

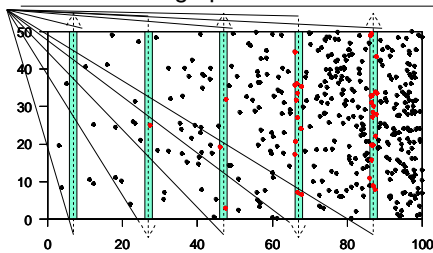
Introduction à la méthode de *Distance Sampling* en transect de ligne



Echantillonnage de l'espace

- Transect de bande (*Strip transect*)
 - Echantillonnage d'une bande d'une largeur définie le long de transects ;
- Transect de ligne (*Line transect*)
 - Utilisation de données supplémentaires pour estimer la largeur de la zone échantillonnée
- Transect de point (*Point transect*)
 - Echantillonnage depuis une série de points avec un rayon fixe ou estimé.

Echantillonnage par transect de bande (*strip transect*)



$$\hat{D} = \frac{n}{a} = \frac{36}{500} = 0,72$$

$$\hat{N} = \hat{D}A$$

$$= 0,72 \times 5000 = 360$$

D = densité
 N = abondance

Taille de la zone d'étude = 5000 (A)

Longueur de transect totale = $50 \times 5 = 250$ (L)

Demi-largeur de bande = 1 (w)

Zone échantillonnée = $2wL = 2 \times 1 \times 250 = 500$ (a)

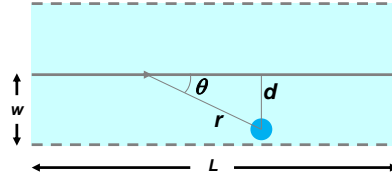
Nombre d'animaux recensés = 36 (n)

$$\hat{N} = \frac{n}{a/A} = \frac{36}{0,1} = 360$$

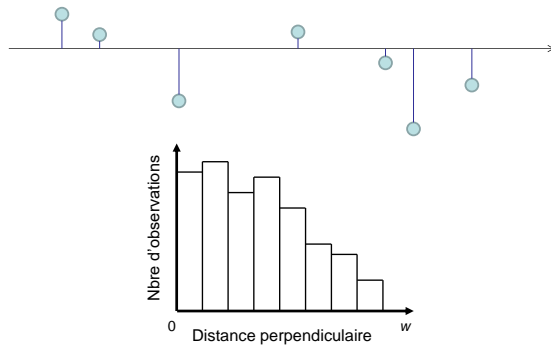
Echantillonnage par transect de ligne (*line transect*)

- Prospection le long de transects en collectant des données sur :
 - La distance parcourue (+ conditions de prospection)
 - Les espèces et la taille des groupes (+ observateur, signal, etc)
- On ne peut pas supposer que la probabilité de détection est égale à 1, donc...
 - Distance perpendiculaire (d)
 - Distance (r) et angle (θ) des groupes détectés $d = r \sin(\theta)$

L = longueur de transect
 w = demi-largeur de bande
 Zone prospectée nominale = $2wL$



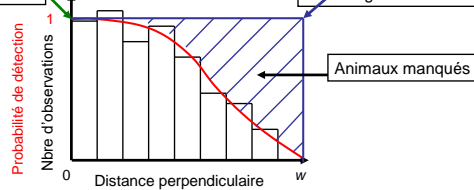
Echantillonnage par transect de ligne (*line transect*)



Estimer la probabilité de détection (*detection probability*)

Suppose que la probabilité de détection = 1 à distance 0

Dans le transect de bande, la probabilité de détection = 1 sur toute la largeur de la bande



Probabilité de détection moyenne : $\hat{p}_a = \frac{\text{aire sous la courbe}}{\text{aire du rectangle}}$

Hypothèse importante : les animaux sont distribués uniformément entre 0 et w

Estimation d'abondance/de densité en transect de ligne

- Estimation de la densité en transect de bande (D) :

$$\hat{D} = \frac{n}{2wL}$$

n = nombre d'animaux observés
 L = longueur du transect
 w = demi-largeur de la bande

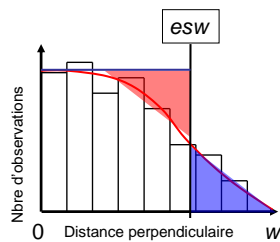
- Estimation de la densité en transect de ligne (D) :

$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$

p_a = probabilité de détection moyenne
 $w \times p_a$ = Demi-largeur effective de bande
effective strip half-width (esw)

- Pour l'abondance (N) : $\hat{N} = \hat{D}A$
 A = taille de la zone d'étude

Demi-largeur effective de bande *Effective strip half-width (esw)*

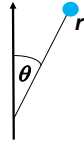


Hypothèses du transect de ligne

- La zone d'étude est prospectée de façon représentative
 - Tous les animaux sur la ligne de transect sont détectés
 - Les animaux ne bougent pas avant la détection
 - Les animaux proches de la ligne de transect sont distribués uniformément autour de la ligne
 - Toutes les mesures sont exactes
 - Les observations sont indépendantes
- Le non-respect de ces hypothèses peut causer des biais...

Toutes les mesures sont exactes

- Longueur du transect
- Identification de l'espèce
 - Mauvaise identification
 - Observation d'espèce non-identifiée
- Distance perpendiculaire
 - Distances radiales
 - Sous ou sur-estimation systématique
 - Erreurs aléatoires
 - Angles
 - Arrondis
 - Erreurs aléatoires
- Taille de groupe
 - Sous ou sur-estimation systématique
 - Évolution du biais avec la distance



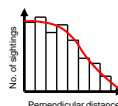
Tous les animaux sur la ligne de transect sont détectés

- A distance perpendiculaire 0, probabilité de détection = 1
 - Par convention : $g(0)=1$
- Si des animaux (ou des groupes) sur la ligne de transect sont manqués :
 - Densité et abondance biaisées négativement
- Deux sources de biais
 - Disponibilité
 - Tous les animaux sur la ligne de transect ne sont pas à la surface
 - Perception
 - Les animaux sur la ligne sont à la surface mais ne sont pas détectés par les observateurs

Les animaux ne se déplacent pas avant la détection

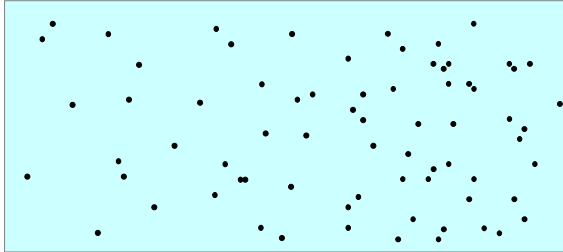
- Et pourtant ils bougent...
 - Si le mouvement est après la détection, pas de problème
 - Le mouvement avant la détection peut être
 - Aléatoire
 - En réponse à la plateforme d'observation
- Le mouvement aléatoire n'est souvent pas un problème
 - A moins que la vitesse de déplacement soit bien plus grande que celle de la plateforme
 - Peut causer un biais positif dans l'abondance
- Le mouvement en réponse est un problème
 - L'évitement entraîne un biais négatif
 - L'attraction entraîne un biais positif
 - Peut être important

$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$



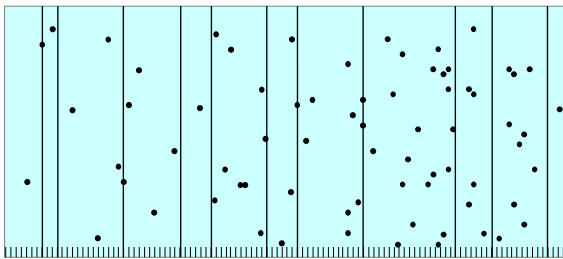
La zone d'étude est prospectée de façon représentative

- Les animaux ne sont pas distribués au hasard dans l'espace
- L'échantillonnage doit être aléatoire et systématique
 - Pour obtenir une probabilité de couverture égale



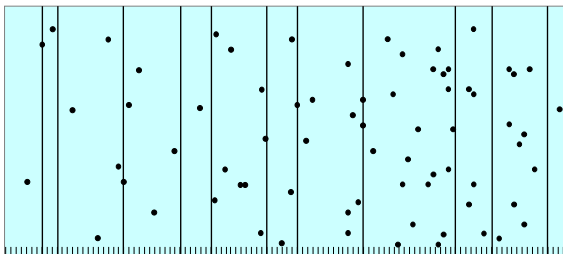
Conception d'une prospection

- Lignes parallèles espacées de façon aléatoire
 - Donne une probabilité de couverture égale



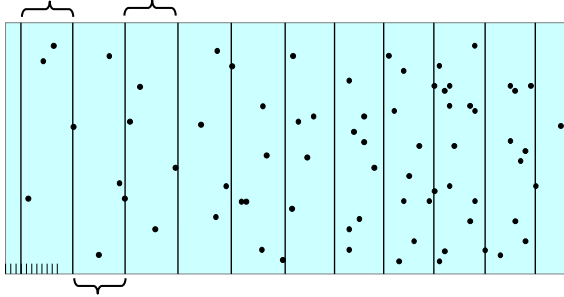
Conception d'une prospection

- Conception aléatoires
 - Peut être plus difficile à mettre en place au niveau logistique
 - Peut entraîner une variance plus élevée



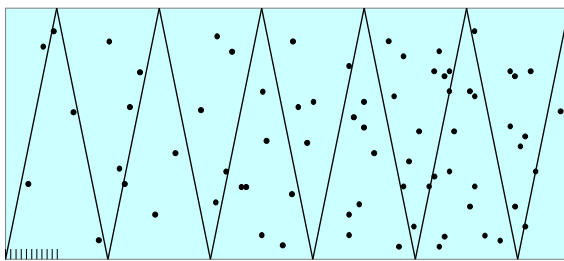
Conception d'une prospection

- Lignes parallèles espacées de manière égale
 - Avec un point de départ aléatoire
 - Souvent utilisé en pratique
 - Mais perte de temps en transit...



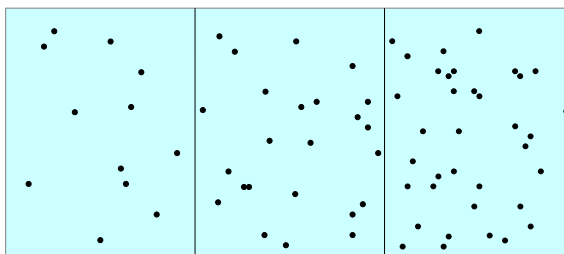
Conception d'une prospection

- Zig-zags systématiques
 - Avec un point de départ aléatoire
 - Approprié pour de grandes zones
 - Problèmes potentiels à la fin des lignes et dans de petites zones



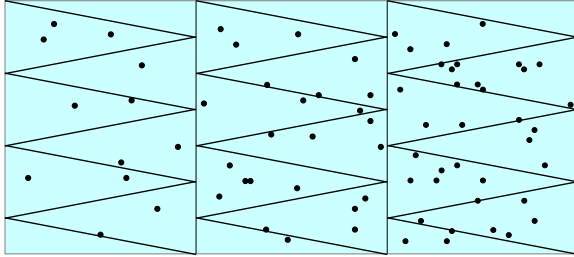
Stratification de la zone d'étude

- Deux raisons à la stratification
 - Raisons logistiques
 - Pour réduire la variance globale
 - Si la densité est connue ou supposée varier



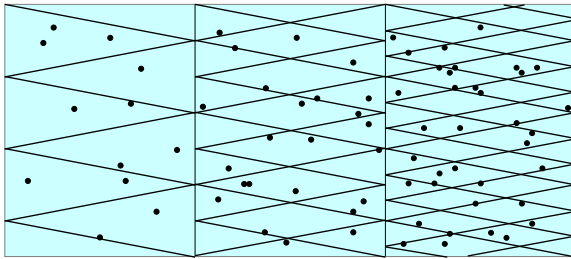
Conception de prospection stratifiée

- Conception stratifiée (avec zig-zags)
 - Lignes de prospection placées perpendiculairement aux contours de densités
 - Caractéristique souhaitable
 - Variabilité inter-transect plus faible dans chaque zone que sur l'ensemble de la zone

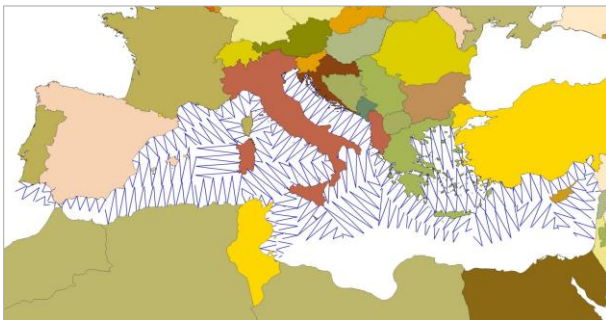


Conception de prospection stratifiée

- Conception stratifiée avec effort inégal
 - Augmentation de l'effort avec la densité attendue
 - Réduit encore plus la variance
 - Mais les couches doivent être analysées séparément



ACCOBAMS ASI transects prévus



L'échantillonnage par transect de ligne en pratique

- Définir la zone dans laquelle l'abondance doit être estimée
- Concevoir la prospection et choisir des lignes de transect
- Prospecter le long des lignes
 - Chercher devant et de chaque côté du transect
- Quand un animal (groupe d'animaux) est détecté:
 - Identifier l'espèce
 - Déterminer la taille du groupe
 - Mesurer la distance perpendiculaire de l'animal/le groupe à la ligne de transect
- Continuer à chercher le long des lignes de transect
- A la fin :
 - Le nombre d'animaux observés est connu (n)
 - Ou le nombre de groupes et la taille de groupe moyenne
 - La longueur totale de transect est connue (L)
 - Utiliser les données de distances pour estimer la probabilité moyenne de détection (p_d)

Conception d'une prospection (Survey design) dans DISTANCE

Le moteur de conception de prospection dans DISTANCE permet d'entrer des informations sur votre zone d'étude, y compris les strates, pour ensuite paramétrer la prospection, calculer la probabilité de couverture de votre zone d'étude et générer les lignes de transect.

Tout d'abord, vous devez créer un projet DISTANCE qui contient les données géographiques de votre zone d'étude et votre conception de prospection.

1. Créer un projet DISTANCE

Ouvrir DISTANCE. Sélectionner **File (Fichier) -> New Project (Nouveau projet)**. Dans la fenêtre **Create Project (Créer un Projet)** qui s'ouvre, sélectionner le répertoire de fichiers où sauvegarder le projet, le nommer et cliquer sur **Create (Créer)**.

Le **New Project Set Up Wizard (Assistant de paramétrage du nouveau projet)** démarre. Dans **I want to (Je souhaite)**, choisir *Design a new survey (Concevoir une nouvelle prospection)*. Cliquer sur **Next (Suivant)**, puis sur **Finish (Terminer)**.

L'onglet **Data (Données)** du **Project Browser (Navigateur du Projet)** s'ouvre pour le projet, montrant une fenêtre **Data Layers (Couches de Données)** sur la gauche contenant une couche de données intitulée *Study area (zone d'étude)* et une fenêtre **Contents of Global layer « Study area » (Contenu de la couche générale « Zone d'étude »)** sur la droite avec des informations sur votre zone d'étude.

2. Système de coordonnées géographiques, projections et unités

Avant toute chose, il faut s'assurer que le système de coordonnées géographiques, la projection et les unités dans DISTANCE sont paramétrés correctement.

Sélectionner **Tools (Outils) -> Preferences... (Préférences)** dans la barre d'outils et sélectionner l'onglet **Geographic (Géographique)** dans la fenêtre **Distance Preferences (Préférences Distance)**. Dans *Geographic coordinate system*, sélectionner **WGS 1984**. Dans *Projection of shapefile data (Projection des données de fichiers de forme)* et *Projection of map (Projection de carte)*, sélectionner **[None] (Aucun)**. Dans *Shapefile units (Unités des fichiers de formes)* et *Map units (Unités des cartes)*, sélectionner **Degree (Degrés)**. Cliquer sur **OK** et fermer la fenêtre.

Sélectionner **File -> Project Properties (Propriétés du projet)** depuis la barre d'outils et sélectionner l'onglet **Geographic**. Pour *Geographic coordinate system*, *Projection of shapefile data* et *Projection of map*, choisir **[None]**. Pour *Shapefile units* et *Map units*, choisir **Meter (Mètres)** car ce sont les unités dont on va avoir besoin pour l'exercice ci-après. Cliquer sur **OK**.

3. Importer des données géographiques de la zone d'étude

Les données géographiques peuvent être importées dans DISTANCE de plusieurs manières : entrer les coordonnées à la main ; copier-coller les coordonnées depuis un fichier texte ; ou importer un fichier SIG dans le répertoire de fichiers du projet. Dans cet exemple, nous allons copier les coordonnées de la zone d'étude depuis un fichier texte.

Ouvrir le fichier **StAndrewsBay.txt**, qui contient les coordonnées de la baie de St Andrews en Ecosse. Les données sont en mètres. **Copier** toutes les lignes.

Dans DISTANCE, double-cliquer sur *Polygon (Polygone)* dans la colonne **Study area** dans la fenêtre de droite pour ouvrir la fenêtre **Shape Properties (Propriétés de formes)**. Sélectionner **Paste from Clipboard (Coller depuis le presse-papier)** dans *Whole shape (Forme entière)* sur le côté droit. Les colonnes **X** et **Y** contiennent maintenant les données du fichier texte. La case **# of vertices (Nombre de sommets)** en haut à droite devrait afficher 257. Cliquer sur **OK**.

4. Afficher les données sur une carte

Sélectionner l'onglet **Maps** dans le **Project Browser** et cliquer sur l'icône **New Map (Nouvelle carte)**, 4^{ème} depuis la gauche sous les onglets. On peut aussi cliquer sur **Maps -> New Map** dans la barre d'outils.

Double-cliquer sur le nom « New Map » et le renommer en « St Andrews Bay ».

Cliquer sur l'icône « **View** » (**Visualiser**), 6^{ème} depuis la gauche) ou cliquer sur **Map -> View Map** dans la barre d'outils pour ouvrir une fenêtre **Map** vide avec comme titre « Map 1: [St Andrews Bay] » en haut. Cliquer sur l'icône **Add Layer to Map (Ajouter une couche à la carte)**, 7^{ème} depuis la gauche, ou cliquer sur **Map -> Add Layer** dans la barre d'outils et sur **OK** pour confirmer l'ajout de la couche *Study area*.

Vous devriez voir une fenêtre montrant la zone de la baie de St Andrews.

Cependant...

Parfois, la carte est bien là, mais très petite. Dans ce cas, agrandir la carte en cliquant sur l'icône **Zoom In** (Zoom avant) de la barre d'outils (4^{ème} à partir de la droite) et en utilisant la souris pour dessiner un cadre plus grand autour de la carte. Utiliser l'icône **Pan** (Main), 2^{ème} à partir de la droite, pour déplacer la carte. Réduire la carte en cliquant sur **Zoom Out** (Zoom arrière), 3^{ème} depuis la droite puis en cliquant sur la carte. Cliquer sur l'icône **Full Extent** pour que la carte redevienne très petite.

Si la carte n'apparaît pas ou semble avoir un problème, vérifier les paramètres des fichiers de formes ou des projections (paragraphe 2.).

Fermer la fenêtre de carte en enregistrant les modifications et retourner au **Project Browser**.

5. Ajouter une couche de couverture

Avant que la moindre prospection ne soit créée, il faut ajouter une seconde couche de données. Cette couche est une grille de points pour estimer la probabilité de couverture (la probabilité moyenne qu'une ligne de transect couvrira chaque point). Cela peut être utilisé pour visualiser à quel point la prospection couvre la zone d'étude de manière égale.

Avant de créer une couche de couverture, vérifier que le système de coordonnées est exact.

Cliquer sur l'onglet **Data** dans le **Project Browser**. Sélectionner **Data -> Data layer Properties (Propriétés des couches de données)** depuis la barre d'outils et sélectionner l'onglet **Geographic data**. *Coordinate system* doit afficher **No coordinate system (Pas de système de coordonnées)**. Si ce n'est pas le cas, cliquer sur **Change coordinate system... (Changer le système de coordonnées)** et choisir **[None]** dans *Geographic coordinate system*. Cliquer sur **OK**.

Dans l'onglet **Data** cliquer sur l'icône **Create New Data Layer (Créer une nouvelle couche de données)**, 5^{ème} depuis la droite, ou sélectionner **Data -> Create Data Layer** depuis la barre d'outils pour ouvrir la fenêtre **Create New Layer**.

Dans **Layer Name (Nom de la couche)**, entrer « Grid » (Grille) et choisir Coverage (Couverture) dans *Layer Type (Type de couche)*.

Pour définir l'espacement entre les points, cliquer sur **Properties...** pour ouvrir la fenêtre **Grid Properties (Propriétés de la grille)**. De façon approximative, un espacement proche de la largeur de bande maximum totale est généralement approprié quand on estime la probabilité de couverture. Pour notre exemple, choisir **1** pour *Distance between grid points (Distance entre les points de la grille)* et **Kilometer** pour *Units of distance (Unités de distance)*.

Cliquer sur **OK** pour fermer les fenêtres **Grid Properties** et **Create Data Layer**. **DISTANCE** indique que l'espacement choisi va créer environ 988 points de grille. Cliquer sur **OK** pour confirmer. La couche Grid a été ajoutée dans *Data Layer Study area* sur la gauche et il y a une nouvelle colonne dans la fenêtre de droite avec les points de grille.

Maintenant, sélectionner l'onglet **Maps** et visualiser la carte de St Andrews Bay ;

Sélectionner de nouveau **Add Layer to Map**, confirmer Grid comme **Layer name** et cliquer sur **OK**. Les points de la grille de probabilité de couverture doivent apparaître sur la carte.

Note : pour des zones d'étude plus vastes, choisir un espacement de grille de la largeur de la bande peut entraîner la création d'un très grand nombre de points (plusieurs milliers). Dans ces cas, il est plus approprié de choisir un espacement plus grand.

Fermer la fenêtre de carte en enregistrant les modifications et retourner au **Project Browser**.

6. Créer une prospection

Par exemple, on peut créer une prospection pour une étude systématique avec des lignes parallèles d'un espacement de 5 km.

Dans l'onglet **Designs (Conceptions)** du **Project Browser**, sélectionner **Designs -> New Design (Nouvelle conception)** ou cliquer sur la première icône (triangle jaune) à droite de **Design:**, sous les onglets.

Dans la fenêtre de gauche, double-cliquer sur le nom « New Design » et le renommer « 5km spacing » (espacement 5 km) ou quelque chose de similaire.

Cliquer sur l'icône **Show details for selected design (Afficher les détails pour la conception sélectionnée)** avec un triangle et une loupe ou sélectionner **Design -> Design Details** depuis la barre d'outils pour ouvrir la fenêtre pour définir les détails de votre conception.

La fenêtre s'ouvre sur l'onglet **Inputs** (Contribution). Il y a aussi les onglets **Log** (Chargement) et **Results** (Résultats) sur la droite.

Dans **Type of Design (Type de conception)**, choisir Line pour *Sampler: (Echantillonneur)* et Systematic Random Sampling (*Echantillonnage aléatoire systématique*) pour *Class: (Classe)*.

Cliquer sur **Properties...** pour ouvrir la fenêtre **Design Properties**.

Ignorer l'onglet **General Properties** pour cet exemple.

Dans l'onglet **Effort Allocation (Allocation d'effort)**, laisser *Edge sampling (Echantillonnage du bord)* en **Minus (Moins)**.

Dans *Allocation by stratum (Allocation par strate)*, choisir Kilometer.

Dans le tableau en bas, entrer **5** dans **Spacing**. **DISTANCE** estime qu'en moyenne, 8 Samplers (Echantillonneurs), dans cet exemple des lignes, vont être créés et que la longueur moyenne en effort sera de 226,2 km.

Angle correspond à l'angle de l'axe de la conception avec lequel les lignes de transect vont être définies. S'il est égal à 0 l'axe est orienté nord-sud, donc les lignes sont orientées ouest-est.

Dans l'onglet **Sampler**, sélectionner Kilometer pour *Line sampler width units (Unités de largeur des lignes d'échantillonnage)* et **0,5** pour *Width (Largeur)*. Ceci est la demi-largeur de bande, donc la zone totale échantillonnée est définie ici comme une bande d'un kilomètre de large.

Dans l'onglet **Coverage Probability**, sélectionner **Estimate by simulation (Estimer par simulation)** avec **100 repetitions** ou simulations.

Cliquer sur **OK** pour enregistrer votre conception de prospection.

Fermer la fenêtre Design et retourner au **Project Browser**.

7. Générer une prospection

Maintenant que les propriétés de la conception de prospection ont été définies, la prospection peut être créée. Cette prospection est une réalisation aléatoire unique de la conception, c'est-à-dire qu'elle aura un point de départ aléatoire et un espacement de 5 km entre des lignes orientée est-ouest.

Cliquer sur **Run selected design (Lancer la conception sélectionnée)**, 4^{ème} icône depuis la droite, ou sélectionner **Design -> Run design** pour ouvrir la fenêtre **Run Design**. Sélectionner la seconde option, **Create new Survey** (Créer une nouvelle prospection) et cliquer sur **OK**.

Une nouvelle prospection est créée, la fenêtre **Survey Details** s'ouvre et l'onglet **Results** (Résultats) s'affiche.

Cet onglet a trois pages.

La première page, **Design engine output (Résultats du moteur de conception)**, résume les informations entrées et les résultats pour cette prospection, notamment le nombre de lignes générées, la longueur de ligne réalisée en effort (« realised on effort trackline length » ou « transect length ») et la longueur de ligne totale (« total trackline length ») soit la longueur parcourue en effort + la distance entre la fin de la prospection et le retour au point de départ. La Proportion de la couche échantillonnée (« Proportion of stratum sample ») sur la ligne du bas correspond à la zone couverte par les échantillonneurs (« realised sampler area coverage ») par rapport à l'aire totale de la couche (« stratum area ») sur les deux lignes au-dessus.

La page **Survey Map** montre la position des échantillonneurs (lignes de transect). Si la carte est très petite en haut de la fenêtre, suivre les instructions du point 4 plus haut pour l'agrandir.

La page **Sampler location list (Liste des positions des échantillonneurs)** donne les points de départ et d'arrivée de chaque ligne. Ces points peuvent être copiés et collés dans un fichier texte pour être chargés sur un GPS.

Fermer la fenêtre **Survey Details** et retourner au **Project Browser**.

Le résumé des résultats de votre prospection sont affichés dans l'onglet **Surveys**.

Dans l'onglet **Data**, les lignes ont été ajoutées au projet dans une nouvelle couche de données, qui possède un fichier de formes associé (ce qui peut être un format plus adapté pour exporter dans d'autres systèmes type GPS). Cliquer sur le nom de la prospection pour voir les données sur la longueur des lignes de transect dans la fenêtre de droite.

Dans l'onglet **Maps**, ajouter une nouvelle couche de données contenant la prospection à la carte.

8. Statistiques de conception

Du fait que les prospections sont générées de manière aléatoire d'après la conception, certaines propriétés comme la longueur de ligne totale et la proportion de temps en effort sont des propriétés aléatoires avec une moyenne et une distribution. Il est intéressant d'examiner la proportion moyenne de ligne de transect totale en effort car plus elle est élevée, plus la conception est efficace. La longueur de ligne totale maximum et minimum qui peuvent être générées est également une chose intéressante à examiner.

Ouvrir de nouveau la fenêtre **Design Details** en utilisant l'onglet **Design**, l'icône **Show details for selected design** (triangle avec loupe) ou via **Design -> Design Details** depuis la barre d'outils.

Cliquer sur **Run... (Lancer)** de nouveau mais cette fois en choisissant l'option **Calculate coverage probability statistics (Calculer les statistiques de probabilité de couverture)** et cliquer sur **OK**.

100 répétitions ont été sélectionnées dans **Design Properties**, donc **DISTANCE** effectue 100 simulations. La barre de progression en haut indique le pourcentage de réalisation. Quand elle atteint 100 %, l'onglet **Results** de **Design Details** devient vert pour accéder aux résultats.

Il y a deux pages de résultats.

La page **Design engine output** contient les résultats en format texte.

La carte **Coverage Probability map** affiche une image où les variations de probabilités de couverture de la zone sont représentées avec un code couleurs. Il se peut qu'il y ait une probabilité de couverture plus faible sur les bords de la zone d'étude. Au nord et au sud, cela est probablement dû à l'angle de la limite de la zone avec l'axe de la conception. A l'est et à l'ouest, cela peut être dû au fait que l'option Minus a été cochée pour la propriété **Edge sampling**. Choisir Plus dans **Design Properties** puis **Effort Allocation** devrait y remédier mais impliquerait des lignes de transect plus longues.

Mais, dans l'ensemble, la probabilité de couverture devrait être homogène, ce qui ne serait pas forcément le cas pour des conceptions en zig-zag ou avec des zones d'étude aux formes plus complexes.

Si la carte de probabilité de couverture fait apparaître des « bandes » de haute et basse probabilité, il est possible de relancer la conception en ajoutant plus de répétitions et / ou en utilisant une grille avec un espacement de points plus petit. Pour cela, vous devez soit retourner dans Design et modifier les propriétés dans l'onglet Inputs soit créer une nouvelle conception (**New Design**).

En retournant dans l'onglet **Data**, cliquer sur la couche Grid dans la fenêtre de gauche : des colonnes de données de probabilité de couverture (en proportion et en pourcentages) ont été ajoutées dans Grid dans la fenêtre de droite.

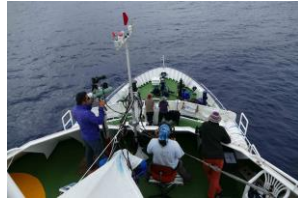
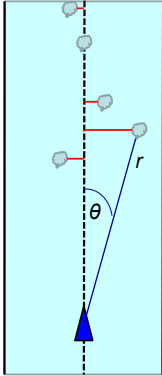
9. Pour aller plus loin

Il est possible d'explorer d'autres conceptions comme :

- Des lignes parallèles avec différent espacement entre les lignes de transect ou avec un axe différent ;
- Des lignes en zig-zag avec un espacement égal (Equal Spaced) et des angles égaux (Equal Angles).

Collecte de données par transect de ligne

Line transect data collection



Objectifs de l'étude

- Pourquoi estimer l'abondance ?
 - Premier aperçu approximatif du nombre d'individus
 - Contribution à un plan de gestion / conservation
 - Prospection unique ou élément d'une série
 - Abondance relative ou "absolue"
- Zone
 - Aussi grande que possible ou zone spécifique (ex : Aire Marine Protégée)
- Espèce(s) / population(s)
 - Une seule espèce cible ou prospection pour plusieurs espèces

Facteurs à considérer - 1

- Navire ou avion ?
 - Disponibilité
 - Coût
 - Efficacité
 - Logistique
 - Littoral / îles
 - Autres considérations
 - Observateurs
 - Formation
 - Renforcement des compétences
 - Données supplémentaires



Facteurs à considérer - 2

- Plateforme :
 - Vitesse de prospection
 - Altitude (avion)
 - Position des observateurs
 - Equipement
- Conditions :
 - De mer
 - Houle
 - Reflets
 - Visibilité générale
- Observateurs :
 - Experience
 - Fatigue
 - Acuité visuelle



Facteurs à considérer - 3

- Observations
 - Taille de l'animal
 - Taille du groupe
 - Animal seul, petit groupe, grand groupe ?
 - Composition / structure – qu'est-ce qu'un groupe ?
 - Vitesse du déplacement (aléatoire)
 - Par rapport à la vitesse de la plateforme
 - Mouvement en réponse à la plateforme ?
 - Evitement ou attraction ?
 - Elément de détection
 - Très visible, cryptique, les deux ?
 - Disponibilité sur la ligne de transect
 - Perception sur la ligne de transect

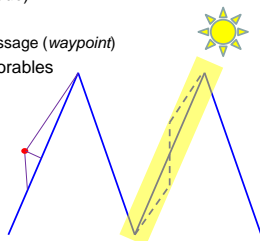


Etablir un protocole de collecte de données

- Effort de prospection
 - Données à collecter
 - Limites des conditions de prospection (conditions de mer, visibilité, etc)
 - Zones à prospecter
 - Rotation des observateurs
- Observations
 - Mode passif (*Passing mode*) ou mode rapprochement (*Closing mode*)
 - Données nécessaires minimum pour chaque observation
 - Méthode de mesures des angles et des distances
 - Données supplémentaires (ex : comportement, jeunes)
- Autres données
 - Espèces autres que cétacés (ex : oiseaux, activités humaines)
 - Données océanographiques
 - Photo-identification

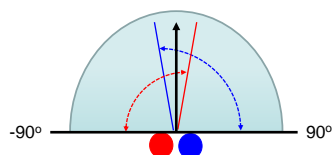
Données d'effort – ligne de prospection

- Suivre les transects désignés
- Enregistrer l'heure et la position régulièrement
- Déviations de la ligne de transect
 - Rapprochement des animaux après détection (si en mode rapprochement, *closing mode*)
 - Retour au transect
 - Route vers le point de passage (*waypoint*)
 - Eviter les conditions défavorables
 - Houle
 - Reflets



Données d'effort – mode d'observation

- Un observateur
 - Recherche devant le navire de -90° à 90° (secteur de 180°)
- Deux observateurs
 - Chacun cherche de 90° de son côté à $\sim 10^{\circ}$ de l'autre côté de la ligne de transect
 - Assure une bonne couverture de la ligne



Données d'effort à collecter

- Nom du navire
- Hauteur de la plateforme d'observation
- Pour chaque enregistrement d'effort
 - Date
 - Heure
 - Position (GPS : latitude et longitude)
 - Mode d'effort
 - En / hors effort
 - Equipe simple ou double
 - Nom des observateurs
 - Conditions de prospection
 - Etat de la mer
 - Houle
 - Reflet
 - Visibilité

Nouvel enregistrement d'effort à chaque fois qu'un de ces éléments change

Données d'observation à collecter

- Date / Heure
 - Position (lat, lon)
 - Angle et distance
- } **Immédiatement**
- Élément de détection (splash, nageoire, souffle, etc)
 - Observateur / position
 - *Aspect* (direction de nage des animaux quand repérés)
 - Espèce
 - Taille de groupe
 - Comportement
 - Présence de jeunes



Protocole de collecte de données d'observation

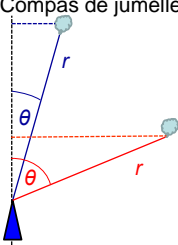
- Quand des animaux sont repérés...
 - Passer en "hors effort"
 - Donner un nouveau numéro d'observation
- Immédiatement :
 - Angle et distance
 - Heure
 - Position
- Puis :
 - Élément de détection
 - Observateur / position
 - *Aspect*
- Mode passif (*Passive mode*) ou mode rapprochement (*Closing mode*) ?

Protocole de collecte de données d'observation

- Mode passif (*Passing mode*)
 - Continuer le long du transect
 - Enregistrer l'espèce et la taille de groupe au plus vite
 - Retourner en prospection "en effort"
- Mode de rapprochement (*Closing mode*)
 - Approcher les animaux
 - Enregistrer l'espèce et la taille de groupe
 - Collecter des données supplémentaires (comportement, photo-id, etc)
 - Retourner en prospection "en effort"
 - Retour au transect
 - En direction du point de passage (*waypoint*)

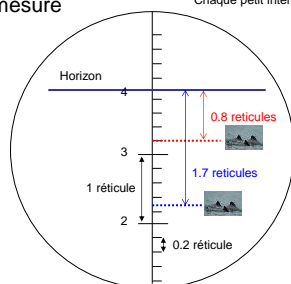
Données d'observation - Angles

- Très important de collecter des données d'angles correctes
- Les mesures d'angle ont le plus gros effet sur la distance perpendiculaire proche de la ligne de transect
- Rapporteurs
- Compas de jumelles



Données d'observation - Distances

- La précision des données de distance est également très importante
- Jumelles avec réticules
- Bâton de mesure

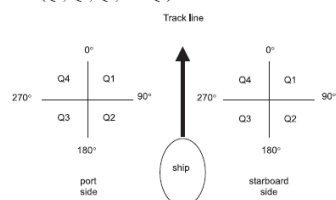


Noter le nombre de réticules entre la ligne de flottaison de l'animal et l'horizon

Données d'observation - Aspect

- Direction de nage des animaux lorsque repérés
 - Par rapport à la ligne de transect
- Peut être utile pour explorer le mouvement de réponse

Fig. 1. Illustration of definition of swimming directions and quadrants (Q1, Q2, Q3, and Q4).



Palka and Hammond (2001). Can. J. Fish. Aquat. Sci. 58: 777-787

Données d'observation – Taille de groupe

- Considérer le meilleur moyen d'estimer le nombre
 - Selon l'espèce, la taille du groupe, etc
- Basse – Meilleure – Haute
- Groupe multi-espèces
- Sous-groupes
- Classes d'âge / de sexe
- Nombre de jeunes



Logiciel de collecte de données

- Utiliser un programme de collecte de données permet :
 - De maintenir le protocole de collecte de données
 - D'assurer la précision des données
- Exemple d'un bon programme pour la collecte de données depuis un navire : **Logger**
 - International Fund for Animal Welfare (IFAW)
 - Distribué par Marine Conservation Research

<http://www.marineconservationresearch.co.uk/downloads/logger-2000-rainbowclick-software-downloads/>

Résumé

- Un protocole strict de collecte de données est indispensable
- Les données doivent être le plus précis possible
- Les données imprécises :
 - Diminuent la précision des estimations d'abondance
 - Peuvent entraîner des biais
 - Identification des espèces
 - Taille de groupe
 - Angles et distances

Saisir des données dans DISTANCE

Le logiciel DISTANCE est la « norme industrielle » pour estimer la densité et l'abondance à partir de données de transects de ligne.

1. Données

DISTANCE nécessite que les données soient importées depuis un fichier texte. La première étape consiste donc à enregistrer le fichier de données Excel comme un fichier de texte (séparateur : tabulation) avec le même nom.

DISTANCE organise les données dans une structure hiérarchique contenant des couches selon le type de données. Les couches de données par défaut sont **Global** (l'étude en général) ; **Stratum/Region** (sous-zones / blocs) ; **Sample/Line Transect** (effort de prospection) ; et **Observation**. Les données sont réparties dans ces couches de la manière suivante :

- **Stratum/Region** : Block, Area
- **Sample/Line Transect** : Transect, Transect Length
- **Observation** : GroupSize, PerpDistance

2. Projet DISTANCE

La première étape dans DISTANCE est de créer un Projet qui contiendra les données et les résultats des analyses.

Ouvrir DISTANCE et sélectionner **File -> New Project** pour ouvrir la fenêtre **Create Project**. Entrer un nom de fichier pour le projet (ex : SCANS) pour accéder à la fenêtre **New Project Setup Wizard**.

Dans **Step 1 : Type of Project**, sélectionner **Analyze a survey that has been completed (Analyser une prospection qui a été réalisée)**.

Le bouton **Help** en bas à gauche renvoie vers l'Aide DISTANCE qui est très complète et très utile.

Le deuxième écran, **Step 2 : Setup for Analyzing a Survey (Mise en place pour analyser une prospection)**, donne des informations sur la structure des données (voir plus haut). Passer à l'écran suivant.

Dans **Step 3 : Survey Methods**, choisir :

- Type of survey : Line transect
- Observer configuration : Single observer
- Distance measurements : Perpendicular distance
- Observations : Clusters of objects (groupes d'objets)

Sur l'écran suivant, **Step 4 : Measurement Units**, sélectionner :

- | | |
|---|----------------------------|
| - Distance : Meter (ou Metre) | [distance perpendiculaire] |
| - Transect : Kilometre | [effort de prospection] |
| - Area : Square Kilometre (Kilomètres carrés) | [bloc / zone d'étude] |

Passer à l'écran suivant.

Step 5 : Multipliers (Mutiplicateurs) est utilisé pour mettre à l'échelle les estimations avec des constantes fournies par le logiciel. Ignorer cet écran et passer à l'écran suivant, **Step 6 : Finished**, pour cliquer sur **Proceed to Data Import Wizard**.

3. Saisir des données

Step 1 : Introduction informe sur ce qui doit déjà être fait avant d'en arriver à ce point.

Dans **Step 2 : Data Source**, sélectionner le fichier de données dans la fenêtre **File Containing Data to Import (Fichier contenant les données à importer)**. Rappel : ce doit être un fichier texte, pas Excel.

Step 3 : Data Destination (Destination des données) fournit des informations à DISTANCE sur la façon d'incorporer les données dans la structure de données.

Dans **Destination data layers (Destination des couches de données)**, sélectionner Observation pour *Lowest data layer (couche de données la plus basse)* et Region (Stratum) pour *Highest data layer (couche de données la plus haute)*. Dans **Location of new records (Localisation des nouveaux enregistrements)**, sélectionner *Add all new records under the first record in the parent data layer (Ajouter tous les nouveaux enregistrements sous le premier enregistrement de la couche de données d'origine)*. Under **Creation of new records in the lowest data layer**, sélectionner *Create one new record for each line of the import file (Créer un nouvel enregistrement pour chaque ligne du fichier importé)*.

Dans Step 4 : Data File Format (Format du fichier de données), sélectionner :

- *Delimiter (Délimiteur)* : Tab (Tabulation)
- *Ignore rows (Ignorer des lignes)* : Do not import first row (ne pas importer la première ligne)
- *Decimal symbol (Symbole des décimales)* : Use « . »

Grid size affiche la taille du fichier de données. Les premières lignes des données sont visibles et doivent être affichées de la même manière que dans le fichier Excel.

Dans **Step 5 : Data File Structure (Structure du fichier de données)**, DISTANCE indique à quelle couche chaque colonne ou champ (« field ») de données doit être attribuée. Ceci doit être fait correctement car il ne sera pas possible de le modifier par la suite dans DISTANCE.

Il faut également donner à chaque champ un nom dans *Field name* et renseigner un type de données dans *Data type*. Dans certaines couches de données il y a des champs obligatoires avec des noms et des types de données spécifiques. Ce sont :

- *Area* dans **Region** (couche Region/Stratum) ;
- *Line Length* dans **Line transect** (couche Sample) ;
- *Perp distance* et *Cluster size* dans **Observation** (couche Observation).

Note : dans la couche **Observation** il est plus simple de commencer par les colonnes de distance perpendiculaire et de taille de groupe avant d'attribuer d'autres champs.

Attribuer les données aux couches et aux champs

Pour attribuer la première colonne de données (Block) :

Cliquer sur [Ignore] dans la ligne *Layer name* en haut de la première colonne de données (Block) et sélectionner **Region**. Cliquer sur la ligne en-dessous, *Field name*, et entrer **Block**. Cliquer sur la ligne suivante, *Field type*, et choisir **Texte**.

Attribuer maintenant la colonne Area.

Cliquer sur [Ignore] dans la ligne *Layer name* et sélectionner **Region**. Cliquer sur la ligne suivante, *Field name*, et sélectionner **Area**. Cliquer sur la ligne suivante, *Field type*, qui se met par défaut en **Decimal**.

Attribuer maintenant les colonnes de données *Transect number* et *Transect length* à la couche **Line transect**.

Transect number : *Layer name* > **Line transect** ; *Field name* > entrer **Transect Number** ; *Field type* > choisir **Integer**.

Transect length : *Layer name* > **Line transect** ; *Field name* > **Line Length** ; *Field type* se met par défaut en **Decimal**.

Pour finir, attribuer les colonnes de données *Species*, *GroupSize* et *PerpDistance* à la couche **Observation**.

Rappel : il est préférable de commencer par la distance perpendiculaire et la taille de groupe.

Distance perpendiculaire : *Layer name* > **Observation**, *Field name* > **Perp distance** ; *Field type* se paramètre en **Decimal** par défaut.

Taille de groupe : *Layer name* > **Observation**, *Field name* > **Cluster size** ; *Field type* est automatiquement classé comme **Decimal**.

Espèce : *Layer name* > **Observation**, *Field name* > entrer **Species** ; *Field type* > choisir **Texte**.

Dans **Step 6 : Finished**, choisir *Existing data – Overwrite existing data*.

Choisir *Save current setting as default* (Enregistrer les paramètres actuels par défaut) si vous le souhaitez.

Cliquer sur **Finish** et croiser les doigts...

4. Les données dans DISTANCE

Si l'importation des données est acceptée, l'écran **Project Browser** apparaît. Comme précédemment, il y a une barre d'outils en haut de la fenêtre DISTANCE avec des icônes en-dessous pour effectuer des tâches particulières. Dans le **Project Browser**, il y a 6 onglets.

L'onglet **Data** a deux fenêtres. Celle de gauche indique les couches de données et celle de droite le contenu de ces couches. Sélectionner *Region* depuis la fenêtre **Data layers** pour afficher ces données. Sélectionner *Line Transect*, puis *Observation* pour voir le reste des données.

Il est possible d'insérer ou de supprimer des couches, des champs, des enregistrements et d'éditer les données tant que le bouton **Lock data sheet (Verrouillage des feuilles de données)**, 4^{ème} depuis la gauche dans la barre d'outils, n'est pas sélectionné.

Il n'est pas nécessaire d'enregistrer les modifications dans DISTANCE, tout est enregistré automatiquement.

Fermer DISTANCE. Le projet est enregistré comme un fichier **.dst** avec un dossier **.dat** qui contient les données dans une base de données Access.

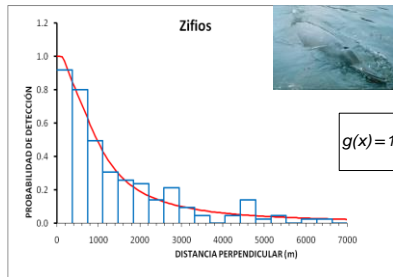
5. Exporter des projets DISTANCE

Pour envoyer un projet à quelqu'un, il faut tout d'abord l'exporter depuis DISTANCE.

Ouvrir de nouveau le projet dans DISTANCE.

Cliquer sur **File > Export project** depuis la barre d'outils. Dans **Type**: choisir d'exporter en tant que fichier de projet DISTANCE (fichier **.dst** avec données associées) ou tout rassemblé dans un fichier zip. L'ensemble du projet (Data, Survey Designs, Surveys, Analyses) sera exporté à moins d'exclure spécifiquement certaines parties en cochant les cases correspondantes dans *Exclude the following parts of the project*.

Transect de ligne – Analyse des données



$$g(x) = 1 - \exp \left[- \left(\frac{x}{\sigma} \right)^{-\beta} \right], \quad x \leq w$$

$$AIC = -2 \log_e(L) + 2q$$

$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$

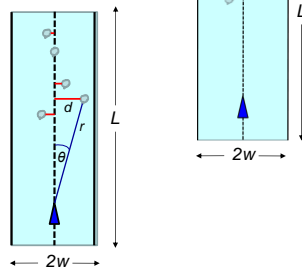
Estimation d'abondance – résumé

- Dans un transect de bande (*strip transect*) on considère que tout est détecté

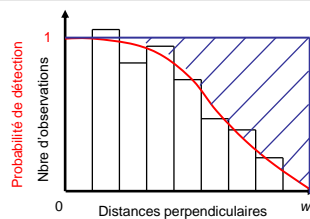
$$\hat{D} = \frac{n}{a} = \frac{n}{2wL}$$

- Dans un transect de ligne, tout n'est pas détecté
 - Besoin d'estimer la probabilité de détection, p_a
 - p_a est estimée avec les données de distances perpendiculaires

$$\hat{D} = \frac{n}{2wL\hat{p}_a}$$



Estimer la probabilité de détection, p_a



Probabilité de détection moyenne : $\hat{p}_a = \frac{\text{aire sous la courbe}}{\text{aire sous le rectangle}}$

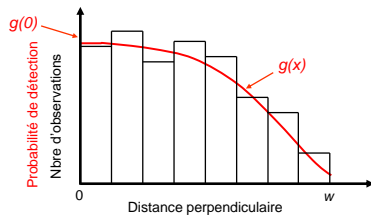
Hypothèse importante : les animaux sont distribués uniformément entre 0 and w

Echantillonnage par transect de ligne - Hypothèses

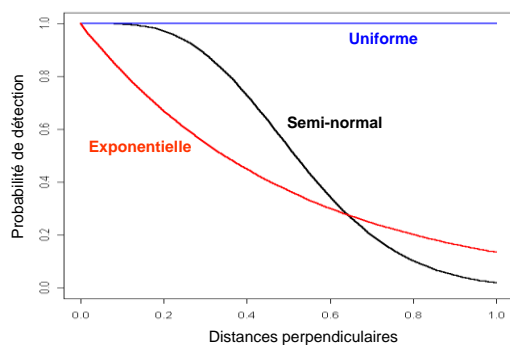
- Toutes les mesures sont justes
- Tous les objets sur la ligne de transect sont détectés
 - Probabilité de détection = 1 à distance perpendiculaire 0
- Tous les objets sont détectés à leur position initiale
 - Pas de mouvement avant la détection
 - Particulièrement en réponse à la plateforme de prospection
- Les données collectées sont représentatives
- Les observations sont indépendantes
 - Pour une estimation de variance non biaisée

Critères pour une bonne fonction de détection

- Un modèle qui :
 - Peut s'adapter à une grande variété de formes possibles (modèle robuste)
 - Peut s'adapter aux données lorsque de nombreux facteurs affectent la détectabilité (*pooling*)
 - A une "épaule" : $g'(0)=0$ (forme)
 - Génère des estimations avec une bonne précision (efficace)



Quelques fonctions de détection (*detection functions*)



Modèles de fonctions de détection

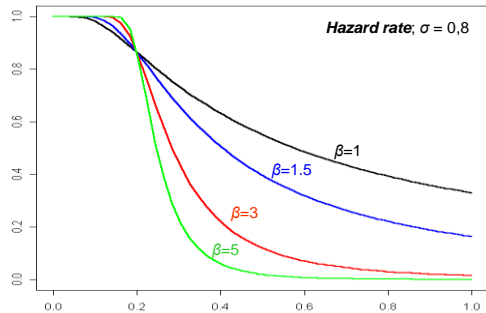
Uniforme : $g(x) = 1, x \leq w$
(Uniform)

Semi-normal : $g(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right), x \leq w$
(Half-normal)

Taux de risque : $g(x) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^\beta\right], x \leq w$
(Hazard-rate)

x = distance perpendiculaire
 w = distance de troncature (largeur nominale de bande)
 σ = écart-type des distances perpendiculaires
 β = paramètre à estimer

Fonction de détection Hazard-rate



Termes d'ajustements (*Adjustment terms*) Améliorer la robustesse

- Ajustements en série (*series adjustments*) de la fonction clé (*key function*) :

$$g(x) = \text{fonction clé} \times (1 + \text{ajustement en série})$$

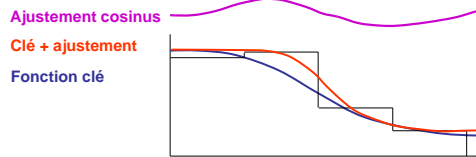
Ajustements en série

Cosinus

Polynôme simple

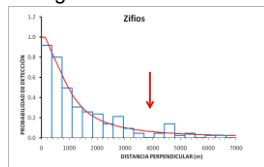
Polynôme d'Hermite

Comment fonctionnent les ajustements ?



Troncature des données (*data truncation*)

- Les observations à grande distance contribuent peu à l'estimation de la forme de la fonction de détection
- Mais peuvent causer une mauvaise adaptation du modèle et une grande variance



- Considérer la troncature avant d'adapter le modèle
 - Pas rare de tronquer autour de 5 % des observations
 - Mais nécessité de décider au cas par cas
 - Une distance particulière est peut-être préférable

Critère d'Information d'Akaike (*Akaike's Information Criterion*)

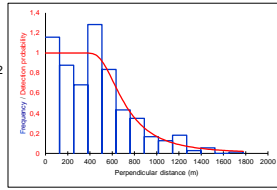
$$AIC = -2 \log_e(L) + 2q$$

- **L** : maximum de vraisemblance (*maximized likelihood*)
 - La vraisemblance est une fonction statistique reliant les paramètres d'un modèle aux données
- **q** : nombre de paramètres du modèle
- Utilisé pour comparer les modèles
- Les modèles avec l'AIC le plus faible est le plus soutenu par les données
 - Deux modèles avec une différence d'AIC de moins de 2 unités sont considérés comme étant soutenu de manière équivalente
- A n'utiliser que si les données sont les mêmes
 - Donc pas de comparaison de modèles avec des distances de troncature différentes
- Mesure d'adaptation relative

Evaluation de l'adaptation du modèle aux données

- Tests de qualité d'ajustement (*Goodness of fit test*)

- Test Khi carré (*Chi-squared*), χ^2
 - Pour données groupées
 - $\sum [(Obs - Exp)^2 / Exp]$



- Diagramme Q-Q
- Test Kolmogorov-Smirnoff (KS)
- Test Cramér-von-Mises (CvM)

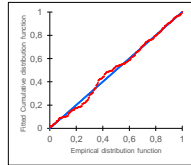
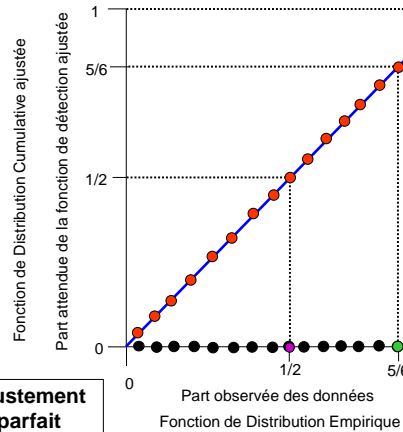
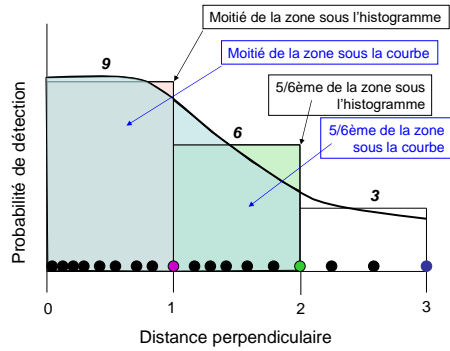
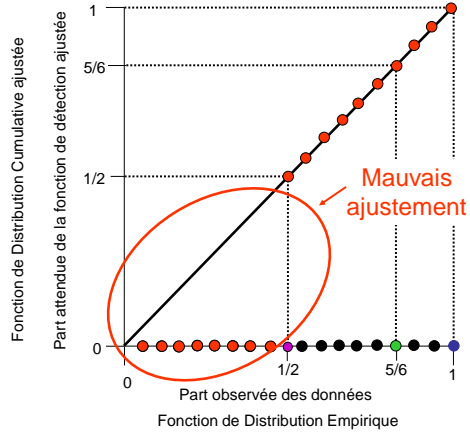
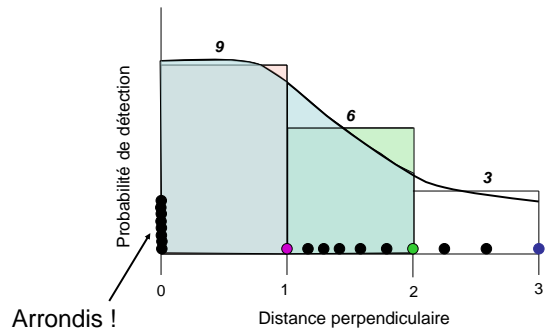


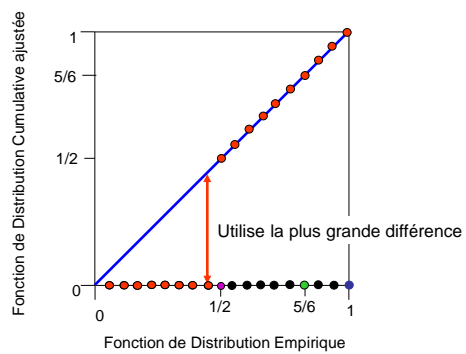
Diagramme Q-Q



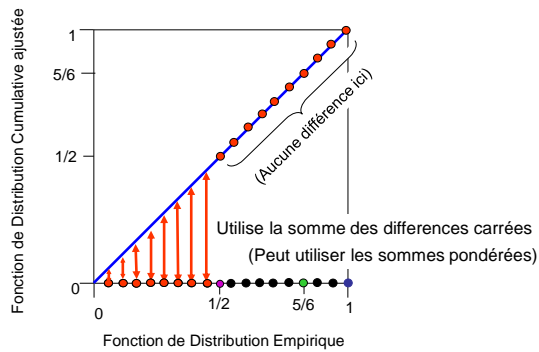
Exemple : Arrondir à zéro



Test Kolmogorov-Smirnov



Test Cramér-von-Mises



Tests de qualité d'ajustement - résumé

- Le Khi carré nécessite des données groupées
 - Les résultats dépendent du regroupement
- Le diagramme Q-Q montre une qualité d'ajustement "haute resolution"
 - Sans regrouper les données en intervalles
- Tests Kolmogorov-Smirnov and Cramér-von-Mises
 - Également sans besoin de regroupement
- Le test Cramér-von-Mises peut être pondéré
 - Pour donner un poids plus important aux observations proches de 0 distance

Résultats

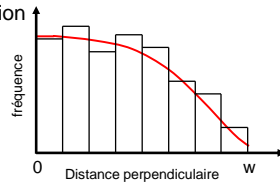
- Estimations des paramètres
 - p_a = probabilité de détection moyenne en-deça de la distance de troncature w
 - esw = *effective strip width* = demi-largeur effective de bande $p_a \times w$
 - $f(0) = 1 / esw$
 - n/L = taux de rencontre
 - D_g = densité de groupes
 - $E[s]$ = taille de groupes
 - D = densité d'animaux
 - N = abondance d'animaux

Estimation de la variance

- Chaque variable ou estimation possède une variance
- Très important :
 - Plus la variance est faible, plus l'estimation est potentiellement proche de la vérité
- Expression :
 - Erreur type (*standard error*)
 - Coefficient de variation (*coefficient of variation*)
 - Intervalle de confiance (*confidence interval*)
- Mode de calcul :
 - Analytique
 - Bootstrap

Résumé

- Estimer la fonction de détection
 - Modèle
 - Fonction clé
 - Termes d'ajustement
 - Tronquer les données
 - Sélection des modèles
 - AIC
 - Qualité d'ajustement
- Estimation de la variance
 - Analytique
 - Bootstrap



Première approche de l'analyse des données dans DISTANCE

1. Analyser les données

Ouvrir DISTANCE en double-cliquant sur le fichier **.dst**. Dans la fenêtre **Project Browser**, cliquer sur l'onglet **Analyses** où sont paramétrés et stockées les analyses. Dans la barre d'outils, il y a des icones :

- **Set** (Organiser) : organiser les analyses
- **Analysis** : créer, supprimer, sélectionner des détails, lancer et déplacer les analyses.

Ces fonctions sont également accessibles depuis **Analyses** dans la barre d'outils principale de DISTANCE.

Sous la barre d'outils du Project Browser se trouve deux fenêtres. Celle de gauche contient les noms etc. des différentes analyses. Celle de droite affiche les résumés des diagnostics et des résultats des analyses.

Pour créer une première analyse, sélectionner **Analysis > Analysis Details** depuis la barre d'outils du **Project Browser** ou cliquer sur la 3^{ème} icône depuis la gauche à côté de **Analysis**: dans la barre d'outils **Analyses**. Cela ouvre l'écran **Analysis Details**. Il y a 3 onglets sur la droite. L'onglet **Input** est ouvert.

1.1 Sélection des données

Tout d'abord, il est nécessaire de sélectionner les données à analyser en créant un **Data filter (Filtre de données)**, qui correspond à la fenêtre au milieu de l'écran. Cliquer sur **Properties...** sur la droite de cette fenêtre pour ouvrir la fenêtre **Data Filter Properties**. Il y a 4 onglets.

L'onglet **Data selection** permet d'indiquer à DISTANCE quelles données analyser. Pour cette première analyse, utiliser toutes les données ; il n'est donc pas nécessaire de faire une sélection. Le reste de cet onglet et les trois autres onglets peuvent être ignorés. Donner simplement un nom au Data filter comme « All data » et cliquer sur **OK**.

1.2 Modèle de fonction de détection

Définir maintenant le modèle pour ajuster la fonction de détection pour estimer la probabilité de détection moyenne.

Cliquer sur **Properties...** sur le côté droit de la fenêtre **Model definition** en bas de l'écran pour ouvrir la fenêtre **Model Definition Properties**. En haut, dans **Analysis Engine**, sélectionner le type d'analyse à réaliser. Les choix sont les suivants :

CDS – Conventional distance sampling	Echantillonnage de distances conventionnel. Analyse basique de données de transect de ligne.
MCDS – Multiple covariate distance sampling	Echantillonnage de distances avec plusieurs covariables. Permet d'introduire des covariables lors de l'ajustement de la fonction de détection
MRDS – Mark-recapture distance sampling	Echantillonnage de distances avec marque-recapture. Utilise un double jeu de données pour estimer $g(0)$ et prendre en compte le mouvement de réponse des animaux
DSM – Density surface modelling	Modélisation de densité de surface. Estimation d'abondance basée sur un modèle.

Choisir **CDS – Conventional distance sampling** pour cette analyse.

Cette fenêtre a six onglets.

Tout d'abord, sélectionner l'onglet **Estimate**. Dans **Stratum definition**, sélectionner **Use layer type** ; la seule option pour **Layer type** est *Stratum*. Dans **Quantities to estimate and level of resolution (Quantités à estimer et niveau de résolution)**, cocher la case *Stratum* en haut pour cocher toutes les autres quantités. Si ces cases

sont cochées, DISTANCE va estimer la densité, le taux de rencontre, la fonction de détection et la taille de groupe séparément pour chaque strate (dans notre exemple, les blocs).

Le taux de rencontre doit être estimé séparément pour chaque strate. Cependant, du fait que la fonction de détection est généralement indépendante des strates, celle-ci est souvent estimée pour l'ensemble des données rassemblées (au niveau Global par exemple). Cocher la case dans la colonne *Global* pour *Detection function*.

Laisser *Cluster (group) size* au niveau *Stratum* bien qu'il soit possible d'estimer la taille de groupe au niveau *Global*.

Laisser *Density* cochée aux niveaux *Global* et *Stratum*. DISTANCE va ainsi calculer une estimation de densité au niveau global et pour chaque strate. L'estimation globale de densité sera la moyenne des densités de chaque strate pondérée par leur surface.

S'il n'y avait qu'une seule zone d'étude (pas de strates), ou si l'on voulait une estimation non stratifiée, laisser **Stratum definition** avec **No stratification** et toutes les estimations seraient faites au niveau global.

Sélectionner l'onglet **Detection function** qui permet de définir comment ajuster la fonction de détection. Il contient quatre onglets.

Utiliser l'onglet **Models** pour sélectionner la forme de la fonction de détection dans la liste des **Key functions** (modèles basiques) et des **Series expansions** (la forme de la fonction d'ajustement). Pour cette analyse, conserver les paramètres par défaut : **Half-normal** pour *Key function* et **Cosine** pour *Series expansions*.

Utiliser l'onglet **Adjustement terms** pour sélectionner la façon d'ajouter les termes d'expansion de série (« *Series expansion terms* »). Pour cette analyse, laisser sur **Automated adjustment** pour laisser DISTANCE choisir le nombre éventuel de termes d'ajustement nécessaires pour ajuster le meilleur modèle en utilisant l'AIC.

Ignorer les onglets **Contraints** (Contraintes) et **Diagnostics**.

Sélectionner l'onglet **Cluster Size** dans la ligne au-dessus.

Dans l'échantillonnage par transect de lignes d'animaux en groupe, il y a une tendance à manquer les plus petits groupes lorsque les distances perpendiculaires sont plus grandes. Cela peut entraîner un biais dans l'estimation de la taille de groupe moyenne. Ce biais peut être pris en compte en régressant la taille de groupe (ou le log de la taille de groupe) par la distance perpendiculaire (voire la probabilité de détection estimée).

Pour ce faire, DISTANCE propose plusieurs options ; choisir la troisième : **Use size bias regression method if regression significant (Utiliser une méthode de régression du biais de la taille si la régression est significative)**. Laisser alpha-level à 0.15. Dans **Size-bias regression method (Méthode de régression du biais de la taille)**, choisir la troisième option : **Regress ln(cluster size) against distance x (Régresser ln(taille de groupe) contre la distance x)**.

Ignorer les autres onglets mais donner un nom au modèle (ex : Half-normal) et cliquer sur **OK**.

La dernière étape consiste à donner un nom à l'analyse en haut de l'écran **Analysis Details**. Utiliser quelque chose de bref mais clair comme « All data HN » (pour représenter l'analyse de toutes les données avec la fonction de détection Half-normal).

Lancer l'analyse en cliquant sur le bouton **Run** en haut à droite de l'écran.

S'il y a des avertissements, l'onglet **Log** va se colorer en orange. Généralement, les avertissements n'invalident pas l'analyse et peuvent souvent être ignorés.

Par exemple :

**** Warning : Parameters are being constrained to obtain monotonicity. ****
(Avertissement : les paramètres ont été contraints pour obtenir la monotonie).

Cela se rapporte au fait de s'assurer que la fonction de détection a la bonne forme (décroissance monotone : la probabilité de détection ne peut pas augmenter avec la distance perpendiculaire). C'est acceptable et l'avertissement peut être ignoré.

Selon la taille du jeu de données (nombre d'observations), il peut y avoir un autre avertissement tel que :

**** Warning : only 500 out of a total of 549 observations will be shown on the qq plot. ****
(Avertissement : seules 500 des 549 observations seront affichées sur le diagramme Q-Q)

Ceci est un paramètre de DISTANCE par défaut. Pour voir toutes les données sur le diagramme Q-Q, sélectionner l'onglet **Input**, ouvrir la fenêtre **Model Definition Properties**, sélectionner l'onglet **Detection function** puis l'onglet **Diagnostics**. Saisir, par exemple, 1500 pour *Maximum number of points in qq plots* et cliquer sur **OK**.

La fenêtre **Confirm Change** apparaît, comme à chaque fois qu'une modification de *Model Definition* intervient alors qu'une analyse a déjà été lancée. Ceci afin d'être sûr que ce n'est pas une modification d'un modèle existant par erreur, plutôt que la création d'un nouveau modèle (erreur facile).

Dans le cas présent, la modification est voulue, donc sélectionner **Yes**.

Retourner au **Project Browser** en sélectionnant **View > Project Browser** ou en fermant l'écran **Analysis Details** (attention de ne pas fermer DISTANCE).

Le point à gauche de cette analyse est gris. Cela signifie qu'il faut lancer l'analyse à nouveau (car *Model definition* a été modifié). Cela sera également le cas pour toute analyse utilisant cette définition de modèle.

Sélectionner **Analyses > Run analysis** (ou le bouton jaune avec le bonhomme qui court) pour lancer de nouveau le modèle.

Ouvrir de nouveau la fenêtre **Analysis Details** en sélectionnant **Analysis > Analysis Details** depuis la barre d'outils du **Project Browser** ou en cliquant sur la 3^{ème} icône depuis la gauche à côté de **Analysis**: dans la barre d'outils **Analyses**.

S'il n'y a pas d'avertissements, l'onglet **Results** sera vert. Si l'analyse échoue, l'onglet **Log** sera rouge. Le point à gauche de l'analyse dans le **Project Browser** sera de la même couleur.

1.3 Results

L'onglet **Results** contient plusieurs pages que l'on peut faire défiler depuis le haut, notamment :

- Les options choisies pour l'analyse ;
- Les détails de l'ajustement du modèle et le meilleur modèle retenu, comprenant un résumé des données saisies :
 - o En haut des pages 2 et 3, *Effort* correspond à la longueur totale des transects, *# samples* correspond au nombre de transects, *Width* à la distance perpendiculaire maximale (ou la largeur de troncature), et *# observations* au nombre d'observations.
- Les résultats pour le modèle le mieux ajusté, comprenant la probabilité de détection moyenne (p) et la largeur de demi-bande effective (*Effective Strip Half-Width, ESW*) ;
- Les diagnostics pour savoir à quel point le modèle de fonction de détection est ajusté aux données :
 - o Diagramme Q-Q ;
 - o Tests de qualité d'ajustement Kolmogorov-Smirnov et Cramer-on-Mises ;
 - o Graphiques de la fonction de détection avec les histogrammes des données observées et résultats du test de qualité 'ajustement « Khi carré » pour plusieurs intervalles de largeur différents.

- Pour chaque strate, les résultats de la régression de la taille de groupe et un graphique (brut) de la ligne de régression ;
- Pour chaque strate, les estimations de densité des groupes (DS), de taille de groupe (E(S)), de densité des animaux (D) et d'abondance (N) avec leurs erreur type, coefficient de variation et intervalle de confiance à 95 % ;
- Des pages de résumés comprenant l'estimation d'abondance globale sur la dernière page.

Lorsque ces différents résultats ont été examinés, retourner au **Project Browser** pour voir le résumé de l'analyse.

Dans la fenêtre de gauche, la couleur du point indique si l'analyse a tourné sans avertissements (vert) ou avec (orange), si elle a échoué (rouge) ou si elle n'a pas encore été lancée (gris). Placer la souris sur le « 1 » de la 4^{ème} colonne pour voir le nom du filtre de données (*Data Filter*) utilisé dans cette analyse. Faire de même avec la colonne suivante pour voir le nom de la définition du modèle (*Model Definition*) utilisé dans cette analyse.

Dans la fenêtre de droite, il y a des colonnes pour :

- # parameters : le nombre de paramètres dans la fonction de détection ajustée ;
- Delta AIC : la différence d'AIC entre ce modèle et le modèle le mieux ajusté ;
- AIC – *Akaike's Information Criterion* : Critère d'Information d'Akaike, une mesure relative de l'ajustement d'un modèle ;
- ESW/EDR : Demi-largeur effective estimée (*Estimated Strip Half-Width*) ;
- D, D LCL, D UCL, D CV : densité d'animaux avec les limites inférieure (LCL, *Lower Confidence Limit*) et supérieure (UCL *Upper Confidence Limit*) de l'intervalle de confiance et le coefficient de variation.

Il est possible de personnaliser cette fenêtre en ajoutant / supprimant des colonnes en utilisant **Analysis > Arrange columns** (ou en cliquant sur le bouton à l'extrémité droite de la barre d'outils). CEPENDANT, une meilleure façon de faire est de paramétrer ces colonnes par défaut pour ne pas avoir à les modifier à chaque nouvelle analyse. Pour ce faire, sélectionner **Analyses > Preferences...** dans la barre d'outils DISTANCE pour ouvrir la fenêtre **Distance Preferences...** Dans *Analysis browser window*, cliquer sur *Analysis browser*. Ajouter des colonnes supplémentaires (ou retirer des colonnes) en utilisant les flèches pour les déplacer de droite à gauche (ou inversement).

Les colonnes utiles à ajouter :

- **GOF K-S p** et **GOF CvM (cos) p** : résultats des tests de qualité d'ajustement Kolmogorov-Smirnov et Craver-von-Mises ;
- **P, P CV** : probabilité de détection ;
- **DS, DS CV** : densité des groupes ;
- **CS, CS CV** : taille des groupes ;
- **N, N CV, N LCL, N UCL, N CV** : abondance avec les limites inférieure et supérieure de l'intervalle de confiance et le coefficient de variation.

Les intervalles de confiance de densité (D CV) et d'abondance (N CV) sont les mêmes donc il n'est pas nécessaire d'avoir les deux.