As mentioned in Appendix S1, maximization is performed on the log-likelihood. In this section we give the derivations of the derivatives of the log-likelihood, which were used to aid optimization. In the main paper we used the symbol $y$ to denote the distance from transect to object. In practice, this distance is a perpendicular distance in the case of line transects and a radial distance for point transects. Here, to avoid confusion, perpendicular distances in the line transect case are denoted by $x$ and radial distances in the point transect case are denoted by $r$.

**Line transects**

Starting from the log-likelihood:

$$l(\theta, \phi; x, Z) = \sum_{i=1}^{n} \left( \log \sum_{j=1}^{J} \phi_j g_j(x_i, Z; \theta_j) - \log \sum_{j=1}^{J} \phi_j \mu_{ij} \right)$$

we derive the derivatives with respect to the optimisation parameters.

**With respect to $\beta_{0j}$**

For the intercept terms (also considering in the non-covariate case, these are just the parameters), the parameters have no effect outside of their mixture (i.e. $\beta_{0j}$ only has an influence on mixture component $j^*$), so we can write:

$$\frac{\partial l(\theta, \phi; x, Z)}{\partial \beta_{0j}} = \sum_{i=1}^{n} \frac{1}{g(x_i, Z; \theta, \phi)} \frac{\partial}{\partial \beta_{0j}} g_j(x_i, Z; \theta_j) - \frac{\partial}{\partial \beta_{0j}} \mu_{ij}.$$  

Now, to first find $\frac{\partial}{\partial \beta_{0j}} g_j(x_i, Z; \theta_j)$:

$$\frac{\partial g_j(x_i, Z; \theta_j)}{\partial \beta_{0j}} = \frac{\partial}{\partial \beta_{0j}} \exp \left( -\frac{x_i^2}{2\sigma_{j}^2} \right),$$

applying the chain rule and remembering that $\sigma_{j}$ is a (trivial) function of the $\beta_{0j}$s:

$$\frac{\partial g_j(x_i, Z; \theta_j)}{\partial \beta_{0j}} = \left( \frac{x_i}{\sigma_{j}} \right)^2 \exp \left( -\frac{x_i^2}{2\sigma_{j}^2} \right).$$

Expressing $\mu_{ij}$ in terms of the error function, Erf:

$$\frac{\partial \mu_{ij}}{\partial \beta_{0j}} = \frac{\partial}{\partial \beta_{0j}} \left( \sqrt{\frac{\pi}{2}} \sigma_{j} \text{Erf} \left( \frac{w}{\sqrt{2\sigma_{j}^2}} \right) \right)$$

$$= \text{Erf} \left( \frac{w}{\sqrt{2\sigma_{j}^2}} \right) \frac{\partial}{\partial \beta_{0j}} \left( \sqrt{\frac{\pi}{2}} \sigma_{j} \right) + \sqrt{\frac{\pi}{2}} \sigma_{j} \frac{\partial}{\partial \beta_{0j}} \left( \text{Erf} \left( \frac{w}{\sqrt{2\sigma_{j}^2}} \right) \right)$$

(2)
To find \( \frac{\partial}{\partial \beta_{0j}} \text{Erf} \left( \frac{w}{\sqrt{2}\sigma_{j^*}^2} \right) \), note that we can write and then apply the chain rule:

\[
\frac{\partial}{\partial \beta_{0j}} \text{Erf} \left( \frac{w}{\sqrt{2}\sigma_{j^*}^2} \right) = \frac{\partial}{\partial \beta_{0j}} S(u(\sigma_{j^*})) = \frac{\partial S(u)}{\partial \sigma_{j^*}} \frac{\partial \sigma_{j^*}}{\partial \beta_{0j}}
\]

where

\[ S(u) = \int_0^u \exp(-t^2)dt \quad \text{and} \quad u(\sigma_{j^*}) = \frac{w}{\sqrt{2}\sigma_{j^*}^2}. \]

Their derivatives being

\[
\frac{\partial S(u)}{\partial u} = \frac{2}{\sqrt{\pi}} \exp\left(-u^2\right), \quad \frac{\partial u(\sigma_{j^*})}{\partial \sigma_{j^*}} = -\frac{w}{\sqrt{2}\sigma_{j^*}^2}.
\]

Given these terms, it is just a case of multiplying them:

\[
\frac{\partial S(u)}{\partial u} \frac{\partial u(\sigma_{j^*})}{\partial \sigma_{j^*}} \frac{\partial \sigma_{j^*}}{\partial \beta_{0j}} = -\frac{\sqrt{2}}{\pi} \frac{w}{\sigma_{j^*}^2} \exp\left(-\frac{w^2}{2\sigma_{j^*}^2}\right)
\]

Substituting into (2):

\[
\frac{\partial \mu_{ij^*}}{\partial \beta_{0j}} = \mu_{ij^*} - \frac{w}{\sigma_{j^*}^2} \exp\left(-\frac{w^2}{2\sigma_{j^*}^2}\right)
\]

Finally, the derivative is:

\[
\frac{\partial l(\theta, \phi: x, Z)}{\partial \beta_{0j}} = \sum_{i=1}^n \left( \frac{x_i}{\sigma_{j^*}} \right)^2 \phi_{j^*} g_{j^*}(x_i, Z; \theta_{j^*}) \frac{\phi_{j^*}}{\mu_i} (\mu_{ij^*} - wg_{j^*}(w, Z; \theta_{j^*})).
\]

**With respect to** \( \beta_{k*} \)

Derivatives with respect to the common covariate parameters are found in a similar way to above. The expressions are slightly more complicated since the \( \beta_{k*} \)s effect all of the mixture components.

\[
\frac{\partial l(\theta, \phi: x, Z)}{\partial \beta_{k}} = \sum_{i=1}^n \left( \frac{1}{g(x_i, Z; \theta, \phi)} \sum_{j=1}^J \phi_j \frac{\partial}{\partial \beta_{k}} g_j(x_i, Z; \theta) - \frac{1}{\mu_i} \sum_{j=1}^J \phi_j \frac{\partial}{\partial \beta_{k}} \mu_{ij} \right)
\]

Every \( \sigma_j \) is a function of the \( \beta_{k*} \)s, so:

\[
\frac{\partial \sigma_j}{\partial \beta_{k*}} = \frac{\partial}{\partial \beta_{k*}} \exp\left(\beta_{0j} + \sum_{k=1}^K z_{ik} \beta_k\right),
\]

\[
= z_{ik} \sigma_j.
\]

Hence:

\[
\frac{\partial}{\partial \beta_{k*}} \exp\left(-\frac{x_i^2}{2\sigma_j^2}\right) = z_{ik} \left(\frac{x_i}{\sigma_j}\right)^2 \exp\left(-\frac{x_i^2}{2\sigma_j^2}\right) = z_{ik} \left(\frac{x_i}{\sigma_j}\right)^2 g_j(x_i, Z; \theta).
\]
And so for the $\mu_{ij}$s:

$$\frac{\partial \mu_{ij}}{\partial \beta_{k*}} = z_{ik*} \left( \mu_{ij} - w \exp \left( -\frac{w^2}{2\sigma_j^2} \right) \right)$$

The derivative is then:

$$\frac{\partial l(\theta, \phi; x, Z)}{\partial \beta_{k*}} = \sum_{i=1}^{n} \left( \frac{1}{g(x_i, Z; \theta, \phi)} \sum_{j=1}^{J} \phi_j z_{ik*} \left( \frac{x_i}{\sigma_j} \right)^2 g_j(x_i, Z; \theta_j) - \frac{1}{\mu_i} \sum_{j=1}^{J} \phi_j z_{ik*} (\mu_{ij} - w g_j(x_i, Z; \theta_j)) \right)$$

**With respect to $\alpha_j$**

First note that we can write the likelihood (1) as:

$$l(\theta, \phi; x, Z) = \sum_{i=1}^{n} \left( \log \left( \sum_{j=1}^{J-1} \phi_j g_j(x_i, Z; \theta_j) \right) + \left( 1 - \sum_{j=1}^{J-1} \phi_j g_j(x_i, Z; \theta_j) \right) \right) - \log \left( \sum_{j=1}^{J-1} \phi_j \mu_{ij} + \left( 1 - \sum_{j=1}^{J-1} \phi_j \mu_{ij} \right) \right)$$

The derivatives with respect to the $\alpha_j$'s of this expression are then:

$$\frac{\partial l(\theta, \phi; x, Z)}{\partial \alpha_{j*}} = \left( \sum_{i=1}^{n} \frac{1}{g(x_i, Z; \theta, \phi)} \left( \sum_{j=1}^{J-1} \phi_j g_j(x_i, Z; \theta_j) \frac{\partial \phi_j}{\partial \alpha_{j*}} - g_j(x_i, Z; \theta_j) \sum_{j=1}^{J-1} \frac{\partial \phi_j}{\partial \alpha_{j*}} \right) \right) - \frac{1}{\mu_i} \left( \sum_{j=1}^{J-1} \mu_{ij} \frac{\partial \phi_j}{\partial \alpha_{j*}} - \mu_{ij} \sum_{j=1}^{J-1} \frac{\partial \phi_j}{\partial \alpha_{j*}} \right)$$

Finding the derivatives is then simply a matter of finding the derivatives of $\phi_j$ with respect to $\alpha_{j*}$ and substituting them back into (4).

$$\frac{\partial \phi_j}{\partial \alpha_{j*}} = \frac{\partial}{\partial \alpha_{j*}} F(\sum_{p=1}^{j} e^{\alpha_p}) - \frac{\partial}{\partial \alpha_{j*}} F(\sum_{p=1}^{j-1} e^{\alpha_p}).$$

Looking at each of the terms:

$$\frac{\partial}{\partial \alpha_{j*}} F(\sum_{p=1}^{j} e^{\alpha_p}) = A_j = \begin{cases} e^{\alpha_j} f(\sum_{p=1}^{j} e^{\alpha_p}) & \text{for } j \geq j*, \\ 0 & \text{for } j < j*. \end{cases}$$

and

$$\frac{\partial}{\partial \alpha_{j*}} F(\sum_{p=1}^{j-1} e^{\alpha_p}) = A_{j-1} = \begin{cases} e^{\alpha_j} f(\sum_{p=1}^{j-1} e^{\alpha_p}) & \text{for } j - 1 \geq j*, \\ 0 & \text{for } j - 1 < j*. \end{cases}$$

So

$$\frac{\partial \phi_j}{\partial \alpha_{j*}} = A_j - A_{j-1}.$$
Point transects

We now provide the corresponding quantities for point transects, starting from the log-likelihood:

$$l(\theta, \phi; r, Z) = n \log 2\pi + \sum_{i=1}^{n} \left( \log r_i + \log \sum_{j=1}^{J} \phi_j g_j(r_i, Z; \theta_j) - \log \sum_{j=1}^{J} \phi_j \nu_{ij} \right).$$  \hspace{1cm} (5)

With respect to $\beta_{0j}$

From (5), one can see that we obtain:

$$\frac{\partial l(\theta, \phi; r, Z)}{\partial \beta_{0j}} = \sum_{i=1}^{n} \left( \frac{\partial}{\partial \beta_{0j}} \log \sum_{j=1}^{J} \phi_j g_j(r_i, Z; \theta_j) - \frac{\partial}{\partial \beta_{0j}} \log \sum_{j=1}^{J} \phi_j \nu_{ij} \right)$$

$$= \sum_{i=1}^{n} \left( \frac{\partial_{\beta_{0j}} g_j(r_i, Z; \theta_j)}{g(r_i, Z; \theta, \phi)} - \frac{\partial_{\beta_{0j}} \nu_{ij}}{\sum_{j=1}^{J} \phi_j \nu_{ij}} \right)$$

the first part of which (the derivatives of the detection function) are as in the line transect case. The derivatives of $\nu_{ij}$ are simpler in the point transect case, since there is an easy analytic expression for $\nu_{ij}$ when $g_j$ is half-normal:

$$\nu_{ij} = 2\pi \sigma_j^2 (1 - \exp(-w^2/2\sigma_j^2))$$

then simply applying the product rule yields:

$$\frac{\partial \nu_{ij}}{\partial \beta_{0j}} = 2(\nu_{ij} + \pi w g_j(w)).$$

Substituting this into the above expression:

$$\frac{\partial l(\theta, \phi; r, Z)}{\partial \beta_{0j}} = \sum_{i=1}^{n} \left( \frac{\partial_{\beta_{0j}} (r_i/\sigma_j)^2 g_j(r_i, Z; \theta_j) - \partial_{\beta_{0j}} \phi_j \nu_{ij}}{g(r_i, Z; \theta, \phi)} \right)$$

With respect to $\beta_{kj}$

Again working from (5), we obtain:

$$\frac{\partial l(\theta, \phi; r, Z)}{\partial \beta_{kj}} = \sum_{i=1}^{n} \left( \frac{\partial}{\partial \beta_{kj}} \log \sum_{j=1}^{J} \phi_j g_j(r_i, Z; \theta_j) - \frac{\partial}{\partial \beta_{kj}} \log \sum_{j=1}^{J} \phi_j \nu_{ij} \right)$$

$$= \sum_{i=1}^{n} \left( \frac{\sum_{j=1}^{J} \phi_j \frac{\partial}{\partial \beta_{kj}} g_j(r_i, Z; \theta_j)}{g(r_i, Z; \theta, \phi)} - \frac{\sum_{j=1}^{J} \phi_j \frac{\partial}{\partial \beta_{kj}} \nu_{ij}}{\sum_{j=1}^{J} \phi_j \nu_{ij}} \right)$$

The derivatives of $g_j$ are as in (3). For $\nu_{ij}$:

$$\frac{\partial \nu_{ij}}{\partial \beta_{kj}} = 2z_{ik} (\nu_{ij} - \pi w g_j(w))$$

Putting that together:

$$\frac{\partial l(\theta, \phi; r, Z)}{\partial \beta_{kj}} = \sum_{i=1}^{n} \left( \frac{\sum_{j=1}^{J} \phi_j \frac{z_{ij}^2}{\sigma_j^2} g_j(r_i, Z; \theta_j)}{g(r_i, Z; \theta, \phi)} - \frac{\sum_{j=1}^{J} \phi_j \nu_{ij} (\nu_{ij} - \pi w g_j(w))}{\sum_{j=1}^{J} \phi_j \nu_{ij}} \right).$$