also displays a single elliptical polygon, based on the best-fit breach distance/coordinates for each transect. This ellipse polygon is provided only for indicative purposes. It is important to note that since a probabilistic approach is used for compliance assessment, no single polygon uniquely describes the modelled footprint.

4 Extending model to multiple, overlapping elliptical inputs

4.1 The above approaches handle single pen group surveys, in addition to multiple pen group surveys where the footprint area is modelled as a single ellipse.

4.2 The EMP guidance does allow the calculated footprint of a multi-pen group farm to be formed of multiple ellipses (one per pen group), where sufficient evidence from the inter-pen group region is provided to constrain multiple ellipses.

4.3 The principles outlined in this document can be adapted to calculate such a footprint. Adequate data to support the bounding of the ellipses in the inter-pen group region will be required.

4.4 This evidence could consist of either:

- One full transect per pen group, which allows for the probabilistic regression model to be used (with a distribution of distances returned); or
- A reduced-analysis approach where two consecutive stations with an IQI of at least 0.64 are returned, and the distance to the first station is used. Further details of these approaches are available in the EMP guidance[2].

4.5 To calculate a footprint area, the approach described in Section 3.4 can be adapted for use with multiple ellipses, such that:

- One ellipse per pen group is constructed for each random draw of breach positions
- These ellipses are dissolved to form a single polygon
- The area of the resulting polygon is diagnosed
- The process is repeated until a full likelihood distribution of footprint areas has been constructed

- $P_{5\%}$ for this area distribution is then used for compliance assessment purposes, as before.

Appendix – Formulation of non-linear models used

Model	Parameterisation	No. of parameters
Shifted Michaelis-Menten	$f(x) = c + \frac{d-c}{1+(^{e}/_{x})}$	3
Four-parameter logistic	$f(x) = c + \frac{d-c}{1+\exp(-b(x-\mathrm{e}))}$	4
Five-parameter logistic	$f(x) = c + \frac{d-c}{(1+\exp(-b(x-\mathrm{e})))^{f}}$	5

Table 1: Formulation of models used. Note that x in all the parameterisations represents the distance from the pen edge.

For information on accessing this document in an alternative format or language please either contact SEPA by telephone on 03000 99 66 99 or by email to equalities@sepa.org.uk

If you are a user of British Sign Language (BSL) the Contact Scotland BSL service gives you access to an online interpreter, enabling you to communicate with us using sign language.

http://contactscotland-bsl.org/

www.sepa.org.uk

03000 99 66 99

Strathallan House, Castle Business Park, Stirling, FK9 4TZ